Supporting Discovery through Contextual Representation: Developing a Tool for Visually Exploring Slave Narratives

Annie T. Chen, Jessica Martell, and Pamella Lach

Abstract— When an individual interacts with a set of documents, their understanding is informed by a process of exchange between reader and text. As such, any tool that facilitates this interaction has the potential to alter the experience in profound ways. In this paper, we explore these potential influences through a case study. The Slave Narrative Name and Place Project sought to develop a comprehensive list of people and places mentioned across the collection of Documenting the American South's North American Slave Narratives (http://docsouth.unc.edu/neh/). The intent was to employ automated extraction via Natural Language Processing, combined with human annotation, to generate ways of exploring the collection. Using a sample of twelve texts, we developed multiple points of entry, displayed in a visually interactive web-based tool: 1) an affect tapestry, generated through automated sentiment analysis; 2) a character tapestry, which visualizes persons mentioned in the narratives by social role (e.g. slave, slaveholder, mistress, religious figures, etc.) based on human annotation work; 3) and indices of people and geographical entities mentioned in the texts. We explore the challenges that we faced in the development of this tool, including creating an appropriate classification schema for visualizing social roles. This project raises important questions about visualizing humanistic objects. How can visual tools be used to support exploration and analysis of historical corpora? What are important concerns regarding the historical and literary specificity in representing persons? By presenting some of the challenges in the development of this visualization tool, we hope our work will inform the development of future text exploration tools in the digital humanities.

Index Terms—Slave narratives, named entity classification, interactive visual analytics, digital humanities

---

1 INTRODUCTION

In recent years, there has been increased interest in and use of text mining and visualization in the analysis of large-scale textual corpora. Visual analytics techniques can be useful for synthesizing information and deriving insight from large volumes of dynamic, ambiguous and potentially conflicting data, and facilitate discovery of the unexpected [1]. In the digital humanities, we have seen examples of the use of tools that combine text mining and visualization techniques to examine linguistic patterns [2, 3], explore word usage in the vicinity of named entities [4], and facilitate identification of documents fitting categories of interest [5]. In particular, it has been argued that text mining can afford a fresh look at familiar texts [2, 5].

However, there are also numerous challenges in developing useful tools for exploring literary and historical texts. For example, historical texts may employ literary, linguistic, and grammatical conventions that may be vastly different from commonly available training data, and place and person names often change over time. Readers engage with texts in their own way and may develop a relationship with the text that is influenced by their own cultural background and sensitivities. The meaning that a text takes on can depend upon the context in which it was produced and experienced, or in which it is later read. When a reader employs a tool to explore a document collection, the impressions and understandings form a complex interplay of viewer – tool – object, and thus, careful consideration of the development of these tools is paramount.

In this paper, we consider one example of how decisions in the data representation and the design of interactive visual analytics features can potentially affect users' experiences of the objects depicted. The Slave Narrative Name and Place Project involved the development of a tool for exploring the slave narratives in the University of North Carolina's Documenting the American South collection (http://docsouth.unc.edu/neh/). This collection is used by K-12 educators, college/university instructors, and scholars in diverse disciplines including History, English, American and African American Studies, and more. The collection contains at least 275 autobiographical, biographical, and fictional texts.

This paper describes the development of a visual tool to facilitate exploration of a sample of twelve texts, beginning with a description of the interface, proceeding on to the method through which the interface was developed and challenges encountered in development, and concluding with implications for interface design.

1.1 The interface

The interface was originally developed to facilitate analysis and exploration of narrative texts for a broad range of users, including those who may not have technical expertise. Each narrative is featured on a separate webpage, which supports exploration of different contextual elements of the narrative: persons, places, and affect expressed. There are two visualizations at the top of each narrative exploration page: an Affect Tapestry and a Character Tapestry. The cities mentioned in the narratives also appear on this page, followed by the text itself. Users are able to navigate through the text by clicking on a Chapter Selection Bar.

The intended purpose of the narrative exploration pages is to facilitate reading and critical analysis of the texts through different types of visual lenses. These are not meant to replace traditional reading styles, but rather, to inspire alternative ways of regarding a text. By linking the visualizations to the text, users can move between the at-a-glance visualizations and closer readings of the narratives. In this paper, we briefly introduce both visualizations but focus on the development of the Character Tapestry.

1.2 Affect tapestry

Affect Tapestries depict the frequency of emotion words throughout a text (Fig. 1). Chapters are plotted along the x-axis and affect prevalence along the y-axis. By scanning from left to right, one acquires a sense of affect prevalence over the course of a narrative. A proxy for affect prevalence was developed based on the number of times keywords from a given affect category appear in a chapter, taking into consideration the length of the chapter. Selection of a chapter using the Chapter Selection Bar results in the display of the corresponding text with affect keywords highlighted.

There are three modes of display: Basic, Advanced, and Custom. The Basic display mode features six categories: hope, joy, anger, anxiety, sadness, and fear; the Advanced display mode includes 12 affect types: anger, sadness, negative-fear, joy, anxiety, hope, love, misery, insecurity, hate, gloom and cruelty. The Custom display mode features conceptual categories that are salient in the slave narratives: religion, torment, punishment, death, escape, and freedom. The keywords for the affect types were drawn from WordNet-Affect [6], and the Custom category lexicon was developed by the first and second authors while reading and annotating the texts. An earlier version of this interface only featured the Advanced display mode; the Basic and Custom modes were added in response to feedback from a survey comparing the needs of K-12 educators and university instructors [7]. The K-12 educators felt that the interface was too complex for their needs; thus, the Basic mode was added to reduce the potential for information overload due to visual complexity. University educators sought more customization and historical specificity; hence we added the Custom category.
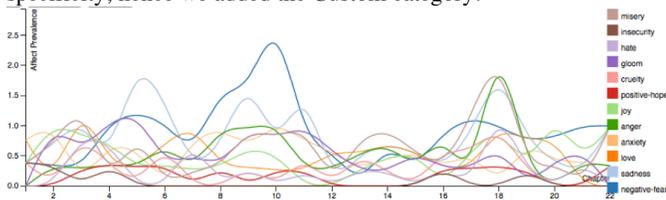

Fig. 1. Affect Tapestry: Northup, Solomon, *Twelve Years a Slave: Narrative of Solomon Northup, a Citizen of New-York, Kidnapped in Washington City in 1841, and Rescued in 1853*.

The Affect Tapestries might be employed to study the flow of a narrative. In the affect tapestry for *Twelve Years a Slave* (Fig. 1), for example, the climax in negative-fear in Chapter 10 corresponds with the violent altercation between Northup and Tibeats.

1.3   Character tapestry

The interface also features Character Tapestries to enable users to study the frequency of mention of persons throughout a narrative (Fig. 2). Person mentions are depicted through circles; the larger the circle, the more frequently a person appears in the chapter. The fill color of the circle is based on a classification of individuals into one of nine categories: slave, slaveholder, overseer, mistress, house slave/servant, religious, relations, abstract, and "other."
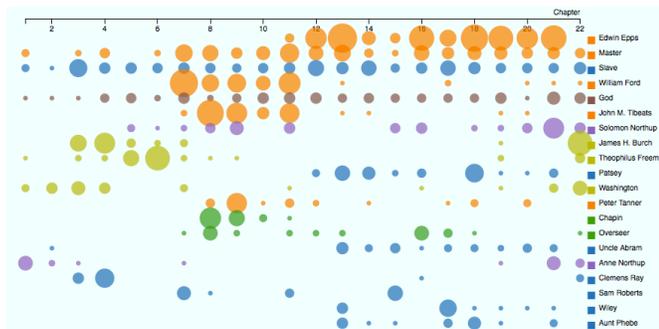

Fig. 2. Social Tapestry: Northup, Solomon, *Twelve Years a Slave: Narrative of Solomon Northup, a Citizen of New-York, Kidnapped in Washington City in 1841, and Rescued in 1853*.

By scanning from left to right, we can visually perceive the appearance and relative prominence of persons in a given narrative. If we consider the progression of orange circles in *Twelve Years a Slave*, we see how Northup was sold from one master to another (Fig. 2). We can also discern the relative frequency of Biblical, fictional, and abstract referents.

1.4   Potential uses of the interface

Considering the distribution of social roles across a text might provide insight into the historical world of the text by encouraging a reader to reflect upon the frequency and time points at which certain persons appear. Comparing narratives might suggest variations in linguistic style due to time period, or strategic uses of language in order to persuade or evoke particular reactions from readers (which arguably might occur in any narrative, but would be interesting to consider in this context since many of the narratives were written before the American Civil War to garner support for the abolitionist movement). Users of the interface can also select narratives that have particular distributions of roles (as depicted through the number, size, and colors of circles) for study, in order to identify parts of the text that may focus on social interactions of interest.

Juxtaposing the tapestries can enable users to identify narratives that may bear similarities in terms of plot, mood, and social setting. For example, the character tapestries for four of the narratives are shown in Figure 3. Two of the narratives tend to include persons who are primarily non-slaveholders (light green) or free (purple), and two of the narratives have social tapestries that are primarily comprised of slaves (blue) and slaveholders (orange). To a certain extent, the Affect and Character Tapestries serve as particular representations of the texts, in perhaps a similar way to 'literature fingerprints' in authorship attribution literature [8].

2   METHOD

The interface for this project employs a system built using PHP and MySQL, in conjunction with the D3.js library for visualization (https://d3js.org/). A workflow including both text mining and manual annotation was developed to prepare the slave narratives for display through the interface.

2.1   Pre-processing and named entity recognition

The corpus for this project came from the North American Slave Narratives digital collection, available on the DocSouth website. At the inception of this project, plain text versions of the slave narratives were not available. As a result, the HTML versions were initially used for pre-processing. As a first pre-processing step, the narratives were parsed into chapters. Subsequently, a pipeline was developed for named entity recognition of person and place names, followed by automated named entity classification of place names, and human annotation of social roles.

2.2   The annotation process

At the inception of this project, we first attempted to extract the named entities from the entire set of 275 narratives using the Stanford Named Entity Recognizer [9], and engage human annotators in the process of verifying these named entities in 35 narratives. In this initial phase, annotators did not actually read, or read carefully, the texts. Rather, they merely reviewed the NER output, turning to the narratives as needed to verify the entities. This resulted in high error rates, while revealing that the NER missed a significant number of entities. This led us to develop an extraction and verification workflow that required the human annotators to engage in close readings of the texts.

The project team members collaboratively and iteratively developed a coding scheme for social roles:

Slaveholder: refers to slave owners (but not limited to plantation owners). Children of slaveholders should be included in this category, but wives of slaveholders should be classified as Mistress.

Non-slaveholder: refers to any white person in the North or South who does not own slaves, excluding the children of slaveholders. This would include both abolitionists and poor white Southerners. White people whose slaveholding status is not clear can be classified as non-slaveholders.
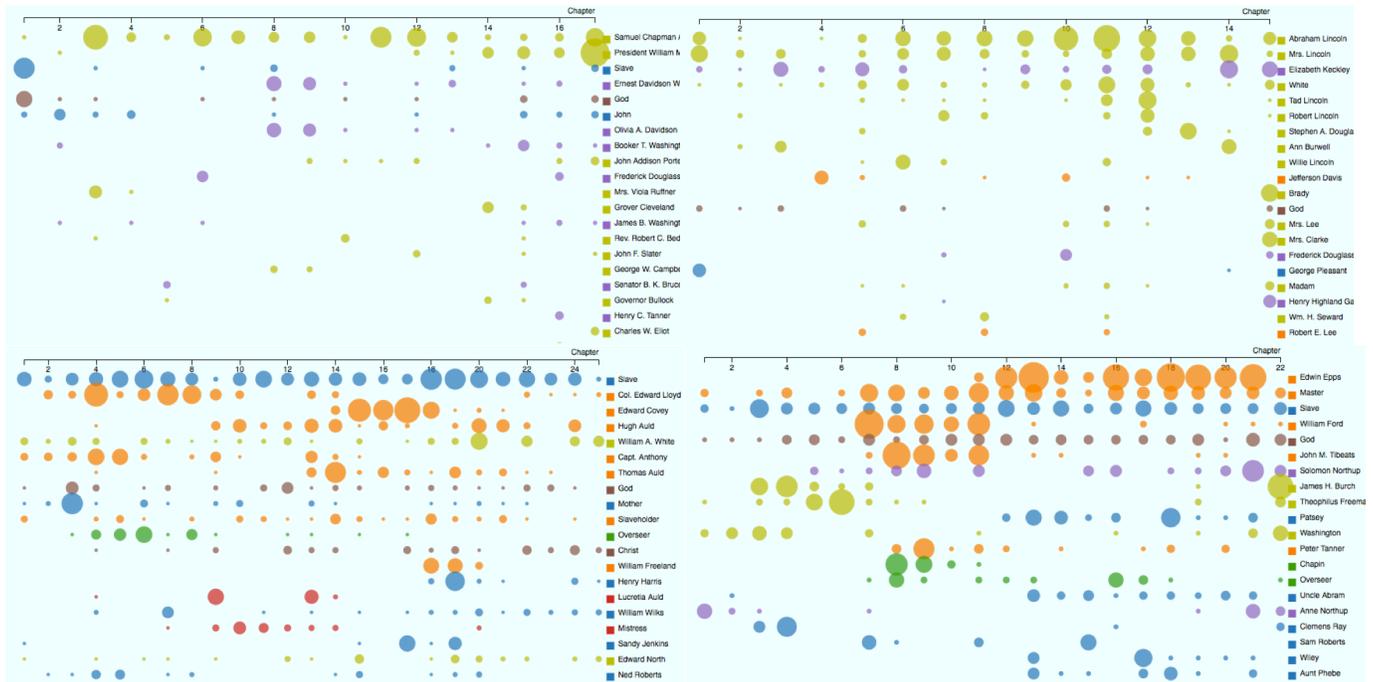
Fig. 3. Social Tapestries. Washington, Booker T., *Up from Slavery: An Autobiography* (top left). Keckley, Elizabeth, *Behind the Scenes, or Thirty Years a Slave, and Four Years in the White House* (top right). Douglass, Frederick, *My Bondage and My Freedom. Part I. Life as a Slave. Part II. Life as a Freeman* (bottom left). Northup, Solomon, *Twelve Years a Slave: Narrative of Solomon Northup, a Citizen of New-York, Kidnapped in Washington City in 1841, and Rescued in 1853* (bottom right).

Mistress: refers to the wife of the plantation owner or master.

Overseer: refers to men charged with managing enslaved laborers, most typically in a plantation setting; frequently described in extremely cruel terms. May be white or black.

Slave: refers to an enslaved person of any age; includes the children of enslaved persons.

Free person of color: refers to a person of color who is not a slave, including people born free and people who are no longer enslaved (manumitted persons); excludes runaway slaves.

Fugitive slave: refers to an enslaved person of color who is in the process of escaping slavery (runaway slave), or has escaped slavery but is not legally free.

Fictional or Biblical: refers to mythical, Biblical, and fictional people mentioned in the narrative.

Other: Refers to all other people that the narrator encounters over the course of the narrative.

3   DISCUSSION

There were various difficulties in the development of a coding scheme, or data representation, in large part due to the unique nature of the corpus. This coding scheme, in combination with the functionality of the interface, shapes how users come to view the narratives through the interface, in both good and limiting ways.

3.1   Challenges of the annotation process

Devising a usable coding schema for annotation was a lengthy and iterative process. First, there were questions about the appropriate level of generality and specificity of the categories. For example, we had discussions about whether and in what ways to differentiate between slaves and fugitive slaves, or whether to distinguish between free persons of color who were born free and those who were born into slavery but eventually manumitted (former slaves).

Temporality proved to be another difficulty; a person may belong to one category at one point in a narrative, and another at a different point. The transition from slave to free, of course, is a common

scenario in many of the narratives, often spanning multiple categories (particularly for fugitive slaves).

In developing a classification schema that would result in useful visualizations specific to this collection, we wanted to consider power dynamics within society. The classification of slaveholders' wives and children, as well as the enslaved persons' children, was critical to our efforts, but particularly thorny. Each narrative seemed to present a different case. Some slaveholders' wives were entirely absent from the narrative, and some children were too young to be implicated within the system. But others grew during the course of the narrative to become equally cruel, or worse, than the masters, whose label we deployed to signify absolute power and all the attendant abuses of human rights that plantation history reveals. Furthermore, some sons who inherited property and people became masters, while other sons did not. And what about wives and daughters, who haunted the backdrop of some narratives? What role did feminine forms of authority play in maintaining the plantation culture, and how could that be visualized in these narratives? Finally, should the children of enslaved persons be universally coded as "slaves" even if freed as adults?

Ultimately, we decided that the children of slaves would be classified as slaves. This was consistent with the legal system, as well as structural relationships, of the time. Male children of slaveholders were classified as slaveholders, a decision that was made both because they would generally inherit slaves and because their position was consistent with future power, according to social customs. Female children of slaveholders proved to be the most problematic. As children, young women lacked formal domestic power (in contrast to adult mistresses, a label we use purposefully to connote particular type of household power). But some of the family's slaves might be included in white women's dowry upon marriage, suggesting informal slaveholding or potential future ownership. In the end, we decided to classify them as slaveholders rather than as mistresses.

Two final issues arose in the development of our annotation schema. It was occasionally not possible to classify an individual

because there was not enough information, e.g., it might not be clear whether a white person was a slaveholder or not; and some narratives had social tapestries which were fundamentally different from the social role classifications we annotated. In the case of Henry Ossian Flipper's narrative, most of the characters could not be classified with antebellum social roles related to slavery because the narrative was written well after Emancipation and the military setting of his narrative varied greatly from other social settings.

3.2 Considering the interface design

Engagement with this project provoked a great deal of thought about how the data representation and interface could potentially affect users' interactions with, and interpretations of, the narratives.

3.2.1 Supporting interpretive work

At the outset, we explained that the interface is a way for readers to explore the text through different visual lenses. One might hope that the interface would encourage readers to consider, reflect upon, and question what they see, and thus use the interface as a vehicle for a deeper and more multi-dimensional engagement. But as digital humanists adopt visualization techniques from other disciplines, there is a danger of depicting data as observer-independent and certain, rather than observer co-dependent and interpretative [10]. Our interface cannot escape this danger. A reader might view the Tapestries as snapshots – a quick way to understand texts at a surface level, without questioning whether alternative representations exist, and without questioning the human decisions that helped create these visualizations. Thus, as Drucker argues, it is important to design applications that portray data "as *capta*, taken and constructed" [10].

To help users understand the schema of social roles, we provide a description of the categories, but additional explanation or visual cues might enable users to more easily recognize the constructed nature of categories, areas of ambiguity, and potential for divergent interpretations. Moreover, we might enable users to create their own role categories and/or modify existing categories. Making the basis of data representations visible and allowing users to shape representations can support engagement with social issues [11].

3.2.2 Representing social context

We began this paper with the observation that a common goal of visual analytics is to derive insight from large volumes of data. This may involve the effort to reduce cognitive overload by reducing the amount of data to be viewed [12]. In the support of humanistic inquiry, this practice can be particularly challenging in the analysis of complex and dynamic social systems, events, and contexts.

The interface presented here is an example of an effort to design a system to facilitate analysis of artifacts that are products of a complex and dynamic social context and history. This system facilitates reflection and questioning about how people, places and affect might be interconnected.

There are also ways to further support a sensitive and nuanced analysis of the corpus. For example, a given individual might simultaneously or at different times belong to different social categories; a system that visually supports this representation could help users to explore the ambiguities in the social roles of the period.

Additionally, the interface currently employs a simplistic measure to represent the presence of persons at different points in the narrative – frequency of mentions. Though frequency as a proxy for salience can be a useful measure, a richer representation, perhaps combining the person mentions with their affective contexts, could help the reader to engage further. Similarly, an absence or a lack might have significance. In her narrative, *Behind the Scenes; or, Thirty Years a Slave, and Four Years in the White House* (1868), Elizabeth Keckley provides name after name of the people that she encounters, and as such, the overt omission of her rapist with only – "I do not wish to say his name" – is telling. How should such an omission be visually represented, if at all? The decisions that we make concerning the data representation, visible or not, affect the questions that we and our users are later able to ask of it.

4 CONCLUSION

In summary, this paper outlined an effort to build an interface for exploring a subset of a historical corpus of slave narratives. We first described the interface and the method through which the interface was developed. Then, we addressed the challenges that we encountered, which led to multiple insights for the provision of visual tools for supporting interpretive work, including communication of the constructed nature of data representations, providing user control over and awareness of the representations, and facilitating the representation of social context.

REFERENCES

[1] D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon, *Visual analytics: Definition, process, and challenges*. Springer, 2008.

[2] T. E. Clement, "'A thing not beginning and not ending': using digital tools to distant-read Gertrude Stein's The Making of Americans," *Lit. Linguist. Comput.*, vol. 23, no. 3, pp. 361–381, Sep. 2008.

[3] A. Muralidharan and M. Hearst, "Wordseer: Exploring language use in literary text," in *Fifth Workshop on Human-Computer Interaction and Information Retrieval*, 2011.

[4] R. Vuillemot, T. Clement, C. Plaisant, and A. Kumar, "What's being said near 'Martha'? Exploring name entities in literary text collections," in *Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on*, 2009, pp. 107–114.

[5] C. Plaisant, J. Rose, B. Yu, L. Auvil, M. G. Kirschenbaum, M. N. Smith, T. Clement, and G. Lord, "Exploring erotics in Emily Dickinson's correspondence with text mining and visual interfaces," in *Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries*, 2006, pp. 141–150.

[6] C. Strapparava and A. Valitutti, "WordNet Affect: an Affective Extension of WordNet.," in *LREC*, 2004, vol. 4, pp. 1083–1086.

[7] A. T. Chen and P. Lach, "Needs and expectations in a tool for exploring slave narratives: a survey of prospective users," *Proc. Am. Soc. Inf. Sci. Technol.*, vol. 51, no. 1, pp. 1–4, 2014.

[8] D. A. Keim and D. Oelke, "Literature fingerprinting: A new method for visual literary analysis," in *Visual Analytics Science and Technology, 2007. VAST 2007. IEEE Symposium on*, 2007, pp. 115–122.

[9] J. R. Finkel, T. Grenager, and C. Manning, "Incorporating non-local information into information extraction systems by gibbs sampling," in *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, 2005, pp. 363–370.

[10] J. Drucker, "Humanities Approaches to Graphical Display," *Digit. Humanit. Q.*, vol. 5, no. 1, 2011.

[11] M. Dörk, P. Feng, C. Collins, and S. Carpendale, "Critical InfoVis: exploring the politics of visualization," in *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, 2013, pp. 2189–2198.

[12] G. Robertson, D. Ebert, S. Eick, D. Keim, and K. Joy, "Scale and complexity in visual analytics," *Inf. Vis.*, vol. 8, no. 4, pp. 247–253, 2009.