A Beginner's Guide to Using Voyant for Digital Theme Analysis

Shawna Ross, Texas A&M University (shawnaross1@gmail.com)
Randa El Khatib, University of Toronto Scarborough (elkhatib.randa@gmail.com)

Abstract

"Digital Theme Analysis" depicts the role that thematic analysis plays in literary criticism, places traditional thematic analysis approaches alongside digital ones, and offers best practices for carrying out digital thematic analysis in the context of Voyant Tools. The chapter identifies thematic analysis as a meaningful pattern that can be traced throughout text(s) that may form the foundation for critical interpretation. Authors draw continuities between traditional theme analysis that revolves around close reading and digital theme analysis that can be carried out using a variety of automated digital methodologies, and indicate how the latter can help accelerate the process while ensuring that the significance of themes are accurately estimated through an empirical, machined approach. Two applications of digital theme analysis are illustrated using Voyant Tools, a popular open-source tool that combines a variety of simple text-mining tools in a single, intuitive graphical user interface. The first case study is an ecological reading of Charlotte Brontë's Jane Eyre, and the second an exploratory thematic analysis of Mary Shelley's Frankenstein, through which best practices for carrying out digital thematic analysis are proposed. Together, the chapter demonstrates how digital theme analysis can serve as a point of entry for an almost on-demand approach to digital theme analysis—providing that the researcher understands the applications of the tool and how to prepare and interpret the data, which they can incorporate alongside close reading to create a sophisticated, fully fleshed out literary argument that takes advantage of the unprecedented speed and scope of distant reading.

A Note on the Text

This article was originally drafted for inclusion in a larger project edited by James O'Sullivan, whom we would like to thank for his editorial labors. Much time has passed since we first drafted this article, but we do hope that it is useful for scholars and teachers who are beginning to experiment with Voyant (https://voyant-tools.org/) for computer-assisted literary analysis.

Introduction

The Digital Humanities, with its wide range of software-assisted approaches for literary analysis, has ushered in a new golden age for what is possibly literary criticism's least sexy technique: the analysis of theme. From bird's-eye view assessments of the novel's shifting preoccupations over time (such as Ted Underwood in *Distant Horizons: Digital Evidence and Literary Change*) to reassessments of particular themes that appear to dominate individual literary movements (Laura McGrath, Devin Higgins, and Arend Hintze's "Measuring Modernist Novelty") to thematic analysis of a single author (Enrique Mallen and Luis Meneses's "Adjoined Conceptual Domains in the Bilingual Poetry of Pablo Picasso"), machine-assisted literary critical methods have made possible a stunning number of new hypotheses about literary themes. This renewal re-energizes what may have seemed a quarter-century ago to be the stuff of secondary education, from pride in Greek tragedy to race in Morrison's novels. Indeed, the work of finding and analyzing themes might seem quaint, trivial, or simple enough not to deserve a second glance. Yet theme analysis is the invisible bedrock upon which we build the superstructures of far flashier approaches—and this is exactly why we should take the time to reevaluate theme. By acknowledging when we rely upon theme to make our argument, by weighing what powers and risks theme analysis entails, and by understanding how Digital Humanities methods resemble or deviate from traditional literary criticism, digital humanists can better appreciate the continuities between their practices and aims and those of traditional literary scholars. Doing so will ensure that digital literary criticism inherits important lessons from non- or pre-digital literary criticism. These lessons are both methodological (involving analytic practices that combine well with digital tools) and, as we hope to show in this essay, ideological (involving an awareness of thematic distortions or elisions). By exploring theme analysis carefully—especially in defining it as a methodology shared by analog and digital criticism—we hope not only to help critics seamlessly combine digital and non-digital approaches to literature, but also to remind us of the oft-overlooked power and flexibility of theme analysis.

What is Digital Theme Analysis?

In theme analysis, a literary critic identifies a meaningful pattern in a single text or across multiple texts. As M. H. Abrams elaborates in his landmark *Glossary of Literary Terms*, "Theme is sometimes used interchangeably with 'motif,' but the term is more usefully applied to a general concept or doctrine, whether implicit or asserted, which an imaginative work is designed to incorporate and make persuasive to the reader" (170). Traditionally, a critic identifies themes by reading carefully, assembling various clues that suggest how a certain topic organizes or underpins a text. These clues might include epigrams, allusions, sources of conflict between characters, key words in the narration or dialogue, prominent symbols, or issues repeated from the author's other works or culled

from the text's historical contexts. This theme—this meaningful pattern—is then traced throughout the text(s) to form the foundation for a critical interpretation. Instances of the theme are collected, collated, and supplemented with research. The critic then uses inductive reasoning to generate a thesis about the theme's significance for the text as a whole. By documenting a pattern and analyzing it as a theme, what might have started as an isolated observation ("There sure are a lot of conversations about pride in this text!") is transmuted into a sophisticated argument ("This text perceives and critiques a major shift in the role played by public demonstrations of personal pride within ancient Greek constructions of political authority"). In sum, theme analysis is a critical approach in which an observed pattern is documented and explored in order to reveal its—or perhaps more accurately, to *imbue* it with—explanatory power.

Even the most sophisticated literary criticism engages in theme analysis at some point in the research process. In its most powerful form, theme analysis is a flexible approach that combines well with any number of other tools—particularly those represented in the other chapters in this collection, particularly "Topic Models" (because topic modeling is one powerful tool that a critic can use to support theme analysis), and "Computational Genre Analysis" (because theme can be explored as one of the constitutive elements comprising a genre). The critic might use any number of Digital Humanities tools to do so, as long as the end point is to analyze theme. This is why we might read the history of literary criticism by classifying various approaches through their treatment of theme: which kinds of themes merit critical attention, which signature technique they use to locate themes, which final interpretive mode that the chosen theme is presumed to access (ethics, aesthetics, history, philosophy, culture). For example, New Criticism dwelled on themes highlighted by the text's stylistic properties in order to craft a unified reading of the text; form becomes the bearer of the theme, telling the reader how to resolve ambiguity or dissonance in the text's thematic registers. Political approaches, such as Marxism, feminism, queer studies, or critical race studies, announce their thematic preoccupations in their very names. Furthermore, texts under their various microscopes are often chosen precisely because of their thematic appropriateness, and arguments are often judged by the degree to which they can showcase, refine, and reflect upon each school's characteristic constellation of related themes.

In other words, thematic analyses of literature have always depended on some form of filtering that efficiently processes texts. Aristotle's formalist machine of genre filtering and Derrida's anti-formalist software of deconstruction control variables, iterate rudimentary processes, and creatively deform texts as much as do Digital Humanities techniques. While we certainly do not equate analog with digital criticism—the timeline, the scope or unit of analysis, the eventual output, and the labor structures involved can differ dramatically—theme analysis is an activity in which both methods equally participate. Digital Humanities methods introducing automation are worth pursuing because analog criticism purchases the power of theme analysis at a high price, as the decision to scan a text for a chosen trend may endow the text with a unity that may be spurious.

Automation—allowing software to identity themes—reduces the probability that a critic will overstate a theme's significance. In the case of politically inflected schools, Digital Humanities methods can be a powerful ally against skeptical readers.

Theme analysis falls into the realm of digital when computational tools are employed to extract recurring patterns from texts in an automated manner. Whereas traditional theme analysis revolves around close reading, digital theme analysis substitutes—during at least one stage of analysis—a "human" way of reading with an empirical, machined approach. Beyond the apparent objectivity that automation lends to textual criticism, another major strength of digital theme analysis is that the critic can use it as much or as little as desired. It can predominate a project, providing the critic's entire methodology and topic of analysis, or it can play a supporting role. For example, as a useful approach to completely unfamiliar texts, digital theme analysis can aid in the earliest stage of analysis, namely, in the basic selection of the works by theme, which can then be used in a more sophisticated research project. Alternatively, a critic who has already begun analog theme analysis might turn to digital methods to clarify a theme's relationship to other themes or discover moments in the text when the theme proliferates or attenuates. Or a critic might use digital methods after drafting an article in order to create corroborating charts and figures. All of these examples share a common practice: the critic at some point relies on computational tools that distant-read texts to locate, document, or analyze theme in a literary text or corpus.

Case Studies Using Voyant Tools

Theme analysis can make use of nearly any software or technique described in this volume. N-grams, topic models, or sentiment analysis may be used to automate the discovery of themes across tens, hundreds, or tens of thousands of texts, while databases may be employed to organize and document particular instances of a theme into a new database or to uncover the thematic patterns that structure an existing database. In some contexts, TEI-XML may be used to encode particular themes in a digital edition, while in other contexts, visualizations produced through network analysis might be more appropriate for demonstrating the presence of a theme in a novel. To begin elucidating interpretive moves a critic might make to engage in digital theme analysis, let us turn to one popular, easy-to-approach tool: Voyant Tools. For the remainder of the chapter, we will focus on Voyant as a means of theme analysis. Although this does mean we have shaped our examples to align with Voyant's particular affordances, surveying many software packages for literary theme analysis would be redundant in the context of the present volume. Moreover, it would lead to an unfortunate predominance of (necessarily ephemeral) instruction sets over our primary mission, which is to define and model a digital methodology for literary theme analysis that can be easily adapted to any software package.

The importance of Voyant to the Digital Humanities cannot be understated. While advanced scholars in the field prefer to use more targeted software packages, construct databases instead of submitting their textual data into a browser-based tool, and/or to develop their own methods of distant reading, Voyant Tools is a popular tool for the classroom, for the general public, and for entry-level DH students and scholars. (And some project developers even run Voyant locally to bypass the website altogether, neatly obviating this drawback.) Voyant was developed by Stéfan Sinclair (McGill) and Geoffrey Rockwell (University of Alberta), whose book *Hermeneutica: Computer-Assisted Interpretation in the Humanities* documents and reflects upon their creation of Voyant. Since Sinclair's death in 2020, the DH community has been reflecting on his importance to the field as a collegial and brilliant teacher, scholar, editor of *Digital Humanities Quarterly*, and president of the *Association of the Computers and the Humanities*. Voyant continues to be in development now, with a renewed appreciation for Sinclair and Rockwell's approachable, influential tool.

Voyant Tools is a free, open-source application that works in-browser. It combines a variety of simple text-mining tools in a single, intuitive graphical user interface, divided into individual windows or panes, which the user may adjust at will. Users may change the relative size of the individual panes, their positioning on the screen, and even the tools displayed on the individual panes. What is extremely useful is that some tools dynamically respond to one another, altering their displays to stay current with the particular term or textual phenomenon that the user is exploring on another pane. In addition, users can customize key features related to each tool, such as by defining stopwords (words left out of tool calculations, such as common articles, pronouns, and prepositions, because they are not relevant for analysis). At the time of publication, the default view presents a word cloud on the top left (the "Cirrus" tool). "Cirrus" may be easily switched, with a single click, to "Terms," which is a simple concordance, or "Links," which generates a simple network graph that shows the relationships between the most frequently used words in the text or texts the user has uploaded. In the top middle is a simple copy of the text(s) the user submitted (the "Reader," which will automatically navigate to relevant passages as the user engages with other panes); users can click quickly to change this pane to "TermsBerry," a kind of network visualization that depicts relations between common terms by representing common words as a berry-shaped collection of circles that become colorized when the user mouses over the term or a related one. At the top right is "Trends," a time-series graph that displays in the form of a familiar X-Y scatter plot to reveal how the frequency of a term changes over the course of the text; it may be replaced with another single, simple click with "Document Terms," a searchable concordance that also displays a thumbnail representation of the "Trends" data for the relevant term. On the bottom left is the "Summary" pane, which reports the total number of words, the number of unique word forms, the vocabulary density (a measure of the variety in the author's diction, generated by dividing the total number of words by the number of unique word forms), the average number of words per sentence, and the top five most frequent words. It

may be quickly replaced with "Documents," which allows users to sort through multiple texts if they are working with a large corpus (that is, more than one text), or "Phrases," a tool that, like topic modeling, identifies multiple groups of commonly co-occurring words. On the bottom right, the "Contexts" pane displays the words that immediately precede and proceed from the terms that the user is currently investigating on other panes; this means that users need not flip through the text on another browser tab or a physical book. "Contexts" may be replaced with "Bubblelines," which uses proportionally sized circles that correlate to word frequency to create lines of varying thickness (it therefore visualizes the same kind of information as the "Trends" pane, but in a more colorful, graphics-based format), or "Correlations," an advanced analytical tool that locates similar patterns of co-occurring words (that is, groups of words that gain or lose frequency roughly alongside one another). Many more tools beyond these reliable favorites are available on Voyant, including the new Veliza chatbot. Based on Joseph Weizenbaum's classic natural language processing program ELIZA, trained to respond to textual inputs as a psychotherapist would, Veliza will generate responses to particular lines in your text, revealing how a psychiatrist might react to the characters' thoughts and problems.

Except for Veliza, most of Voyant's constituent text-mining tools use word-frequency tabulations to visualize a text—that is, they count the most frequent words (MFW) used in text and generate some sort of chart or graphic, often interactive, that represents this frequency (most familiarly, the word cloud). To begin, the user must submit a textbase by copy-and-pasting an electronic text(s), by providing a URL(s) to a website containing the text(s) in a format that Voyant recognizes (for URLs, PDFs work only on a page-by-page basis), or uploading the text(s) in a variety of formats (.doc, .rtf, .pdf, .html, .xml, or a zipped file containing files of any of these formats). Voyant allows users to export their results, making it unnecessary to reload the text each time the critic wishes to view the visualizations. This convenience, it must be noted, *does* make it incumbent on the critic to have on hand software that can parse the resulting file, to save and process images with enough detail to render clearly when viewed through different platforms, and to record any relevant metadata (e.g., the edition accessed, the date when the visualization was created, the original format and size of the downloaded graphics or dataset). Though a critic may feel tempted to use a tool like Voyant simply to confirm hypotheses and immediately turn back to drafting a written argument, the data generated depend on the critic's each and every input and interaction with the results. Consequently, visualizations therefore may not be replicable, so always save results for future consultation.

To put it another way, just as with any tool used for digital theme analysis, each of Voyant's affordances requires specific actions on the critic's part so that it does not become a liability. For example, librarian Megan Welsh has critiqued it for a number of difficulties, including its uneven loading times and the lack of standardization preventing easy exportation (96-97). Yet successful critics will anticipate such challenges and adjust their experimental design to reflect

their needs and available resources, including text file format and reliability, relevance of producing "real-time" results (in pedagogical contexts), mode of dissemination (online versus hardcopy), and type of data required (statistics, dataset, or image). Knowing the limits of any digital tool used for textual criticism is crucial for the early stages of digital theme analysis. In addition, knowing how the tool works—what algorithms or processes it runs, or in other words, what kind of data it tracks and what kind of data it produces—will manage expectations about what a tool can reliably do for the critic. It will also, perhaps even more importantly, avoid a "black box" style of scholarship that blindly accepts the outputs of a tool without properly acknowledging the tool's limitations. Regarding the present example, Voyant, the scholar should understand that word frequency lists are very good at suggesting the text's linguistic preoccupations, but less good at indicating word proximity, discerning between different denotations or connotations of the same word, or relating those preoccupations automatically to the text's plot or structure, to its historical context, or to word usages in comparison texts. This does not mean to avoid Voyant, but to supplement its results with other critical methods.

Our first case study shows how Voyant can be used to conduct a study of a text with a specific lens (theory) in mind. Let's say that a scholar is interested in an ecological reading of Charlotte Brontë's *Jane Eyre* and has already completed an initial round of analog close-reading. When submitted Voyant, with irrelevant data deleted (a preface not in the first edition, licensing information from the text available on Gutenberg) and stop words selected, the top 50 most frequent words apparently deal with interior themes—"house" (168 instances) and "door" (155)—rather than ecological themes. However, consulting the "Word Trends" box, which produces a line graph for individual word frequencies over the course of the book, yielded an interesting pattern in which "house" and "door" rise and fall rhythmically, with three distinct troughs that indicate that certain portions of the book that might deal more with the outdoors—suggesting not only that the novel fluctuates in privileging the outdoors, but also that the scholar should look in those "troughs" for relevant passages to close-read. Before devoting time to closely examining these passages, it is necessary to filter out irrelevant passages by considering the multiple meanings of such superficially simple words. Does Brontë use "door," a liminal object located on the boundaries between indoor and outdoor spaces, more frequently to indicate entering or leaving a house? Does she use the term figuratively or literally? Does she focus in interior or exterior doors? Voyant's "Keywords in Context" window (which reveals the handful of words both before and after the instance of the keyword in the critic is interested) can help answer these questions. One of the most powerful functionalities in Voyant, this window allows the critic to combine distant and close reading easily and quickly. Consulting this window reveals that Jane is often being blocked from a space to which she desires entrance. Thematically, doors are related to violence and alienation for Jane—a negatively connoted word not associated with freedom or happiness but with feelings of isolation and exile.

So far, if the critic wishes to make an ecological reading of the novel, Voyant suggests that it is not possible to support such a reading by establishing Jane's unequivocally positive relationship to the outdoors. Connecting Voyant's results to salient passages in the text shows that it is probably not possible to champion Jane as an environmentalist. Reflecting on the novel's major transitional moments, the critic then locates passages in the novel regarding this proposition: the book opens with Jane being pleased that "there was no possibility of a walk that day" (Brontë 2006, 9), and Brontë underlines the horror of Jane's exile from Rochester by showing Jane weak and hungry from her few homeless nights spent on the moors. Further down the word frequency list are "home" and "nature" in close proximity, yet inspecting the sentences in which such words are found, "nature" appears to denote temperament, personality, type, or a "state of nature" rather than to animals, plants, or the outdoors. It is only further in the word frequency list—"wild" (58), "air" (56), "wind" (47), "moon" (42), "sky" (39), "wood" (37), "trees" (34), "flowers" (30), "garden" (28)—that a sense emerges of *Jane Eyre*'s focus on nature as an elemental force and as a subject for artistic representation. At this point, only after carefully sifting through Voyant's word frequency visualizers, a mature theme analysis of nature in *Jane Eyre*—one that eschews the politically inflected idealism that analog theme analysis might seem to support—begins to emerge. To pursue Jane's ecology of aesthetics, the scholar could then use Voyant to search for "window" and "painting," as both focus on natural subjects as something to perceive, as well as their derivatives ("window-seat," "bow-window," "paint," "draw") and a word relevant to both: "frame." This investigation yields a constellation of significant passages to analyze, which the critic will then connect to other critical readings of the text, and it also yields ideas for further digital analysis, such as topic modeling (to check whether natural imagery and artistic/window imagery cluster), statistical analysis (to verify the statistical significance of these word counts), or n-grams (to compare *Jane Eyre* to other texts). Significantly, these words were all suggested not directly from these Voyant results, but rather from dipping in and out of the text, reading passages indicated by Voyant, emphasizing the degree to which closely interacting with the text and existing literary criticism is still necessary to transform the quantitative data into mature theme analysis.

Our second case study illustrates how one may approach a text even if one has not engaged in any systematic close reading of it. Consider Mary Shelley's novel *Frankenstein*, which has been adopted by modern culture through countless remakes. This novel is colloquially associated with the story of the "monster," the "ungodly" evil creature that turns against his creator, Victor Frankenstein. We would therefore expect to find words pertaining to monstrosity, science, horror, and revenge. To test these (ostensibly) predominant themes, the critic might run the text of the novel through Voyant. What immediately stands out are the numerous repetitions of words relating to sentiments, which significantly outnumber those related to horror, science, or alchemy. In fact, the predominant words, when viewed in the "Keywords in Context" window, portray a humanoid with deep emotions (or at least a strong desire for them) rather than a violent

drone. The following MWF capture these emotions: “felt” (79), “feelings” (76), “heart” (76), “dear” (72), “love” (59), “feel” (50), “hope” (50), “happiness” (49), “happy” (46), “joy” (41), “affection” (40), “good” (37), “pleasure” (37), “soul” (36), “spirit” (34)“gentle” (34), and “kind” (34). Solely considering the plot of Shelley’s *Frankenstein* yields a story of murder and loss, yet the word counts stressing this theme are, in fact, relatively minor.

The critic’s next step is to consult extant criticism. One representative example is Maurice Hindle’s “Vital Matters: Mary Shelley’s Frankenstein and Romantic Science,” which argues, “Few I think would fail to use the word ‘scientist’ to describe the monster’s creator, whether they had read Mary Shelley’s novel or not. Yet if one turns to the text of the book, this word is nowhere to be found [...]. The fact is, the word ‘scientist’ had not even been coined in 1818” (Hindle 1990, 29). This is not a fluke but rather the first step in recognizing that we cannot import an ahistorical reification of science as we know it today. A Voyant user could advance Hindle’s argument by identifying a different theme for analysis—one that *is* justified by digital theme analysis. For example, the two most frequently used words are “man” (131) and “life” (115), while the words “creature” (66) and “monster” (32) are evoked far less frequently than “man” (131) or”human” (71). At the same time, it quickly becomes apparent that nature, through human sensory perception, is one of the creature’s primary sources of acquiring knowledge: “country” (54), “nature” (53), “sun” (45), “world” (45), “scene” (44), “ice” (42), “light” (37), “mountains” (37), and “earth” (36) — an empirical approach, to be sure, but one that lacks a sci-fi element associated with the genre when tracing the creature’s origins and the language used to describe his worldview. Rather, one of the MFWs used by the creature when contemplating his origin is “father” (113), indicating Victor’s role as the absent “father” from his life. This is closely followed by the word “eyes,” (102) where by looking in the “Keywords in Context” window, the critic finds that it is the primary lexical connection between creator and creation. This is even more poignant because the creature’s inability to engage with others renders sight the primary basis of his social life. Looking at Voyant’s density graph, which traces the frequency of abstracted word repetitions across a text and allows the critic to experiment with chunking the text (divide it into sections for easier interpretation), it becomes evident that many similarities link Victor and his creature. Strikingly, many quotes are almost identical, such as one of the creature’s closing sentences: “I cannot believe that I am the same creature whose thoughts were once filled with sublime and transcendent visions of the beauty and the majesty of goodness” (Shelley 1992, 200). Victor indulges in similar contemplations because he blames himself for creating a “monster,” emphasizing how the two defend the same humanistic values.

This digital theme analysis shows that Shelley’s *Frankenstein* does not prioritize the language of horror and science, but a language of humanism. Further, it does not suggest that *Frankenstein* is a horror story in the modern sense; in fact, there is a noted lack of words related to violence. The horror is located in the monster’s appeal to our emotions and logic and in the unexpected closeness of

his state of mind and intentions to Victor's, which makes the critic question who the monster really is—demonstrating how digital theme analysis moves quickly between quantitative data and the larger questions of traditional criticism.

Sample Methodology

This *Frankenstein* case study demonstrated how a critic can incorporate a simple tool effectively yet relatively quickly. And indeed, straightforward applications like Voyant do reward an almost on-demand approach to digital theme analysis—providing that the researcher understands the limitations of the tool, uses an accurate dataset, responsibly interprets the data, and documents the process. However, if the critic uses more arcane or specialized tools, or if digital methods will form the primary mode of analysis for an extended piece of criticism, the critic will be rewarded by using a more formalized methodology. Although the following list is not exhaustive, the scholar should proceed like so:

1. Define the research question and set the parameters for an acceptable answer (scope, granularity, accuracy, format).
2. Review available digital tools and select the most useful and feasible one.
3. Prepare the data, making sure to "clean" it (check for accuracy and minimize noise) and render it into the right format.
4. Submit the data to the tool for distant reading. (If using a large corpus, the critic may want to run a sample set before preparing the entire dataset.)
5. Identify any flaws in the results from a technical standpoint (glitches, dataset errors). Fix them and rerun the data. If this cannot be fixed, document the process for later incorporation in the resulting study.
6. Save relevant proof (statistics, datasets, images), along with metadata, in a durable, functionally interoperable format. In other words, avoid proprietary formats, and eschew compression. For unstructured textual data, plain text format (.txt) is preferable. For tabular data (that is, structured textual or numerical data), comma-separated values format (.csv) is ideal. For images (including maps, line drawings, and photographs), default to TIFF (.tiff) or scalable vector graphics (.svg).
7. Run a "gut check" by checking results with the text or text corpus and/or with results from other digital tools. Rerun the software or rethink the distant reading method if it seems advisable.

There are many considerations that should guide Step 2, the process of reviewing and selecting among a range of potential software packages or platforms to use for theme analysis. These mandatory conditions that an appropriate tool must satisfy include the following requirements:

1. It must satisfy the research question.
2. It must use technologies and formats the scholar is familiar with or is willing to become trained in.
3. It must be able to handle the dataset (both its size and format).

4. It must allow the critic to access, adjust, and save the data as many times as needed.
5. It cannot require financial, material, spatial, or personnel resources that are impossible to access
6. It cannot suffer from limitations (whether from glitches or from the tool's inherent limitations) that would invalidate the conclusions in the mind of a critical peer reviewer.

At this point, the critic's work has truly just begun, as the data must be incorporated into a cohesive argument. First, the scholar should write an account of the tool usage. Its length, use of technical terms, and position (inline with your analysis or included as a footnote, appendix, or companion digital content) should match the audience's expertise and interest in technological niceties. Without delving into an unnecessary amount of detail, the writer should disclose any difficulties incurred or any gaps in the data. Once enough detail about the technological methodology has been provided, in order to retain the reader's trust and interest, the scholar should then incorporate results as seamlessly as possible into an argument about literary theme. Pair the digital tool results with passages from the literary text, allowing the scholar to provide a close reading that will convince readers that the theme analysis is justified by analog and digital methods. The scholar might also supplement these statistics or figures by incorporating data from another digital tool or comparing these results with existing literary criticism on the particular theme in question.

Conclusion

Digital literary studies might be understood partly as an approach that uses computational methods to try to locate, explore, and substantiate themes with computationally enforced objectivity, with greater efficiency, or with a significantly expanded scope than a critic using only analog methods. These digital methods should not be seen as the critic's sole or most important work, but as a point of entry at which the critic steps in actively to analyze the results, using close reading to create a sophisticated, fully fleshed out literary argument that takes advantage of the unprecedented speed and scope of distant reading. Being armed with these tools allows us to expand our literary territories by plunging into larger battles without fear of atrocity—that is, without a total loss of the text itself. We can ask bigger questions with less time.

References

Abrams, M. H. *A Glossary of Literary Terms*. 7th ed. Heinle & Heinle, 1999.

Brontë, Charlotte. *Jane Eyre*. Edited by Stevie Davies. London: Penguin, 2006.

Hindle, Maurice. "Vital Matters: Mary Shelley's Frankenstein and Romantic Science." *Critical Survey* 2.1 (1990): 29-35.

Mallen, Enrique, and Luis Meneses. "Adjoined Conceptual Domains in the Bilingual Poetry of Pablo Picasso." *Digital Studies/Le Champ Numérique*, vol. 9, no. 1, Dec. 2019, p. 20. *www.digitalstudies.org*, doi:10.16995/dscn.320.

McGrath, Laura, Devin Higgins, and Arend Hintze. "Measuring Modernist Novelty." *Journal of Cultural Analytics*, 2018. doi:10.22148/16.027.

Shelley, Mary. *Frankenstein*. Edited by Keith Carabine. London: Wordsworth Editions, 1992.

Sinclair, Stéfan and Geoffrey Rockwell. *Hermeneutica: Computer-Assisted Interpretation in the Humanities.* MIT Press, 2016.

Sinclair, Stéfan, Geoffrey Rockwell, and the Voyant Tools Team. *Voyant Tools*. 2012.

Underwood, Ted. *Distant Horizons: Digital Evidence and Literary Change*. University of Chicago Press, 2019.

Welsh, Megan. "Review of Voyant Tools." *Collaborative Librarianship* 6.2 (2014): 96-7.