

Digital *Frankenstein*

Proposal Author: Giorgina Paiella

Submitted for Professor Alan Liu's Intro to DH Project Prospectus Assignment

March 14 2017

Table of Contents

I. Abstract.....	2
II. Narrative.....	2-6
III. Environmental Scan.....	6-7
IV. Methods and Technology.....	7-11
V. Work Plan and Schedule.....	12
VI. Project Budget.....	13
VII. Method of Project Evaluation.....	14
VIII. Final Product and Dissemination.....	14
IX. Sustainability.....	14

I. Abstract

I am seeking a grant of \$26,179.20 for the creation of a website, *Digital Frankenstein*, dedicated to a comprehensive, digital humanities exploration of Mary Shelley's *Frankenstein*. As we approach the 200th anniversary of *Frankenstein*'s publication, we can reflect upon the far-reaching cultural legacy of Shelley's narrative in literary studies, cinematic history, and popular culture. Despite its popularity, however, *Frankenstein* has not been the subject of extended DH analysis. In light of the bicentennial of the text's publication, this project will use DH tools and methodologies like text mining and visualizations—coupled with written analyses exploring the outcomes of these experiments run on the text—to reveal new insights into Shelley's novel and hidden histories that would otherwise be overlooked by traditional methods of literary analysis.

II. Narrative

Frankenstein has inspired many succeeding narratives and continues to produce a prolific amount of literary scholarship since its first publication in 1818. DH work is valuable because it reveals previously overlooked features of a text, using machine readings and visualizations to uncover new insights.¹ This innovative project, called *Digital Frankenstein*, would be the first comprehensive DH exploration of *Frankenstein*, resulting in a website featuring text analysis and visualization experiments run on the novel and complementary explications of the outcomes of each investigation. While DH tools and methodologies can be applied to any text or corpus, *Frankenstein* is particularly well suited to DH methodologies for a number of reasons.

In addition to being one of the most prevailing cultural narratives, *Frankenstein* is also one of the most frequently misread, misinterpreted, and misunderstood texts in the

¹As Franco Moretti states of visualizations in *Graphs, Maps, Trees: Abstract Models for Literary History*, “these maps will be *more than the sum of their parts*: they will possess ‘emerging’ qualities, which were not visible at the lower level” (53).

history of the canon. The abundant political, scientific, and gendered contexts that Shelley wove into the novel—and the multiplicity of entry points and possible readings that emerge as a result—contribute to the text’s mythic status. Also contributing to this status are the many literary and cinematic adaptations and remediations of Shelley’s novel. As literary scholar Chris Baldick contends, the “series of adaptations, allusions, accretions, analogues, parodies, and plain misreadings which follows upon Mary Shelley’s novel is not just a supplementary component of the myth; it *is* the myth.”² This project will accordingly explore the variety of intertexts (preceding, contemporaneous, and succeeding) with which Shelley’s text is in conversation. In addition to the corpus of *Frankenstein* derivative works, the novel itself was published in two editions: the original 1818 publication and the revised 1831 edition. Topic modeling and other text-analysis tools will allow for an in-depth look at the changes made to the text between the 1818 and 1831 published editions. Exploring these alterations on a semantic level will allow me to discover whether text mining confirms existing interpretations of the text’s revision or reveals new insights. Text mining will also be useful for a cross-textual analysis of topics in the two editions of *Frankenstein* as well as other valuable intertexts to the novel, like Mary Shelley’s journals and Milton’s *Paradise Lost*. As a text that has been persistently misread as a tale of technological prophecy, text analysis and visualizations will serve as a means of wading through the many scholarly readings and interpretations of *Frankenstein* to see how DH can illuminate what is unrevealed by literary readings of the text.

Frankenstein is also an ideal candidate for DH applications because it is a novel deeply invested in spatiality, geographic place, and movement. Visualizations would bring to life the epistolary structure of the novel and the relationship among the three nested narratives for students and scholars of the text. Below, I discuss the DH tools and methodologies that will structure my investigations and the project outcomes that they will produce. I detail the technological methodology behind these tools in the “Technology and Methods” section of this proposal. While I am breaking down these topics to detail the various facets of the *Digital Frankenstein* project, many of these

² Baldick, “The Monster Speaks: Mary Shelley’s Novel” (4).

methods complement each other, and the data generated from one investigation will often be integrated into another (for example, the topics generated from topic modeling will also be used to produce a social network analysis map visualizing the relationships between certain topics and the texts in which they appear).

Topic Modeling

I plan to run topic modeling with MALLET on the 1818 and 1831 editions of the text, Mary Shelley's journals, and intertexts like *Paradise Lost*. Topic modeling will allow for an exploration of the differences between the different editions of the text to see if one has "more" of a given topic than another. Results will then be analyzed to see how these outcomes coincide with or revise existing literary scholarly examinations of the revisions to the 1818 text. Topic modeling can be used to test a number of possible revision theses; it can track on a semantic level, for example, whether the 1831 edition hinges upon fatalism more than the 1818 edition and whether Victor is held less responsible for his actions.³ Other important areas of topic modeling exploration will include an examination of whether the 1818 edition has more topics relating to responsibility and Victor's personal failure to love his creature, and whether the 1831 edition has more topics relating to remorse, destiny, fate, or hubris. Another thesis to be tested is whether there are more topics relating to an organic conception of the universe in the 1818 edition and a mechanistic view of the universe in the 1831 edition. Evaluating shifts in topics pertaining to gender will also be valuable. Sentiment analysis of key changes to the novel will complement this topic modeling work.

GIS Mapping and Timelines

GIS mapping will be used to track the global character journeys in the narrative and visualize mirrored paths in the novel (for example, the intersection of Victor and the monster's paths). Mapping character movement can also be used to test a number of theses relating to mobility, including an examination of whether the mobility of female

³ As Anne K. Mellor argues in "Revising *Frankenstein*."

characters in *Frankenstein* is more restricted in contrast to their male counterparts. Visualization can reveal whether there are any patterns to the shape or structure of character journeys, and the challenge of mapping the nested structure of the narrative will be a unique contribution to the mapping of novels and allow for a renegotiation of the boundaries and limitations of visualization tools used to visually represent literature. Mapping will make the nested narrative structure easier to understand and more clearly represent how the novel chronologically unfolds. A timeline set against the backdrop of a geographic map of character paths will also aid this goal and allow for an interactive experience of the text. GIS mapping can also be used to evaluate if there is a geographic correlation between topics generated from topic modeling and certain locales in the novel.

Social Network Analysis

Producing a social network map of the novel would be an innovative application of social network analysis because of the nested, frame narrative structure of *Frankenstein*. The novel is narrated exclusively from one character's perspective through recounted events, so a social network map of the novel will offer some interesting challenges to this tool and allow me to explore how visualizations can better represent non-traditional narratives. Social network modeling will also be used to model the MALLET topics generated on the novel and related narratives in Gephi. The final application of social network modeling for *Digital Frankenstein* will be a social network map of *Frankenstein* scholarship and scholarly networks generated based on the prolific body of criticism written on the text. This social network map would reveal which texts and scholars have been in conversation and track the rising and falling trajectories of certain topics over time.

Voyant Tools Analysis

The variety of tools offered by Voyant will produce visualizations of the novel, including cirrus cloud word frequencies, knot diagrams, bubble diagrams, and network

maps. These tools can track changes across the two editions of the text and explore whether certain words are more common in one edition compared with the other.⁴ These visualizations would offer an important counterpart to written scholarship on the text.

III. Environmental Scan

Thus far, the only other DH work conducted on *Frankenstein* has consisted of digital archiving projects. The most high profile project that digitally preserves the text is the Shelley-Godwin Archive, which provides digital access to the manuscripts of Percy Shelley, Mary Shelley, William Godwin, and Mary Wollstonecraft, uniting the legacy of this prolific family of writers. The NEH funded project consists of a partnership between the Maryland Institute for Technology in the Humanities and The New York Public Library, with source contributions from the Huntington Library, Bodleian Library, the British Library, and the Houghton Library.⁵ The archive features Shelley's *Frankenstein* in various editions to track the evolution of the text's production and publication, including drafts of volumes one and two of the 1818 edition of the text and fair copies of volume three of the novel. *Digital Frankenstein* would be a great complement to this digital archival work because it would focus on an in-depth analysis of *Frankenstein* with DH tools, illuminating the potential of DH applied to the text aside from TEI projects.

The topic modeling facet of this project is inspired by DH work like Cameron Blevins' "Topic Modeling Martha Ballard's Diary," a text-mining project dedicated to text analysis of Ballard's extensive diaries and consisting of a blog post series discussing the project's progress.⁶ Blevins conducted MALLET topic modeling on Ballard's diaries, consisting of daily entries composed over the course of twenty-seven years. He generated topics on the diaries, labeling each topic and then using the programming language "R" to generate matrixes on each diary entry, topic, and presence of the topic in a given entry.

⁴ Word frequency counts produced through Voyant can also be supplemented by a corpus analysis using AntConc.

⁵ See the Shelley-Godwin Archive website: <http://shelleygodwinarchive.org>

⁶ <http://www.cameronblevins.org/posts/topic-modeling-martha-ballards-diary/>.

Blevins then produces a series of visualizations on the data, exploring, for example, how weather-related topics follow seasonal patterns and how peaks in certain topics correspond with events in Ballard's personal life. I will use a similar technique and blog post method to topic model Mary Shelley's journals along with *Frankenstein*, exploring how Shelley's journals can provide insights into the novel and vice versa.

For the visualization aspect of this project, it is worth mentioning that maps and timelines of *Frankenstein* are commonly used in the classroom for instruction purposes or have been produced using traditional visualization tools (like the map and timeline), but not in a DH context. This project would utilize cutting-edge DH tools that are interactive in nature and all featured on one platform for easy access and exploration. *Digital Frankenstein* will therefore bridge the gap between traditional visualization studies and DH.

IV. Methods and Technology

The various objectives of utilizing these tools to achieve project outcomes are discussed in the narrative portion of this proposal. This section discusses the technical methodologies behind these tools and how they will be used for the *Digital Frankenstein* project. The products of these text mining and visualization experiments—in addition to the blog posts analyzing them—will all be published on a free, open-source website and content-management system like [Wordpress.com](https://www.wordpress.com) for ease of use, open access, and affordability.

Topic Modeling

As a sample run, I ran topic modeling on the 1818 and 1831 editions of Shelley's *Frankenstein*. The first topic model was run using Topic Modeling Tool. As described earlier, the purpose of topic modeling *Frankenstein* is to reveal on a textual level which topics may be more present in a given text and to more quantifiably track changes between the 1818 to the 1831 editions of *Frankenstein*. I will experiment with a variety of different lengths of “chunked” sections of text and number of topics to see how this influences output. In order to conduct meaningful comparative topic modeling work on

the two editions of the novel and intertexts like Shelley's diaries, topic modeling will be run on all texts relevant to the project under one topic model.

For the preliminary test run, I created plain text files of the 1818 and 1831 editions of the text chunked by chapter (with the exception of Walton's introductory and concluding letters, which are grouped together into introductory and concluding chunks respectively and comprise what would otherwise equal the length of a single chapter). I generated twenty topics on the two editions of the text. Topics of interest include topic six, which pertains to the origin story of the novel. A closer look at this topic reveals that more than double the amount of words associated with topic six (594) are found in the 1831 introduction and preface to the text compared with the 1818 preface (223). This corresponds with Shelley's new introduction to the 1831 edition of the text, which details how she came "to dilate upon, so very hideous an idea."⁷

Figure One: List of Topics Run on the 1818 and 1831 editions of *Frankenstein* with Topic Modeling Tool

List of Topics

0. justine poor william father murderer elizabeth dear lake discovered victor evidence misfortune god picture return unhappy tears condemned murdered child
1. father elizabeth mind idea love beautiful sensations dear river england mountains change days boat journey heard bound valley victor lake
2. time eyes thought continued night long length father horror felt hands joy morning beautiful form knew place wretch dark rest
3. clerval miserable scene city visited work london switzerland journey arrived days oxford lakes proceeded mountains hills mind filled edinburgh scotland
4. felix man agatha cottage feelings cottagers words stranger heard protectors safie understood delight appeared language friends knowledge learned sweet gentle
5. science natural philosophy father studies elizabeth made modern professor names greatest life knowledge soul day clerval books felt world krempe
6. story spirit nature imagination things possessed tale author house subject speak geneva mind ghost write changed read ideas human means
7. sea land journey fiend vengeance words sun ice sledge ocean suddenly spirits heart hours moon fatigue sound wind peasants deserts
8. father elizabeth happiness love day letter marriage explanation cousin victor union misfortunes friend threat secret tranquillity case place affection disturb
9. night death misery life passed heard heart despair thought miserable feelings eyes hope fear dear human horror friends joy elizabeth
10. ice man vessel dear sea years margaret sledge voyage stranger sailors enterprise 17 hear sister day grief strange soul greatest
11. found cottage wood man day fire young light night food sounds eyes pleasure cottagers gentle youth snow ground morning garden
12. elizabeth father child innocence mother years judges confessed court committed beaufort lady cousin heaven passed sister continued girl poor support
13. time man life friend eyes happiness place thought found heart thoughts day made return love mine feelings mind present country
14. man thy ice creature clouds rage miserable listen moved thou creator destroy mountain demand power hands tale make existence air
15. felix safie father daughter turk paris lacey de country leghorn part deliverer arabian agatha girl time lover cottage heart arrived
16. clerval justine dear favourite ill feelings cousin poor happy mother aunt affection country long henry felt eyes care time young
17. life human nature study discovery pursuit death great light father ardour object made engaged creation found science feelings soul mind
18. frankenstein friend ice feel voice return vengeance die death hopes thou england find love september glorious called sister imagination tale
19. room father kirwin mr magistrate body mind man light death clerval shore agitation felt put called prison fiend life frightful

⁷ From Shelley's introduction to the Standard Novels 1831 edition of *Frankenstein* (5).

Figure Two: Topic Six (Genesis of the *Frankenstein* narrative)

TOPIC : story spirit nature imagination things possessed tale author house subject speak geneva mind ghost write changed read ideas human means ...

top-ranked docs in this topic (#words in doc assigned to this topic)

1. (594) [Preface%20and%20Introduction%201831.txt](#)
2. (512) [Chapter%202%201831.txt](#)
3. (498) [Chapter%202%201818.txt](#)
4. (223) [Preface%20and%20Introduction%201818.txt](#)
5. (0) [Walton%20Letters%20Introduction%201831.txt](#)
6. (0) [Walton%20Letters%20Introduction%201818.txt](#)
7. (0) [Walton%20Letters%20Ending%20Continuation%201831.txt](#)
8. (0) [Walton%20Letters%20Ending%20Continuation%201818.txt](#)
9. (0) [Chapter%209%201831.txt](#)
10. (0) [Chapter%209%201818.txt](#)

Twenty topics generated with MALLETT on the same chunked sections of the novel also uncover interesting insights into the text. Also looking at the preface of the 1818 edition and the preface and introduction of the 1831 edition of the text, there is a substantial difference in the frequency of topic eight, relating to the origin of the narrative (61% in the 1818 edition of the text and 36% in 1831), which reveals that this discrepancy is likely due to the far more extensive 1831 introduction. These test runs demonstrate the potential of more detailed, future topic modeling work on the novel, and more in-depth topic modeling exercises will be executed on all relevant texts for the project.

Figure Three: List of Topics Run on the 1818 and 1831 editions of *Frankenstein* with MALLETT

```

frankenstein_keys.txt
0 0.25 found wood cottage fire man young night sounds food light pleasure cottagers gentle day trees girl pain youth hovel ground
1 0.25 room night father magistrate kirwin body account murderer corpse put called geneva frightful black man kind agitation neck shore bed
2 0.25 thy mountains creature ice valley thou tale mountain listen head scene soul rain thee begone precipices sensations effect chamounix abhor
3 0.25 scene visited clerval mountains journey days henry return idea elizabeth miserable city beautiful dear london england promise happiness rhine friend
4 0.25 sea found left sun determined rage felt time light town promise quit sat free waves work fiend leave wind man
5 0.25 frankenstein friend vengeance hopes voice ice men apothecary england imagination thou idea called find sister tale glorious attempted virtue return
6 0.25 clerval dear justine long ill time henry favourites felt cousin friend length poor spirits journey subject mother children ingolstadt feelings
7 0.25 time made mind man day appeared words passed world great eyes happy thought affection part place countenance men thoughts joy
8 0.25 true heart story progress labours friends nature geneva friend tale point task success summer amusement beginning plan food fancy undertaking
9 0.25 night journey path eternal vengeance pursued ice dead fatigue spirits spirit sight sledge food sea weeks direction gained immense departed
10 0.25 friend ice margaret sailors vessel years cabin enterprise voyage man sea deck letter undertaking stranger sledge sympathy noble sister courage
11 0.25 father marriage country happiness day union daughter father's mother died years misfortunes paris girl heart turk event native beloved death
12 0.25 life human found death creation pursuit frame desire study work feelings light native engaged object nature occupied mind feel secret
13 0.25 justine elizabeth child innocence house poor committed cousin victim passed court despair guilty dear judges confessed lady spoke heaven guilt
14 0.25 man sun rage suddenly revenge human hands moved replied rose power clouds destroy creator companion sensation hid consent demand destruction
15 0.25 life eyes death feelings heart miserable felt misery feel horror despair thought friends monster lost peace rest heard happiness existence
16 0.25 felix safie agatha cottage language man lovely arabian father learned friends stranger understood cottagers feelings protectors conversation blind lacey virtue
17 0.25 elizabeth william father lake victor justine poor dear geneva hope murderer mountains fear night calm discovered wandered brother letter god
18 0.25 found greater means mine door months continued enter creature life hands directed amiable poor circumstances fears sorrows unfortunate beaufort poverty
19 0.25 natural science studies philosophy father spirit walton krepps modern life soul agrippa knowledge elizabeth years names changed nature professor cornelius

```

GIS Mapping and Timelines

I will conduct GIS mapping of events in the novel and character paths using platforms like Esri Storymaps, Neatline, and StoryMap JS. Timeline tools like TimelineJS will complement the geographic mapping of the novel. Information to be mapped will be generated from a close reading of *Frankenstein* and careful log detailing character journeys and the chronology of the frame narrative.

Social Network Analysis

Using visualization tools like Gephi, I will generate social network visualizations on the novel by extracting information from the novel on dialogue and character relations. MALLET topic model outputs generated on the corpus of *Frankenstein* and related texts will also be used to generate Gephi visualizations using tools like MALLET-to-Gephi Data Stacker. To generate a social network model of *Frankenstein* scholarship networks, I will mine works published on *Frankenstein* and examine scholarly relationships using databases like Google Scholar, Project Gutenberg, and SNAC.

Voyant Tools Analysis

I ran some test runs on the 1818 and 1831 editions of *Frankenstein*. While more detailed analysis will be conducted on word frequencies in the text (and will be supplemented by AntConc as needed), there are some initial observations from running word cloud frequencies on the two editions of the text. “Time” appears more frequently in the 1831 edition of the text, as does “death,” and “heart.” “Father” is more frequently used in the 1818 edition, which may help support the thesis, for example, that Victor’s neglect of his creature and association as the monster’s father figure is more prominent in the 1818 edition.

V. Work Plan and Schedule

Project duration: May 1, 2017 - January 2018 (nine months)

This project will span from May 2017 to January 2018, ensuring that *Digital Frankenstein* will be completed in honor of the bicentennial for dissemination of the project at events and conferences relating to the 200th anniversary of the text's first publication. Text mining and visualization parts of the project will be generated incrementally and blog posts written after the completion of each given method and project outcome, beginning in stage three.

Project Stages

First stage: Conversion of all literary texts to plain text; close reading of the texts to record character journeys and chronology; data mining for social network map of scholarship networks.

Second stage: Topic modeling of *Frankenstein* and relevant intertexts generated with MALLET; using “R” to model data outputs.

Third stage: Creation of visualizations (GIS maps and timelines; social network visualization of the novel and MALLET topics; social network visualization of scholarship networks; Voyant Tools visualizations); writing of blog posts analyzing outcomes of data mining and visualization experiments.

Fourth stage: Project wrap-up and reflection; any necessary follow-up tasks; preparation for final product dissemination.

VI. Project Budget⁸

Item	Title	Description	Planned Expense Period	Total Hours	Hourly Pay Rate	Funding Needed
A. Salaries						
Name						
Giorgina Paiella	Project Director	RA for data mining; creation of visualizations; writing of analytical posts	5/1/2017-1/1/2018	720	\$25	\$18,000
Subtotal: Salaries						\$18,000
B. Benefits						
Name					Benefits %	
Giorgina Paiella	Project Director	RA for data mining; creation of visualizations; writing of analytical posts	5/1/2017-1/1/2018		.01	\$180
Subtotal: Benefits						\$180
Total Direct Costs						\$18,180
Indirect Costs		44% of total direct costs				\$7,999.20
Total Costs						\$26,179.20

⁸Fringe benefits and indirect cost rates are drawn from the UC Santa Barbara Office of Research Budget Preparation website: <http://www.research.ucsb.edu/spo/budget-preparation/#IndirectCosts>. This project runs on a non-traditional timeline to be timely for the 200th anniversary, extending over the summer months and part of the academic year. For simplicity, I use the academic year fringe benefit rate for a graduate student worker (1%) versus the summer month rate (3.1%).

VII. Method of Project Evaluation

This project will be evaluated at each stage to ensure *Digital Frankenstein* is on track for completion within the nine-month funding period. After the first preparatory stage is completed and the second stage underway, written blog posts will start to be composed in the third and fourth stages of the project and published gradually on the *Digital Frankenstein* website. Future directions for the project include the addition of more intertexts to the corpus of texts to be analyzed and expanding the written scope of the project to a scholarly article to be published on results.

VIII. Final Product and Dissemination

This project will be completed in time for the bicentennial of *Frankenstein's* publication. The major output of the project is the *Digital Frankenstein* website. I also plan to present at conferences and events related to the 200th anniversary of the text's publication to discuss the project, its outcomes, and future directions.

IX. Sustainability

This project will be completed in a nine-month time frame, but will be useful to future scholars of the text beyond the bicentennial. The results of these investigations and various visualizations will be posted on a free website platform like [Wordpress.com](https://www.wordpress.com). Sustainability and preservation of content should not be an issue considering the time frame of the project, but another website platform can be used in the future to preserve content.