PoeIt: A Poetry Visualization

Paige Johnson*
University of Colorado

Caleb Hsu†
University of Colorado

ABSTRACT

We set out to discover a way to analyze poetry at a high level in an exploratory way using data visualization. By focusing in on the works of two very different American poets (Emily Dickinson and Walt Whitman) who happened to be contemporaries of each other, we have built several visualizations that allow users to dig into poetic structure, patterns, and themes. These visualizations reveal statistics like average line length and word frequency, and visualize structure by visualizing the poetic choices made by authors, including words by line and syllables per line. Through this, we have discovered that by using distant reading with poetry, even subtle differences between poets become apparent. This works allows users to come to their own conclusions about the two poets without having to read through the poems word by word.

1 INTRODUCTION

Within the past few decades, the advancements of technologies have allowed for us to to develop computational tools and that support distant reading. Distant reading is a term coined by literary scholar Franco Moretti to describe critical approaches that seek to understand literature and literary history by aggregating and quantitatively analyzing large text corpora. [6] Poetry is a unique form of literature that we believe can greatly benefit form distant reading. We decided to take the writing of two very similar yet different poets to demonstrate this.

Emily Dickinson and Walt Whitman are two famous American Poets that both lived during the 19th century. While both poets pioneered modernism in American poetry, they are famous for having their own unique style of writing. Walt Whitman lived a carefree and easy going life, which is reflected in his poems, while Emily Dickinson was very structured and conservative. We took famous works by both of these poets and visually compare and contrast them. Two ways that we believed we could show the differences in the styles of poetry was by focusing on the length of the poems as well as the vocabulary used.

2 RELATED WORKS

One of the most well know and polished visualizations in field of poetry is Poemage. [6] Poemage is a visualization system for exploring the sonic topology of a poem. It allows for distant reading by focusing on the phonetic structure of a poem. Poems are rich with unique ways to interpret data. There are three types of information that we can analyze in poetry. Formal information (e.g., lines, stanzas), phonetic information (e.g., meter, intonation, timing), and semantic information (e.g., genres, words, repetition, sentiment). [1] The visualization that we propose analyzes both formal and semantic information.

With the growing access to raw information, there comes a need for non-technical people and scholars to be able to access that information in a visual and intuitive way. [5] While the ability to search for keywords or phrases in a collection is now widespread, such search only marginally supports discovery because the user has to decide on the words to look for. On the other hand, text mining results can suggest interesting patterns to look at, and the user can then accept or reject these patterns as interesting. [3]

One area of poetry that we have focused on is the frequency of words used. Observing repetition in literary prose comes naturally during the normal process of reading. [8] The idea of using word frequency to gain an understanding and overview of text is not new. Word clouds are a straightforward and visually appealing visualization method for text. [4]There have even been more unique visualizations like a parallel word cloud that are designed to help users understand the differences between two different massive text sources at a glance. [2]

Since poetry can be complex with many layers, researches are proposing multiple levels of visualization, starting from verses and moving up to poems and comparing thousands of poems together. [7]We decided to start of looking at individual poems and their composition, as well as comparing the word frequencies across many works.

3 DESIGN

PoeIt consists of four main data components: an overview table, a poem structure visualization, a syllable frequency graph and a word bubble visualization. The overview table displays basic information related to the poem, including the title, number of lines, and average line length. Although this overview is not presented as a visualization, we chose to reveal the data as numbers because it gives users simple and concise intuition into the kind of poem they will be dealing with. From the get-go, a user comparing Dickinson with Whitman will notice that Dickinson's poems average a much lower number of total lines and line length compared to her more verbose counterpart.

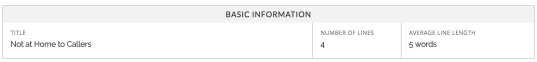

Figure 1: Poem Overview

The next data component is a poem structure visualization. As you can see, we turned each word in a poem to the same length block and then display it on the appropriate line.

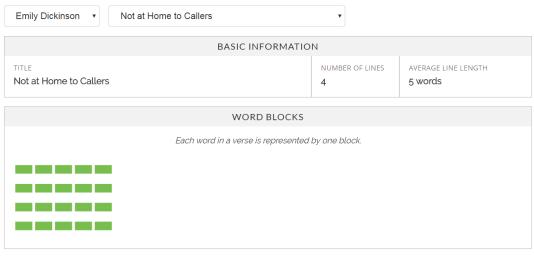

Figure 2: Dickinson's Block Poem

The poem structure visualization reveals the line-by-line structure of each poem by representing each word as a block with the same di-

*e-mail: pajo3557@colorado.edu
†e-mail: chhs9974@colorado.edu

mensions. Because we are more concerned with structure than word length, the length of each block is not dependent on the length of the word. With this visualization, it becomes immediately apparent when punctuation is used stylistically – for instance, Dickinson's poem "Not at Home to Callers" makes use of caesuras to bring each line to 4 "words". The resulting visualization as seen in Figure 2 indicates a very uniform and concise style of poetry.

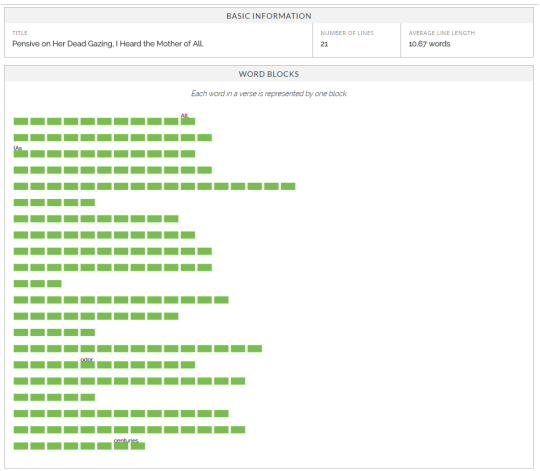

Figure 3: Whitman's Block Poem

On the otherhand, Walt Whitman's carefree and verbose style of poetry can be deduced from a wordblock of his poem "Pensive on Her Dead Gazing, I Heard the Mother of All."

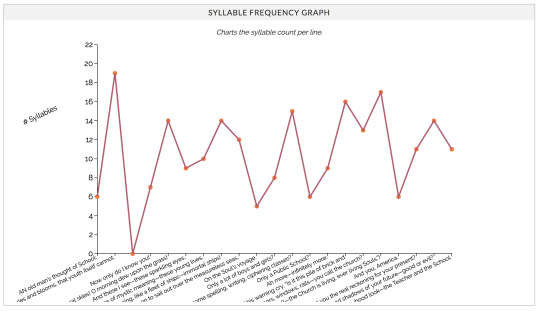

Figure 4: Whitman's Poem Line Graph

The next visualization is a syllable frequency graph. This is a line graph that visualizes the number of syllables in each line of a poem. Each point is consists of its line on the x-axis, and the syllable count for that line on the y-axis. Here in Figure 4 you can see Walt Whitman's poem "An Old Man's Thought of School," and its dynamic structure simply by taking a glance at the graph. Upon closer inspection, users can dig into patterns between syllable counts, and more advanced literature users may be able to infer the meter a poet is using based on those counts.

The word bubble visualization displays all the words used by an author across their body of work and represents frequency by bubble size (We ignore the 30 most common words from both to eliminate common words such as "a," "I," "the," etc). One of the advantages of this type of visualization is that it reveals the sorts of topics and themes that a poet is concerned with. For example, the casual user can observe that Whitman's poetry is often centered on people and scenes in their daily lives, as shown by a high usage of pronouns and 'common' words. Based on the popular movements of his era, one could correctly conclude that Whitman was a humanist, or someone who attaches high value to human needs and problems as opposed to the divine.

The bubbles are either attracted to each other or repelled based of the "Jitter" interactive slider. Since there are more bubble than will fit on the screen, the user is constantly making new discoveries every time the visualization is loaded or the Jitter variable is changed.

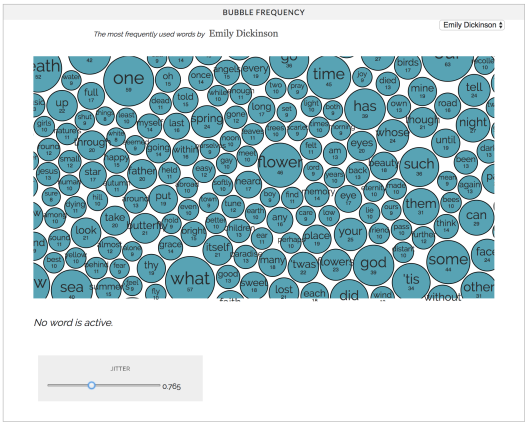

Figure 5: Word Frequency Bubble Chart

4 Conclusion

While there are an infinite amount of ways to explore poetry, we believe that we have created a new and unique way for people to discover poetry. We let the user explore our visualizations and come to their own conclusions by beginning with a broad overview, and revealing more details progressively in the subsequent visualizations. When testing our visualizations with exploratory evaluations, we have received spontaneous insights from our users, seeing as our data does not necessarily lead a user to a specific conclusion over time. Users were able to pinpoint Emily Dickinson's conservative and structured style and Walt Whitman's carefree work simply from word blocks, syllable graphs, and frequency bubbles. We believe that expanding our visualizations to other authors would yield the same insights. We are exited about PoeIt and the possibilities of distant reading in poetry.

References

[1] A. Abdul-Rahman, J. Lein, K. Coles, E. Maguire, M. Meyer, M. Wynne, C. R. Johnson, A. Trefethen, and M. Chen. Rule-based visual mappings with a case study on poetry visualization. *Computer Graphics Forum*, 32(3pt4):381–390, 2013. doi: 10.1111/cgf.12125

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

[3] A. Don, E. Zheleva, M. Gregory, S. Tarkan, L. Auvil, T. Clement, B. Shneiderman, and C. Plaisant. Discovering interesting usage patterns in text collections: Integrating text mining with visualization. In *Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management*, CIKM '07, pp. 213–222. ACM, New York, NY, USA, 2007. doi: 10.1145/1321440.1321473

[4] F. Heimerl, S. Lohmann, S. Lange, and T. Ertl. Word cloud explorer: Text analytics based on word clouds. In *2014 47th Hawaii International Conference on System Sciences*, pp. 1833–1842, Jan 2014. doi: 10.1109/HICSS.2014.231

[5] G. Lord, M. N. Smith, M. G. Kirschenbaum, T. Clement, L. Auvil, J. Rose, B. Yu, and C. Plaisant. 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 (JCDL '06)*, pp. 141–150, June 2006. doi: 10.1145/1141753.1141781

[6] N. McCurdy, J. Lein, K. Coles, and M. Meyer. Poemage: Visualizing the sonic topology of a poem. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):439–448, Jan 2016. doi: 10.1109/TVCG.2015.2467811

[7] A. Mittmann, A. von Wangenheim, and A. L. dos Santos. A multi-level visualization scheme for poetry. In *2016 20th International Conference Information Visualisation (IV)*, pp. 312–317, July 2016. doi: 10.1109/IV.2016.64

[8] S. Ruecker, M. Radzikowska, P. Michura, C. Fiorentino, and T. Clement. Visualizing repetition in text. *Digital Studies / Le champ numrique*, 1(3), 2009.