The aim of this work was to create a usable data set from botanical data gathered in the 1800’s by Henry David Thoreau for visualization and scientific purposes. A special focus was to preserve the work completed by Ray Angelo in his botanical index for future work. A Python script was written to scrape each HTML page on Angelo’s website to gather the botanical information and index information provided. The data has been useful in drawing out patterns in Thoreau’s exploration of botany. Specific plants are watched more closely, though his interest wandered throughout the botanical world. Future data work will be needed or helpful, including attribution of dates to specific pages and volumes, as well as updated botanical terms.

Further research is needed to complete the data sets. For example, dates (specially month and year) would be helpful in determining what Thoreau cited in winter months vs summer months. In addition, Ray’s index uses the Grey’s Manual of Botany 8th Edition. Since its use, other editions have been printed, and plant taxonomy has continued to improve. Updating botanical terms will be required.

Acknowledgments

This work was completed with significant help from Dr. Peter Hook, Cornell University, and Dr. Timothy Bowman, Wayne State University.