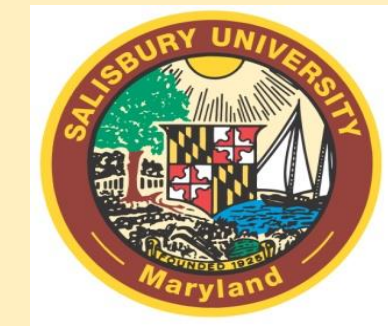# Software Tool Development for Analyses and Visualizations of Textual Information

Lauren Olson (Undergraduate)[1], Eliezer Volkis (Undergraduate)[2], Randall Cone (Faculty Advisor)[2]

[1]Smith College Department of Computer Science, [2]Salisbury University Department of Mathematics and Computer Science

Motivation

The goal of this project was to take a new approach in designing an artificially intelligent expert system for natural language processing.
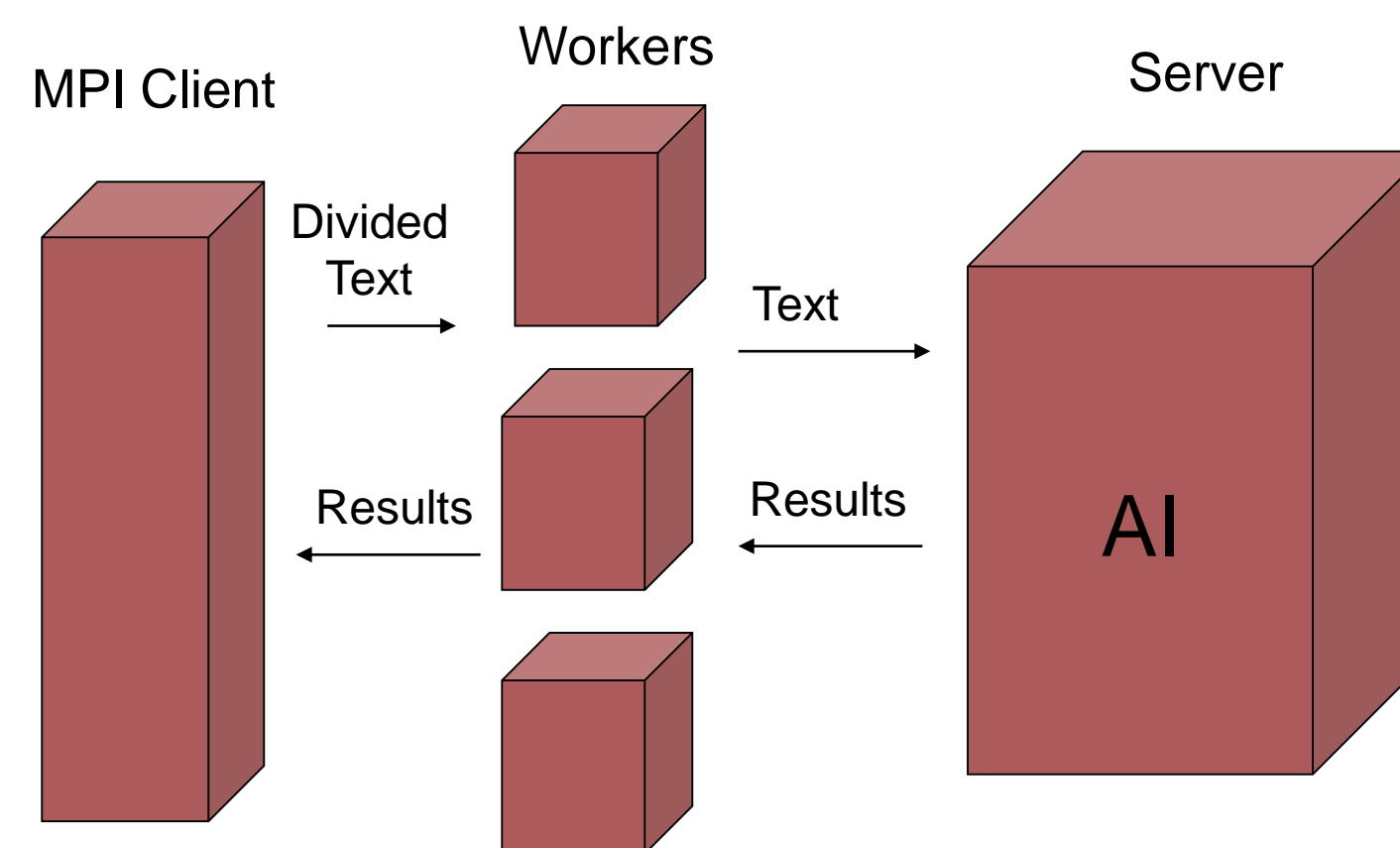
Introduction

Our AI analyzes the emotional meaning of textual information. The AI has a bootstrap, comprising of 60 words, as established by Watson and Clark in the PANAS-X (1994). The bootstrap then connects to WordNet for a complete analysis of target words. The AI then outputs the analysis on the text.

Method

We created a Google-Form survey, and took surveys from a sample population. We then converted the results into a JSON file, which is passed to the AI. Using this information, we can customize the AI to tailor to individual personalities. The AI operates off a server and receives connections from clients to analyze the individual words of the text. Similarities are then computed for each word with the 11 emotional categories from PANAS-X. These similarity values are multiplied by the average score for words in each category on the survey. The 11 emotional values will then be returned as a vector for that word, and will be kept in a dictionary for future text analysis. Results of the analysis will then be returned to the client.

Basic Project Design

Sample Output

This image was generated from Dr. Cone's previous work in detecting emotional content in text. The text analyzed is displayed along the bottom of the image and the graph shows the emotions strongly related to each word. The image is a custom SVG generated from Python code. It is a sample of a potential visualization for the output from the AI of this project.

Image from: "Perchance to Dream: Art, Mathematics, and Shakespeare"

Details

Our AI system was developed from scratch using Python. It makes use of libraries such as NLTK (natural language tool kit), gspread, and mpi4py. NLTK is a library for processing text files; it allows for easy tokenization of words, eliminating stop-words, and classifying parts-of-speech. It also has a built-in access to WordNet, and has functions to view synsets (synonyms) and similarities of words. Gspread is a library connecting Python programs directly to Google Drive via the Google Drive API, allowing us to access the Google sheet containing personalized results. Mpi4py links Python to OpenMPI, allowing us to write parallelized programs.

Future Work

To improve upon the current AI results, links to Word2Vec and ConceptNet may be added to the program. For future investigations, we can try to come up with a quantitative average of different personalities in a random sample. The results from the AI may also be used to potentially provide visualizations alongside classical texts. Other applications for the AI could be an email visualization software that provides a snapshot of the emotional content of the email.
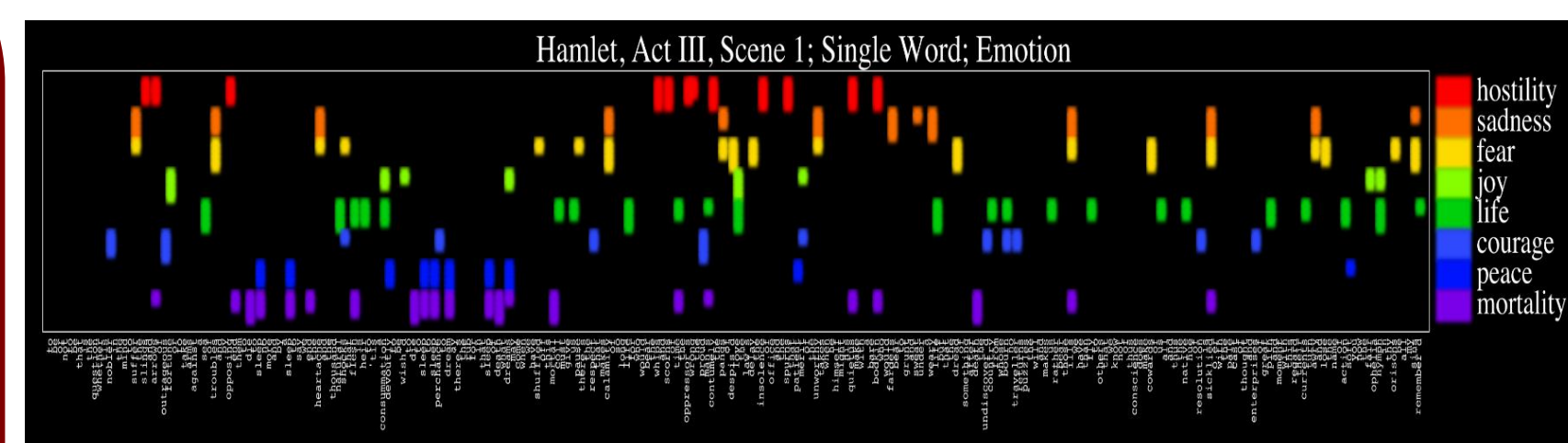
.

References

Watson, David and Lee Anna Clark. *THE PANAS-X: Manual for the Positive and Negative Affect Schedule - Expanded Form*. Ames: The University of Iowa, 1994. https://www2.psychology.uiowa.edu/faculty/clark/panas-x.pdf Accessed 26 June, 2017.

Cone, R. E. "Perchance to Dream: Art, Mathematics, and Shakespeare" Journal of Humanistic Mathematics, Volume 7 Issue 2 (July 2017), pages 4-36. DOI: 10.5642/jhummath.201702.03 . Available at: http://scholarship.claremont.edu/jhm/vol7/iss2/3

Princeton University "About WordNet" WordNet. Princeton University. 2010. http://wordnet.princeton.edu

Acknowledgements