Vizbooks

Emotions within Literature

Nishant Agrawal
Sai Priya Jyothula
Sreeraj Rimmlapudi
Vishnusri Tannamala

Introduction

Visualizing text has become a field in itself within the Viz community. Academia, in general often has to deal with a lot of textual data. But, why do we need to visualize text at all? Dr. Jason Chuang from Stanford University says that we visualize text primarily to achieve four tasks – Understanding, Grouping, Comparing, and Correlating. With the overhaul in the way people have begun to pursue media, textual data is all around us. Apart from the age-old- articles and books we now have twitter feeds, blog posts, comments and tags, text messages and even computer programs!

Early forms of textual visualizations include word clouds, word trees/sequences. We have now come a long way since then.

For this project, we have decided to visualize 21 books made available to us for free through Project Gutenberg. Most of the books available on the website are those whose Intellectual Property rights have expired and are hence under public domain presently. This gave us a unique dataset rich in classical fiction; those including the works of the erstwhile Jane Austen, Conan Doyle, and Mark Twain to name a few.

The project aims to visualize various emotions and sentiments within a book. In particular, we are trying to see
(a) How emotions flow within a book all the way down to a paragraph level from the beginning to the end of the book and
(b) Mapping an individual character to a list of emotions to summarize what they undergo through the course of the book and finally,
(c) Encoding the interaction between two characters and intuitively represent the intensity of the conversation and what kind of the emotional spectrum it will fall into.

The ultimate aim of this undertaking is for it go beyond the realms of a classroom. Our intention was to create a framework with a strong base that people even with a slight interest in literature could use and enjoy. Having read most of these books ourselves, it was intriguing to see them in a completely visual format. This piqued our interest further into trying to include as many possible tasks as possible. Most researchers who attempt to visualize literature are often hindered by the lack of the appropriate tools that are friendly and promote a certain level of ease-of-use. We aim to bridge that gap. Stefanie Posavec had to go through Jack Kerouac's book manually. So did the researchers at University of Calgary [1]; while working on *The Bob Gibson Collection of Speculative Fiction*, the team had a set of English literature scholars and students to parse through the collection and create metadata that can be used. This effort can now be slowly replaced by utilizing a variety of NLP packages[5] and emotion-word lexicons[6].

Video Link: https://youtu.be/p7DgjtwQlxE
Direct Link: http://saipriyajyothula.me/vizbooks/

Installation

Download the NLTK corbora (use **nltk.download** in python)

pip install -r requirements.txt

Extraction of Data

We used NLTK (Natural Language Toolkit) to extract (named) entities within the text. These were invaluable in figuring out the characters in the book. We also used a dictionary with keys as words and the values as an emotion vector which helped in calculating the total emotion thereby, giving us a quantified value that could be utilized throughout all visualizations. In order to get interaction between two distinct characters, we considered their co-occurrence in a single sentence or paragraph, i.e., if two characters appear in the same sentences, we assume they are conversing with each other.
In order to get the sentiment of each character individually, we consider a ten-word window before and after the character occurrence in the text. We believe that the character can be described by words surrounding him.

Understanding the Visualizations

Our dashboard has three different visualizations apart from the home page. The home page pays the ultimate homage to book visualizations. The home page is a matrix of all the covers of the books we are visualizing. Book covers play an extremely important role when a reader buys a book. It has the capacity to change a consumer's decision to buy the book or not. All in all, the book cover is a crucial element of the reading experience.

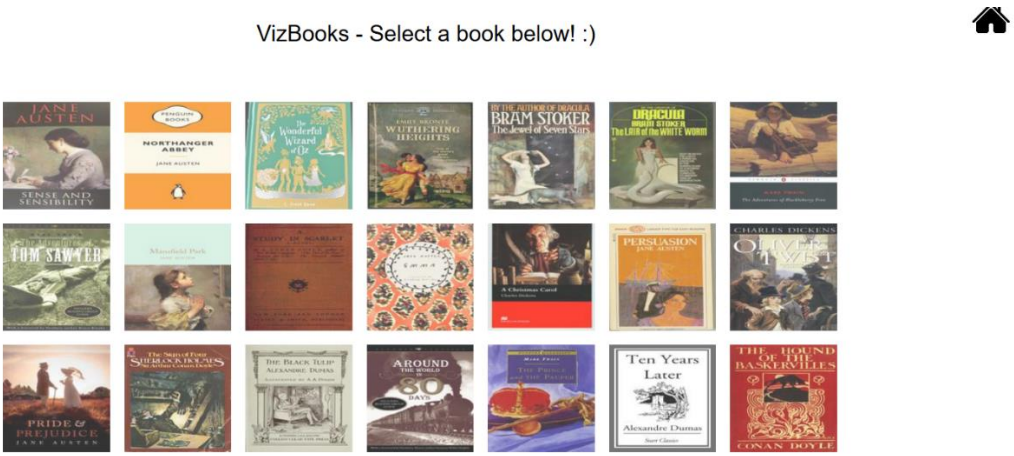

Figure 1: Landing page

Visualization one

Upon selecting a book, the user is redirected to a page containing three different visualizations.

The first is a radial tree containing spokes that represent chapters (segments) and sub-spokes that represent the paragraph segments. The sub-spokes have the interaction element to them.

A user can either select to see the overall ebb and flow of positivity and negativity in the book or (individual) emotions such as anticipation, sadness, disgust, joy, anger, surprise, fear, and trust. Each of these emotions is color-coded with intuitive shades for easy understanding. Red for negative, blue for positive. An emotion can be checked from the given options to view the paragraphs with a majority of sentences with just that emotion. Multiple such emotions may be selected or checked.

Upon selecting one spoke of the paragraph-segment, a pie chart next to the radial tree will show the division of all the emotions within that segment. Otherwise, the pie-chart will show the overall sentiment distribution for all the un-collapsed chapter-segment spokes.

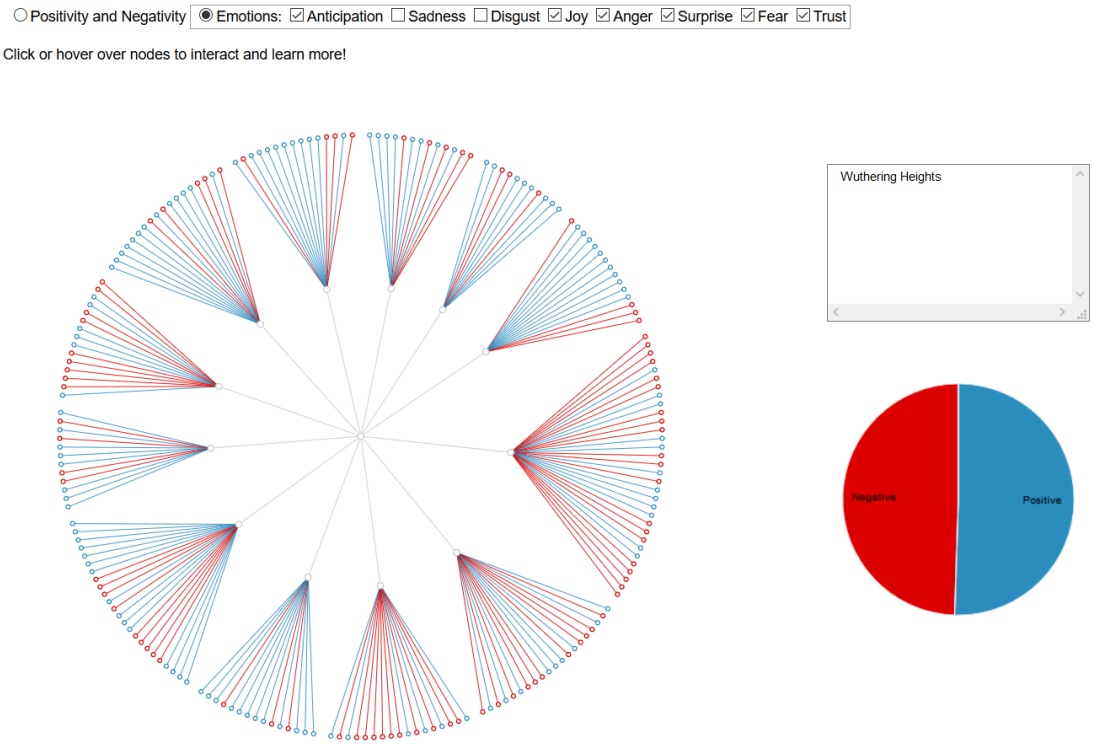

Figure 2: Overall positive and negative sentiment.

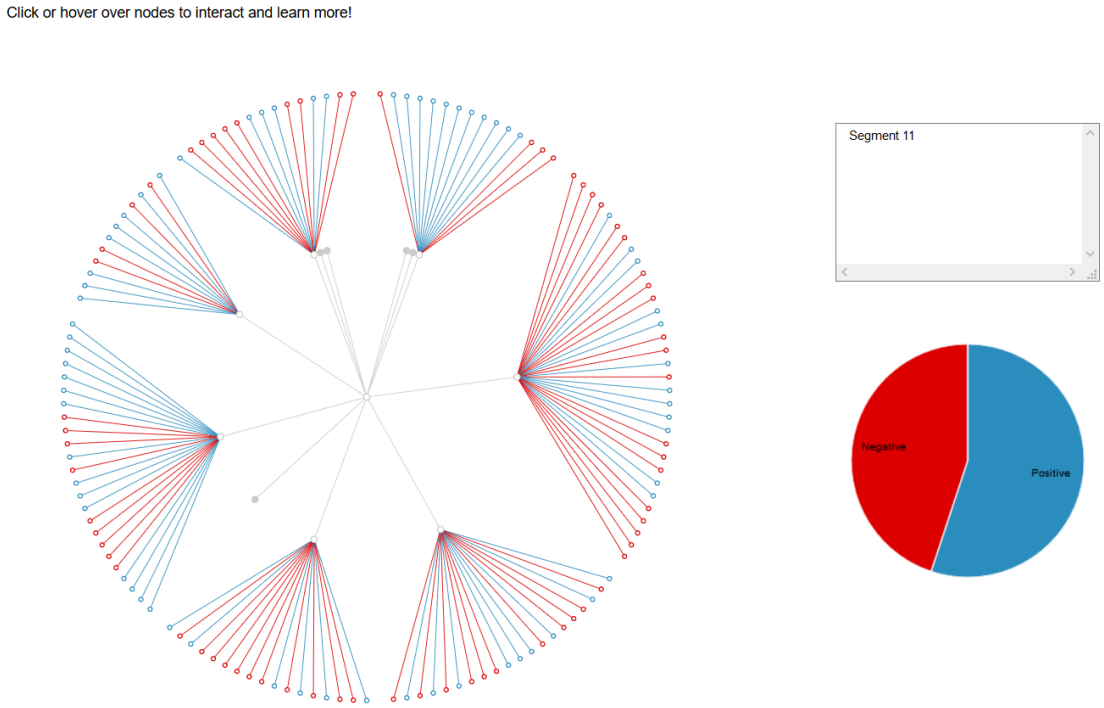

Figure 3: Same book example, with some chapters collapsed. Notice the pie-chart!

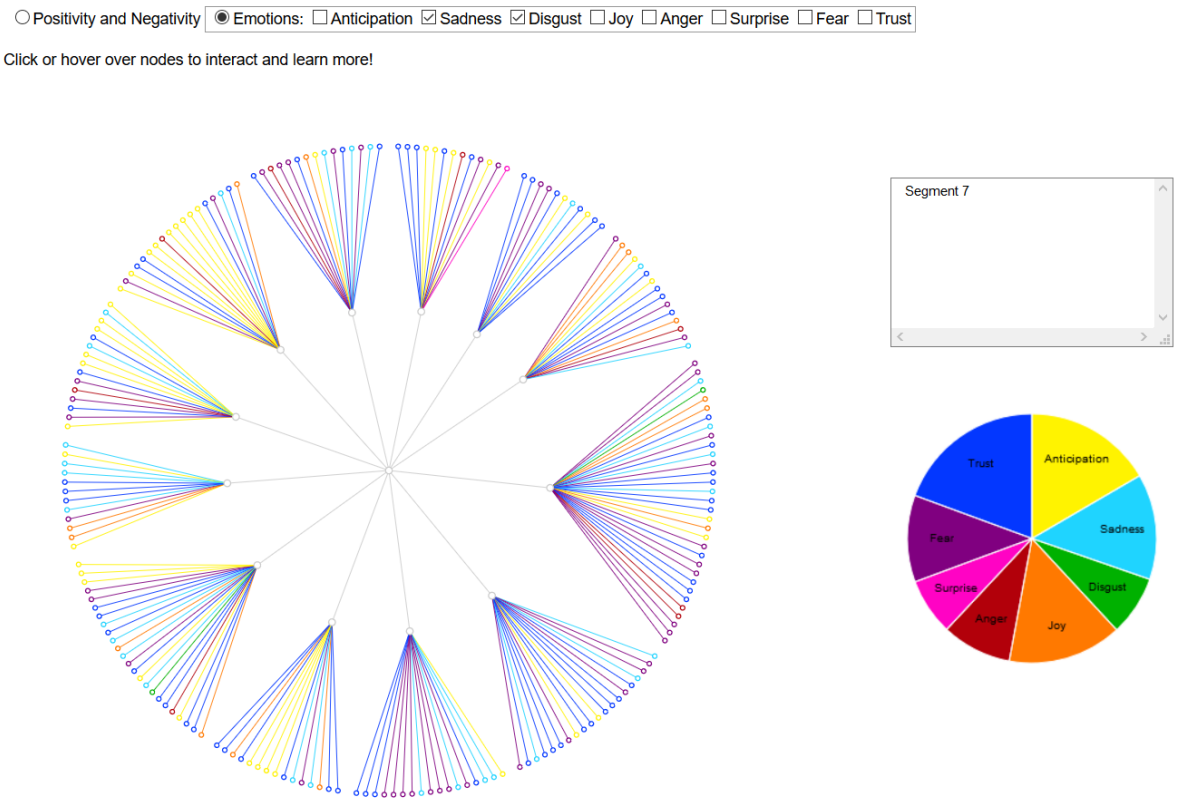

Figure 4: The same book visualizing overall emotions within a book. Pie-chart shows distribution.

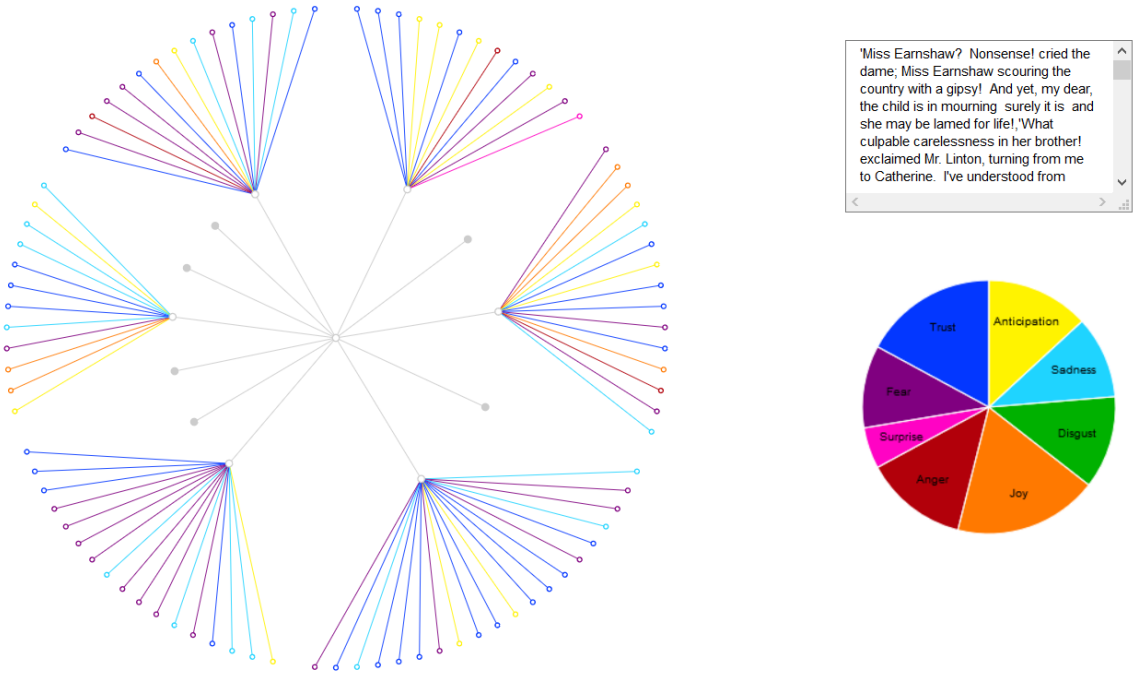

Figure 5: Text displayed when selecting the leaf node. Slight changes in the pie-chart after collapsing some spokes can also be noticed.

This visualization supports all search and most analyze tasks. Within *search*, the user can intuitively explore through the myriad of emotions within the book, browse and locate in case they are familiar with the book, and look up if they have a certain emotion in mind and are exploring just that. Within *analyze*, the visualization supports tasks such as present, discover, enjoy and annotate.

The second and third visualizations **rely heavily** on the chapters that are selected here. Changing the selection of the chapter here, in the first visualization affects the structures of the second and third visualizations.

Visualization two

The second visualization draws on data from the first visualization. Here, we try to map all the emotions the ten key characters in the book are either experiencing or associated with. This humanizes the project further as more often than not, readers are attached to and grow fond of a single (or multiple) character/s. This facilitates them to select just a single character and see the intensity of the emotions (that are

conveyed by the thickness of the path) as they flow from the character's name towards the emotion spectrum.

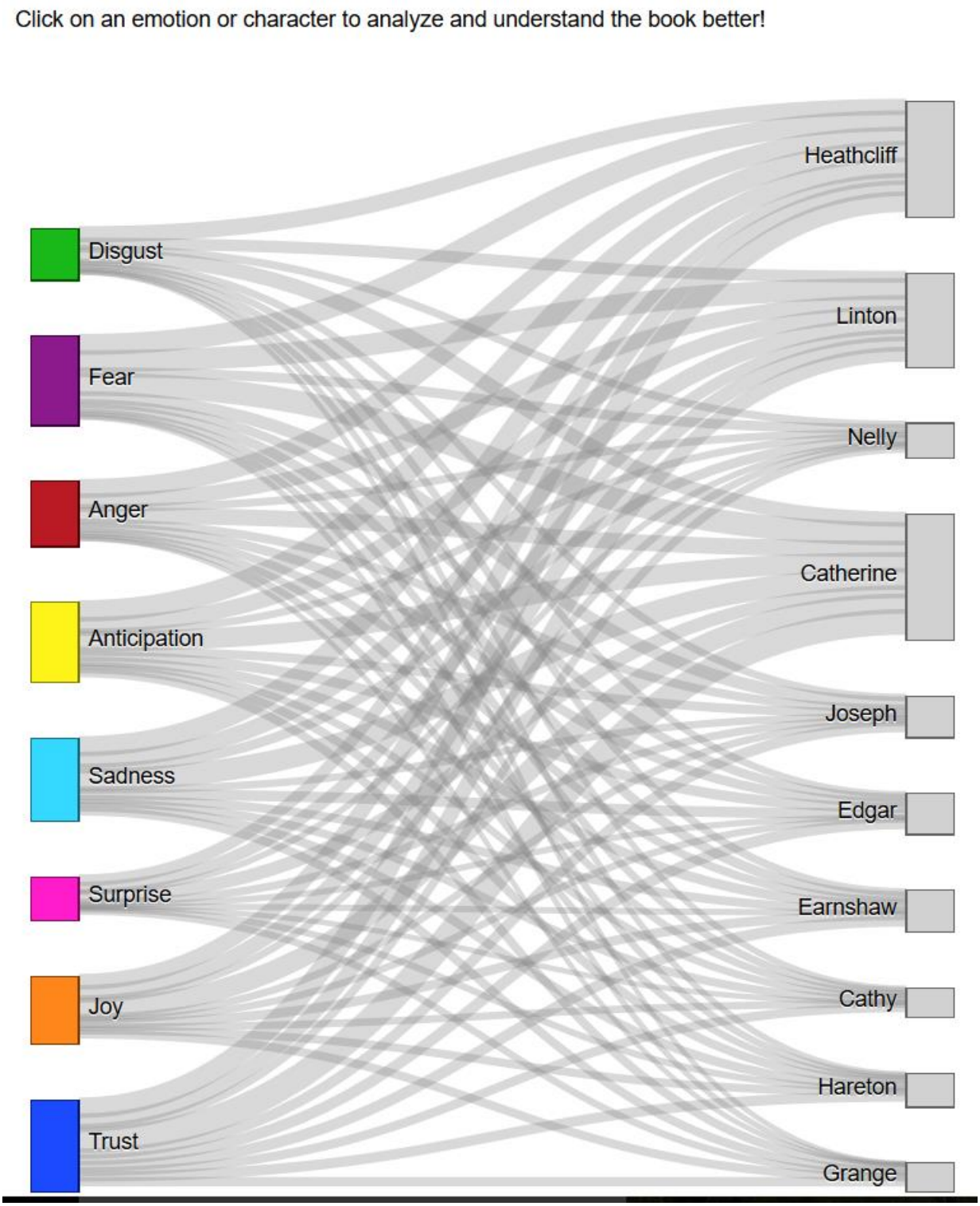

Figure 6: Overall Character-emotion association diagram

Interaction includes choosing just one chapter in the first visualization to see the character-to-emotion map just for that chapter. However, it can be a standalone visualization in itself. We also have a fun aspect to this visualization! You can click any of the emotions or the characters and drag them around and place them elsewhere, anywhere on the svg. Apart from the fun part, it also help in prioritizing emotions or characters that you want to focus on.

Furthermore, brushing and linking has been implemented both ways: we can click on simply one character to change the force-layout graph of the third visualization on the right. Similarly, we can choose only one of the emotions on the left side of the second visualization to change the force-directed layout on the right to change its structure.

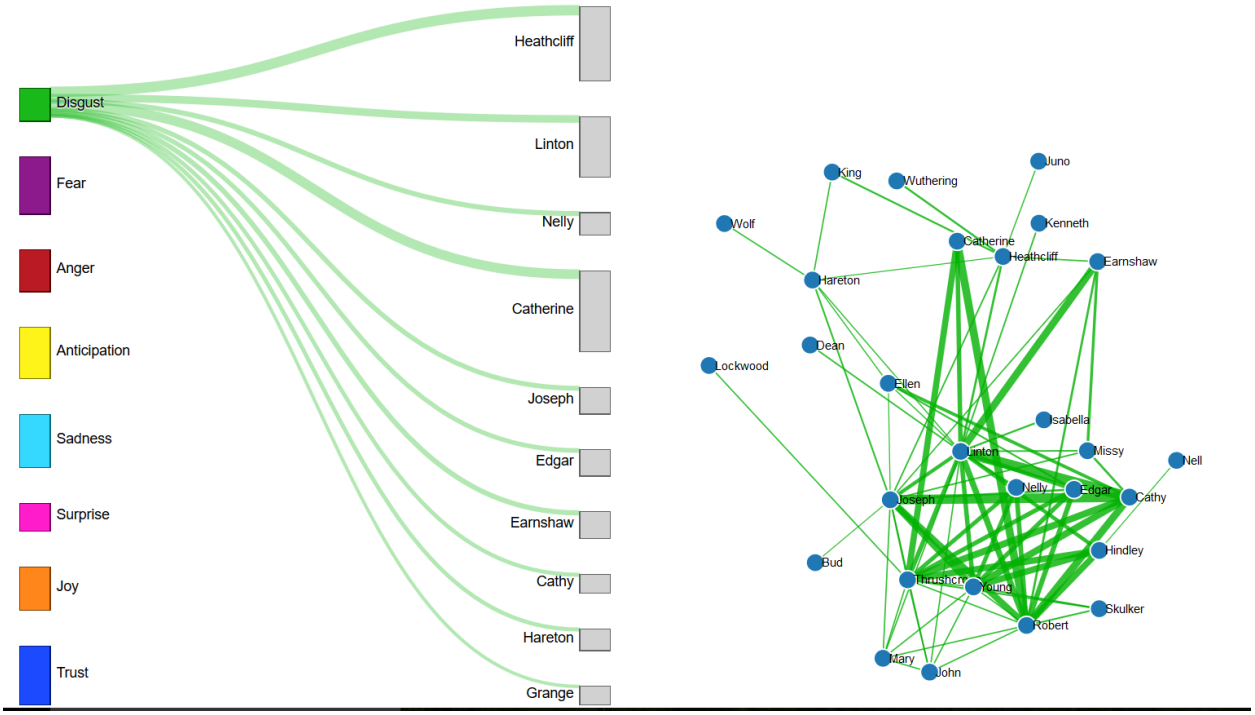

Figure 7: Single emotion to character association. Also shown, the corresponding character interaction layout of that emotion.

These second and third visualizations have a relationship with the first one as well. The second and third visualizations only show the filtered data of the chapters that are not collapsed in the first visualization.

It supports some tasks in *analys*ing and most tasks in *query* actions. Few such tasks include comparing, summarizing, identifying apart from discovering and enjoying. The visualization supports all the *search* tasks – lookup, locate, browse and explore.

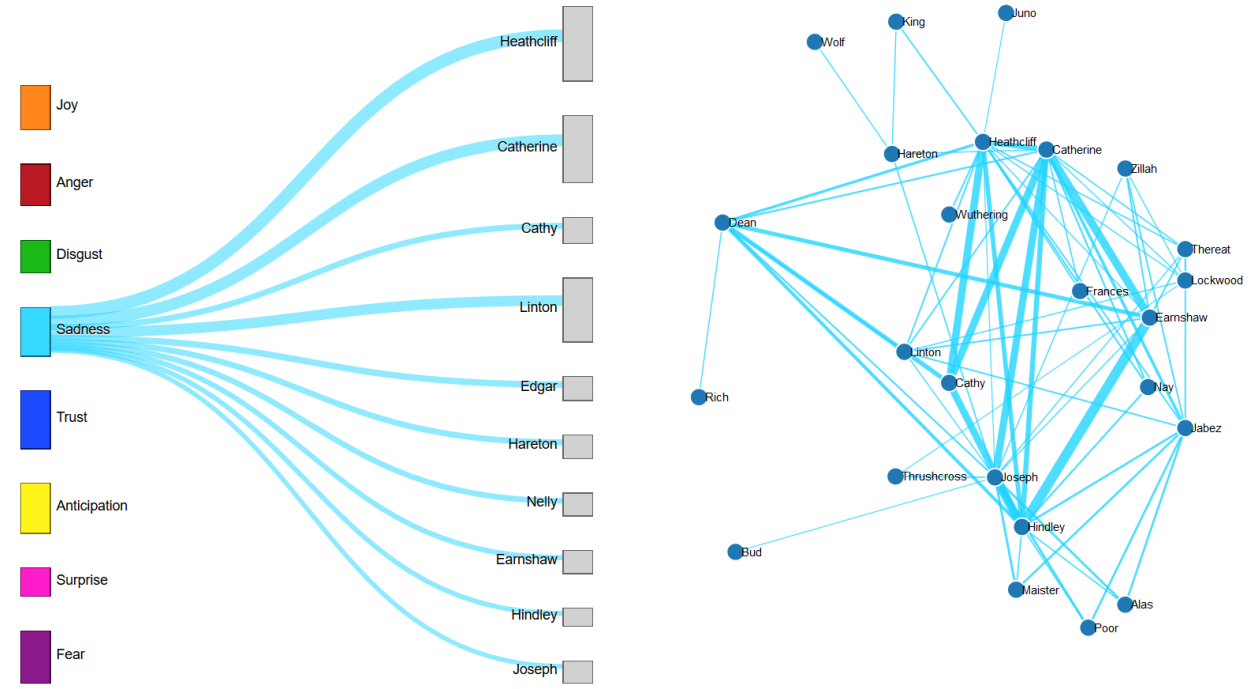

Figure 8: Another example of a single emotion to character layout

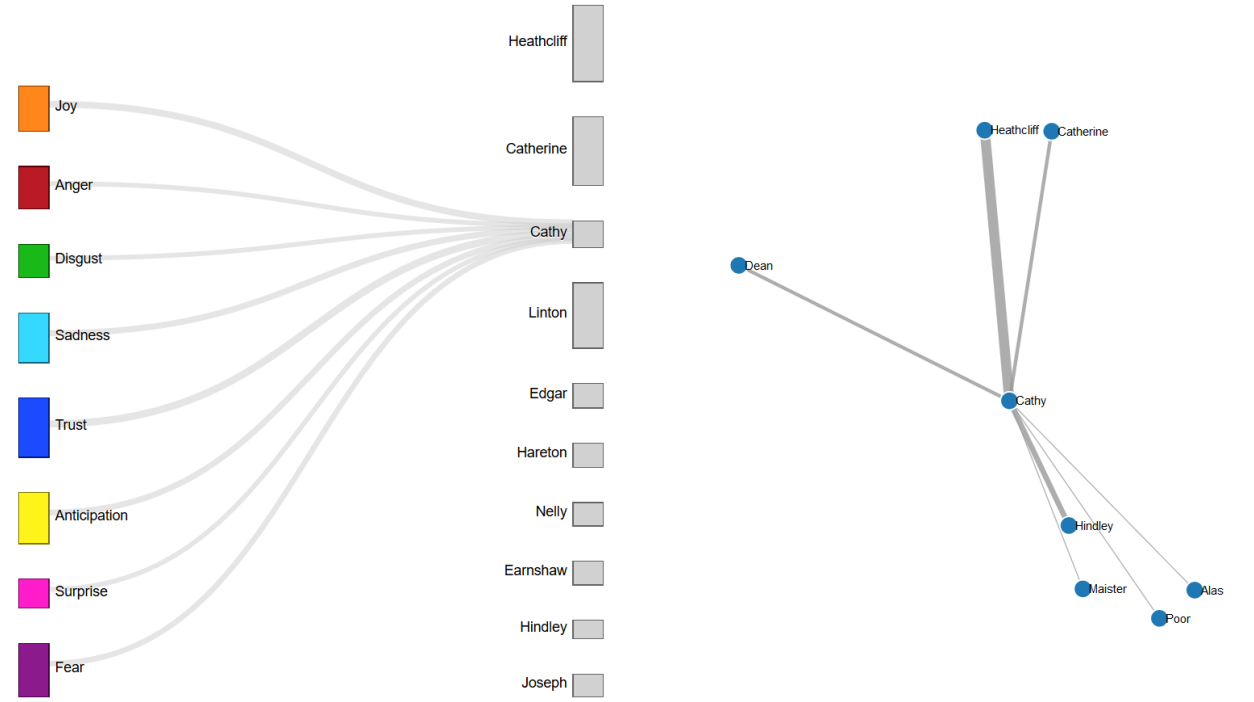

Figure 9: A single character-emotions diagram on the left and the corresponding character interaction diagram on the right (Notice how Cathy is associated with Alas and Poor even though they are not characters)

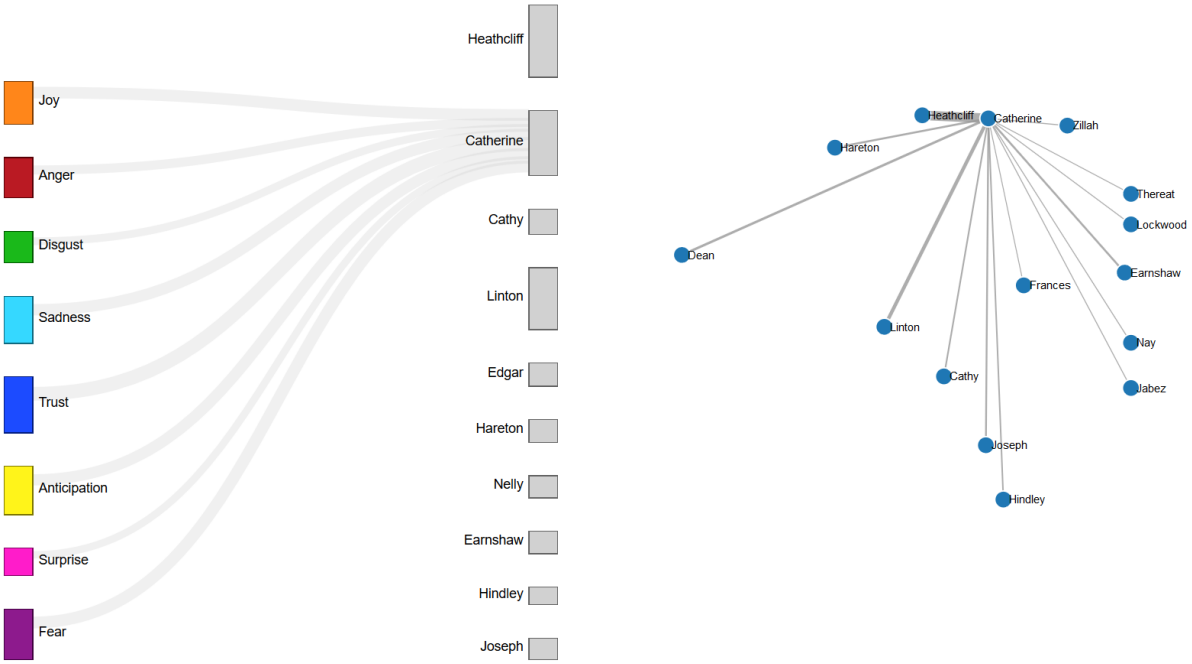

Figure 10: Another character-emotions association example

Visualization three

The third visualization is a force-directed graph illustrating the emotional intensity of character interactions. Each character is represented by a point mark and the interaction is represented between two nodes using a line mark. Channels control the overall layout and understanding of the visualization. More the thickness of the line, greater the score of that particular emotion. Colors of all the emotions are the same throughout the three visualizations in order to maintain uniformity across the dashboard. The thickness between two characters indicates the density of their interactions. The distance between two nodes conveys nothing.

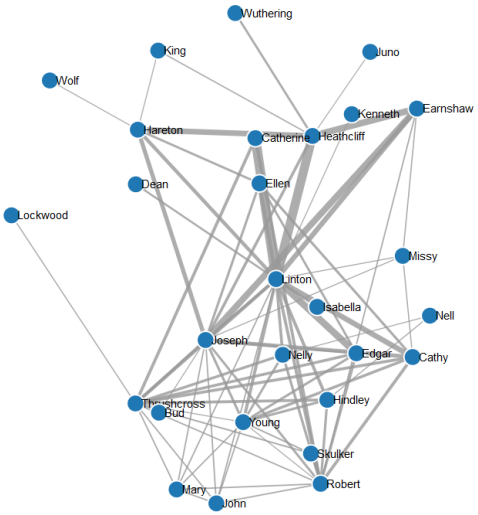

Figure 11: Overall key character interaction

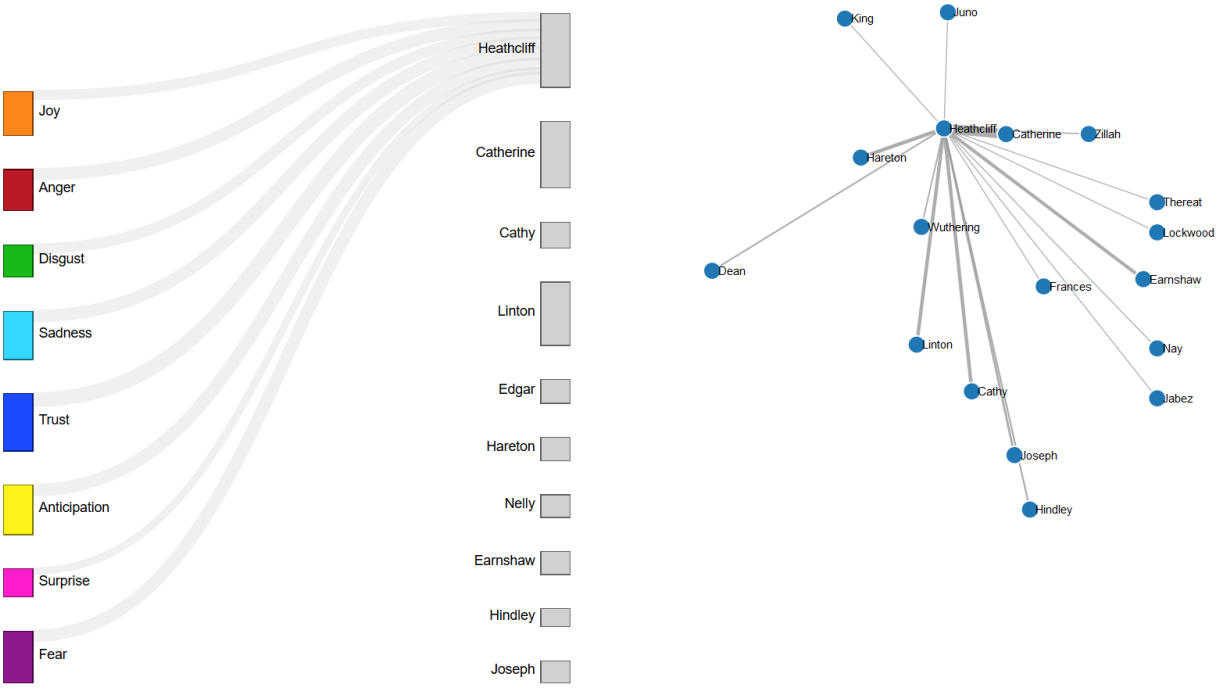

Figure 12: Heathcliff's emotion map and his interaction map with fellow characters

The structure of the graph layout changes as we select chapters in the first visualization and emotions within the second visualizations, thereby, implementing brushing and linking. Most of the task abstractions are preserved throughout this visualization as well. Major tasks that are supported are compare and identify within the *query* action; explore within the *search* action and finally present, discover, enjoy within the *analyze (consume)* action.

Inspirations

There are quite a few text visualizations that have inspired us while doing the project. The most influential one perhaps is Stefanie Posavec's *'Literary Organism'*. Her project 'Writing Without Words' visualized Jack Kerouac's *On the Road.* Having manually gone throughout the book, Ms. Posavec visualized many facets of the book such as sentence length, themes, parts-of-speech, sentence rhythm, punctuation, and the underlying structure of the text. [2] Although Ms. Posavec's visualizations are visually pleasing, they fail to have an interaction element to them. It only supports actions such as present, locate and browse. We have tried to incorporate interactions that have the element of enjoyment to them in our dashboard.

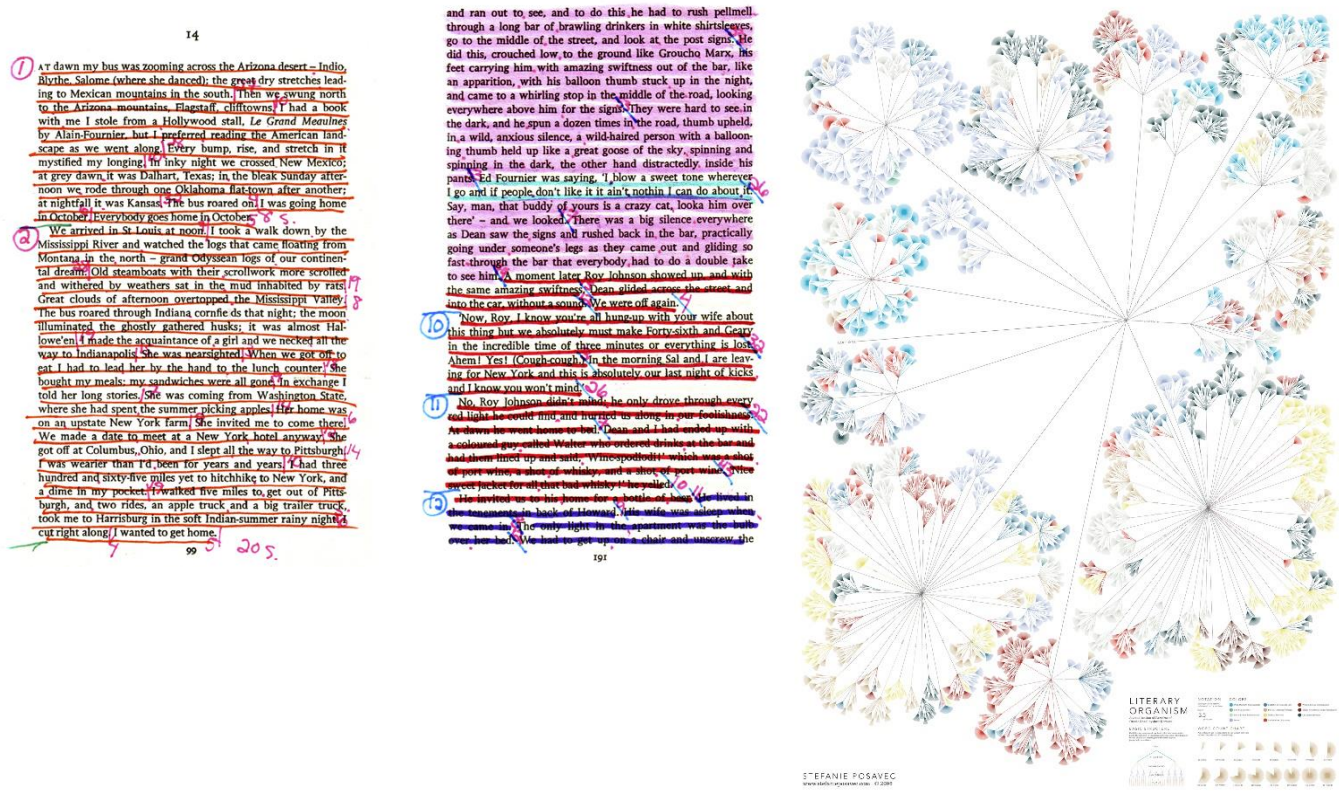

Figure 13: Manually analyzing text and the final 'Literary Organism' of *On the Road*

Another Inspiration was a *'Booker Prize Infographic'* created by Johanna Kamradt in 2011.[3] In this visualization, all of the 2011 Booker Longlist novels were broken down by various themes. All the 13 books were mapped to 20 different color-coded themes. Some interesting themes that the books were mapped to include an escaped tiger, biological terrorism, cannibalism, homicidal cowboy brothers, and nanny trust issues apart from the usual themes of death, love and betrayal. This visualization is appealing purely because of its varied and exotic selection of themes. However, most lexicons do not support such eccentric themes, which prompted us to stick with a standard emotional framework.

'The Bohemian Bookshelf', a project by Alice Thudt, Uta Hinrichs, and Sheelagh Carpendale was another influence in terms of interactivity and implementing brushing and linking.[4] It was fascinating to see how one selection in one visualization affected the other four visualizations, thereby bring the user closer to the goal of 'serendipitous discoveries'. We tried to implement a similar concept of visualizations changing based on selections (character selections in the force-directed graph and the plotlines visualizations)

The Team

Our team was made up of four members Nishant Agrawal, Sai Priya Jyothula, Sreeraj Rimmalapudi and Vishnusri Tannamala. Even with individual tasks assigned to each of us, we still heavily relied on each other's help and expertise throughout the project. Right from the initial brainstorming sessions while looking for datasets to all the final design decisions – were all taken together collectively.

While Sai Priya and Vishnusri worked together on cleaning the data (raw text files downloaded from www.gutenberg.org) and creating appropriate data sets that can facilitate easy visualizations. They also worked on the first visualization (radial tree of emotion). Nishant Agarwal and Sreeraj Rimmalapudi worked on getting the mining tools in place. They also worked on the second and third visualizations i.e., the key character interaction force-directed graph(s) and the key character emotion Sankey diagram. The final integration was done by everyone.

Room for Improvements

During the course of the past 4 weeks, we did encounter quite a few setbacks. One obvious one was the lack of a highly powerful Natural Language Processing package.

Another major hiccup was how a package perceives names. For instance, in Pride and Prejudice, the key character, Elizabeth Bennet is often referred to as Lizzy. In the Force-directed visualization that portrays character interactions, one can see two nodes- one for Elizabeth and another for Lizzy, although they are the same person. Another such example is with the works of Conan Doyle where there are two nodes for the famed detective Sherlock Holmes (Sherlock and Holmes).
‘

Another challenge we faced was perhaps being a little over ambitious. The force-directed graph turned out to be a hairball when we visualized interactions between all characters in a fairly long book. This was when we came up with the design decision of showing only the interactions between key characters.

References

[1] http://www.utahinrichs.de/uta/uploads/Publications/Publications/Hinrichs_scifi_2015.pdf
[2] http://www.stefanieposavec.co.uk/writing-without-words/
[3] http://www.slow-journalism.com/plot-lines
[4] http://innovis.cpsc.ucalgary.ca/Research/BohemianBookshelf
[5] http://www.aclweb.org/anthology/W/W11/W11-1514.pdf
[6] http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm