

# 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 and word lexicons.

Video Link: <https://vishnusri-gmail.tinytake.com/sf/MTA5MDU3M180MzUxMDMw>

## Installation

Download the NLTK corpora (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.

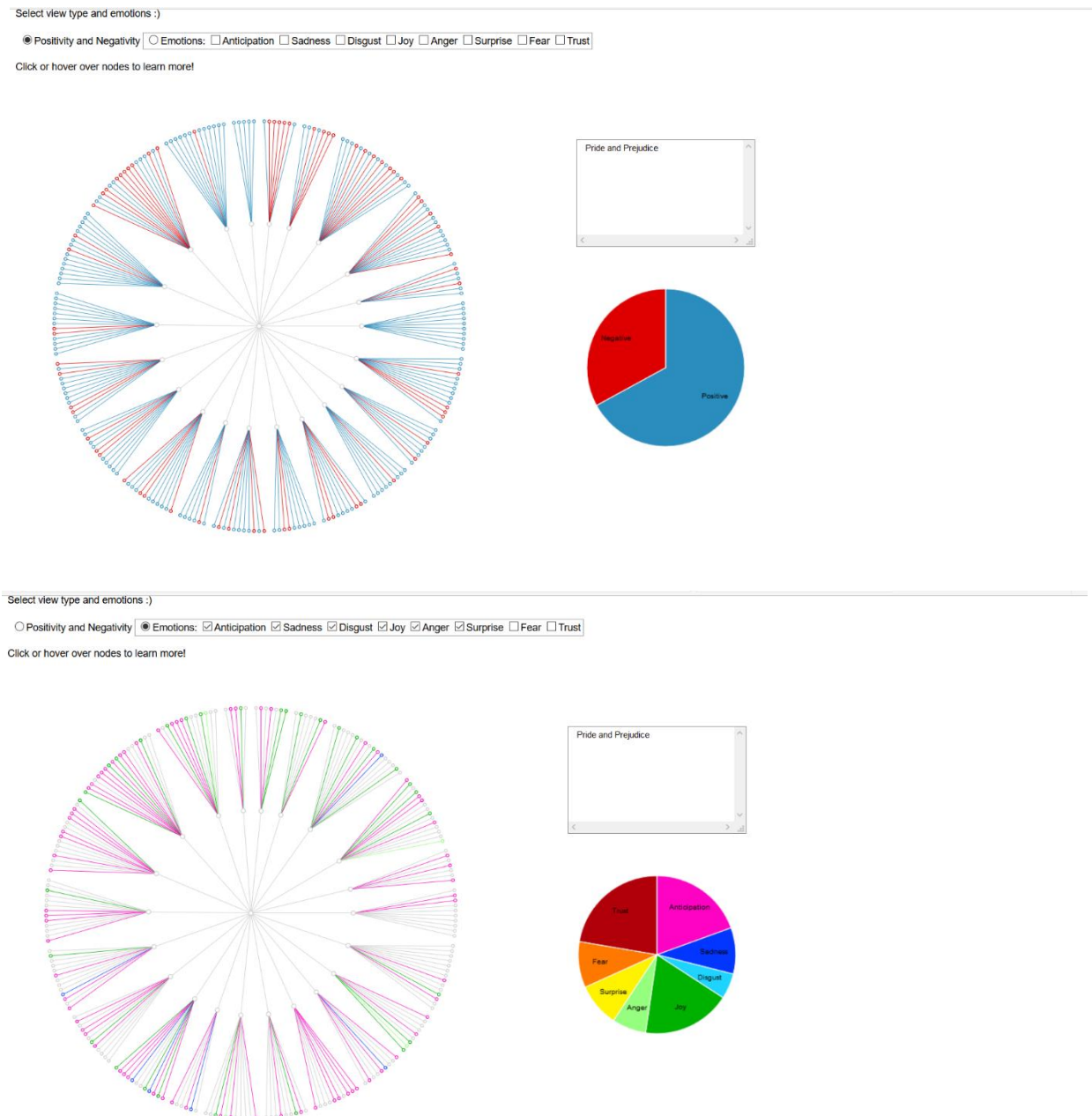
### Visualization one

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. An emotion can be checked

to view the paragraphs with a majority of sentences with just that emotion. Multiple such emotions may be selected.

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.



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

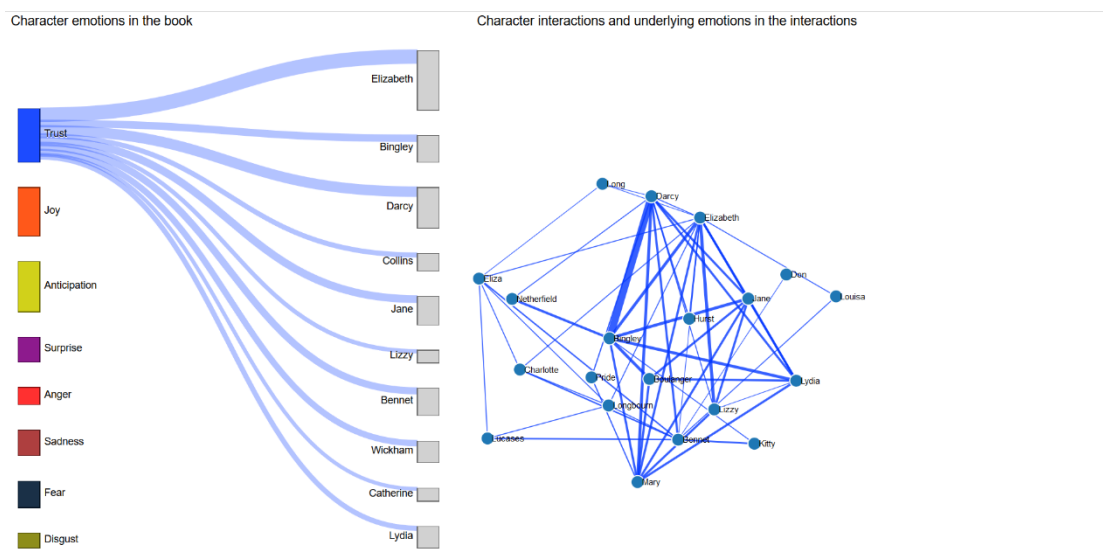
The second visualization draws on data from the first visualization. Here, we try to map all the emotions the ten key characters in the book are either experiencing or associated with. This humanizes the project further as more often than not, readers are attached to and grow fond of a single (or multiple)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

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.

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.

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.

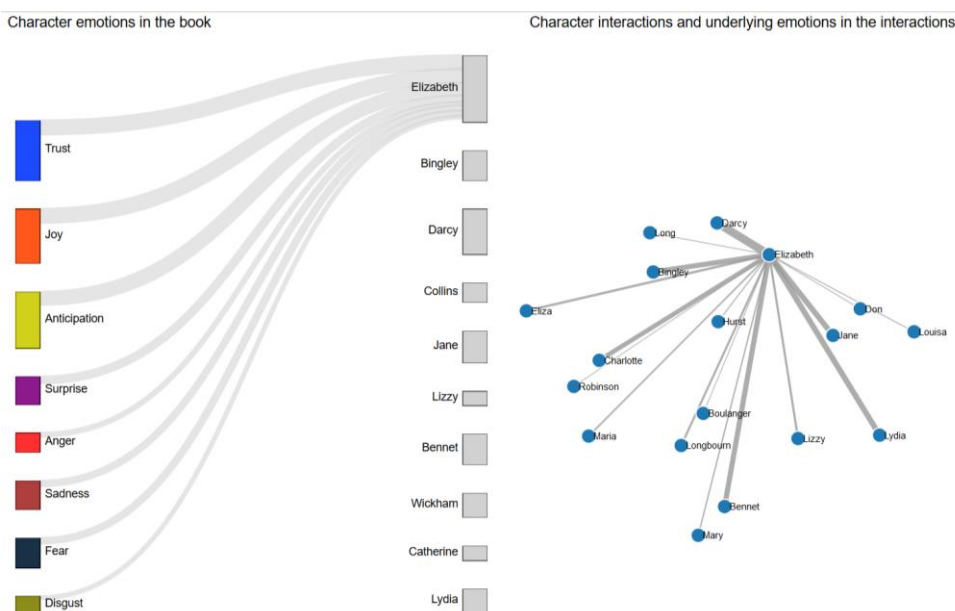
It supports some tasks in *analysing* and most tasks in *query* actions. Few such tasks include comparing, summarizing, identifying apart from discovering and enjoying. The visualization supports all the *search* tasks – lookup, locate, browse and explore



### 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 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



① At dawn my bus was zooming across the Arizona desert - Indio, Blythe, Salome (where she danced); the great dry stretches leading to Mexican mountains in the south. Then we swung north to the Arizona mountains. Flagstaff, Clifton. I had a boat with me I stole from a Hollywood store. The Grand Manitou by Alain-Fournier but I preferred reading the American landscape as we went along. Every bump, rise, and stretch in it mystified my longing. In larky night we crossed New Mexico; for a day, it was Dalhart, Texas; in the bleak Sunday after-dinner we rode through one Oklahoma flatdown after another; at nightfall it was Kansas! The bus roared on! I was going home in October! Every body goes home in October.

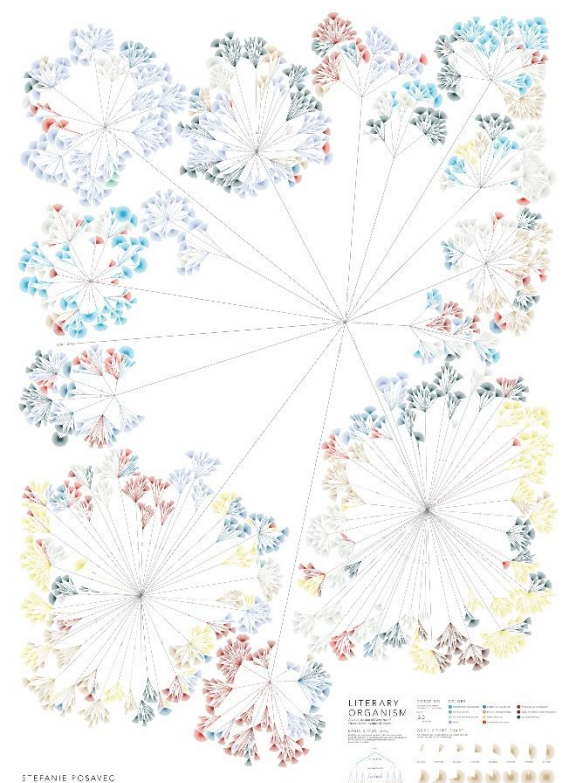
② We arrived in St Louis at noon. I took a walk down by the Mississippi River and watched the logs that came floating from Montana in the north - grand Olympian logs of our continental dream! Old steamboats with their scrowlwork more scrooled and withered by weathers sat in the mud inhabited by rats. Great clouds of afternoon overtopped the Mississippi Valley. The bus roared through Indiana corrañe at that night; the moon illuminated the ghostly gathered husks; it was almost Halloween! I made the acquaintance of a girl and we neeked all the way to Indianapolis! She was nearghous. When we got off, I had to lead her by the hand to the lunch counter; she bought my meals; my sandwiches were all gone. In exchange I told her long stories. She was coming from Washington State where she had spent the summer picking apples (the home was on an update New York farm. She invited me to come there. We made a date to meet at a New York hotel anyway. She got off at Columbus, Ohio, and I slept all the way to Pittsburgh - was wearier than I'd been for years and years. I had three huge meals and sixty-five dollars in my pocket. I took a bus and a dime in my pocket. I walked five miles to get out of Pittsburgh, and two rides, an apple truck, and a big trailer truck, took me to Harrisburg in the soft Indian-summer rainy night. I cut right along I wanted to get home.

99

and ran out the door, and to do this he had to rush pell-mell through a long bar of brawling drinkers in white shirtsleeves, to the middle of the street, and look at the post sign. He did this, crouched low to the ground like Groucho Marx, his feet carrying him, with amazing swiftness out of the bar, like an apparition, with his balloon thumb stuck up in the night, and came to a whirling stop in the middle of the road, looking everywhere above him for the sign. They were hard to see in the dark, and after a good while he saw the road, round up into a wide, lonely silence, a wild-haired person with a ballooning thumb stuck up like a great goose of the sky, spinning and spinning in the dark, the other hand distractedly inside his pants. Ed Fournier was saying, "I know a sweet note wherever I go and if people don't like it it ain't nothin' I can do about it. Say, man, that buddy of yours is a crazy cat, looks him over there - and we looked. There was a big silence everywhere as Dean saw the sign and ash rushed back in the bar, practically going under, someone's legs as they came out and giddling so fast through the bar that everybody had to do a double take to see him. A moment later Roy Johnson showed up, and with the same amazing swiftness, Dean pulled across the street and into the car, without a sound. We were off again.

"Now, Roy, I know you're all hungover with your wife about this thing but we absolutely must make Forty-sixth and Gary in the incredible time of three minutes or everything is lost. Ahem! Yes! (Cough-cough.) In the morning Sal and I are leaving for New York and this is absolutely our last night of Hicks and I know you won't mind."

No, Roy Johnson didn't mind, he only drove through every night-club scene first and second hand, and he had a good time. Dean and I had ended up with a cold glass and Ed-Walter wore ordered drinks at the bar and had them fixed up and said, "Wine-o'clock!" like a man and a woman were at the table, and a shot of port wine. "Vince sure jacked off that bad with that thing," he said, and then he invited us to his home for a bottle of his own. We took the taxi to the back of Howard's. The wife was asleep when we came in. The only light in the apartment was the light from the door. The wife was asleep when we came in. The only light in the apartment was the light from the door. The wife was asleep when we came in. The only light in the apartment was the light from the door.



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](http://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](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>