

# Log Data Visualizer

Aditya Gandhi, Ashish Sunny Abraham, Rishabh Srivastav, and Yash Goel

**Abstract**— We leave a lot of digital footprints when we interact with any type of technology. We can get time-stamped log data of these interactions. Analyzing the way a user interacts with any data conveys a lot about what they are thinking. But this is a very cognition-intensive task where we will be required to remember or annotate a lot of information to keep track of the interactive activities along with the data that was interacted with. The log data itself being highly textual makes the process even more tedious. If there was a way to represent this kind of data visually then it will be possible to understand the human rationale and strategies to a greater extent, even in a limited amount of time. Our application – Log Data Visualizer – aims to provide an effective solution for this very issue. We have created a visualization structure that will convert a large amount of textual log data into a visual summary so that it is easy to interact and gain insights from the given log data. This paper presents our current design, along with our future plans. We filtered and processed the sample dataset and created word clouds. We created a framework with two levels – upper and lower. The upper level shows a Treemap and the lower level shows a bar chart, bubble chart, word cloud, and a table of details. We explored the various aspects of data representation and integrated multiple visualizations in our design model to accommodate the different types of data and their connotations. We also discuss the pros and cons of this tool and the scope for improvement of this visual design.

**Index Terms**—Visual Summary, Visual Analysis, Log Data, Interaction Data, Treemap, Word Cloud, Human Analyses and data exploration.

---

## INTRODUCTION

The way a user interacts with any data reveals a lot about their thought process and their general behavior. Each person has a unique way of thinking and some unique behavioral characteristics that are developed based on their upbringing and their surroundings. Since every individual is different, therefore it is important to understand the user interaction data as it opens up a variety of use cases to implement the inference from this kind of data. We can use it to analyze, for instance, how the users interact with a website, and then on finding some trend in the log data, we optimize the website to provide a more intuitive interface and access to information. This is how we can create better solutions to cater to the needs of the respective audiences. Another use case is that we analyze the interaction log data of analysts and try to recreate the flow of thought with which they came to a conclusion. This is what our application aims to accomplish.

Our application aims at understanding the log data of analysts that comprise some specific interactions with documents, over a time period. Examples of such interactions include opening documents, reading documents, searching, highlighting, or the number of topics changed. These interactions give us insight into what the analyst is thinking or wondering. Analyzing these will help in understanding how the analyst arrived at a conclusion and what data did he consider the most important for his hypothesis.

But the challenge is that analysts are not keen to take notes while doing their analysis. It disturbs the flow of thought and they revise their hypothesis multiple times, so it seems a waste of time to take notes until they are sure of one thing. But when it comes to presenting their findings to their superiors or collaborating with their peers, it becomes difficult to explain their work without proper notes. There are ways to pre-process the log data but to make that amount of data available in an efficient way that facilitates deducing patterns catered to our needs, requires more research and innovation.

Our application's goal is to process the interaction log data and

then generate visual summaries that highlight the important interactions of the user. This way the analyst does not have to worry about taking notes of what he is doing as they are automatically recorded, and then we can extract that data, process it and present it later. The way we are achieving this is by first pre-processing (cleaning) the log data. Then we use the cleaned data to generate a word cloud for each analyst's behavioral analysis. After we are able to secure all the word clouds, we run those through the Natural Language Processing (NLP) models to categorize and then highlight the significant keywords. This result is used in our final visualization design to produce a visual report that will be able to present all the important activities concisely. The data and design exploration process has been done using tools such as Tableau which has features to create an interactive dashboard. We created a framework with two levels – upper and lower. The upper level shows a Treemap and the lower level shows a bar chart, bubble chart, word cloud, and a table of details. We explored the various aspects of data representation and integrated multiple visualizations in our design model to accommodate the different types of data and their connotations.

The paper is structured as follows: in the Section 1, we cover the literature review we did to form the foundation of our visualization design process. Section 2 describes our approach to creating the design, covering the process and exploration of both data and design. It also talks about the interface and the steps the viewer will take to navigate it. Section 3 states the evaluation performed by conducting user studies. We then conclude by discussing the limitations and future work in Section 4 and 5.

## 1 RELATED WORK

There is a considerable amount of research done in the efforts to find the best method to clean raw data and visualize the log data. But there has always been a need for a new improved visualization design for summarizing the log data, due to its textual and monotonous nature. Also, niche data such as interaction log data from analysis tools require a specific visualization design that would better represent the given data. We took note of many research works to build a foundation of knowledge so that we can take inspiration from each and try to find any solutions to their shortcomings in our implementation.

- 
- Aditya Gandhi. Student at University of Florida. UFID: 1782-3334
  - Ashish Sunny Abraham. Student at University of Florida. UFID: 6388-7782
  - Rishabh Srivastav. Student at University of Florida. UFID: 7659-9488
  - Yash Goel. Student at University of Florida. UFID: 5193-9756

First, we looked at the papers related to data mining and processing as that was the first step in our approach. Wesam Bhaya shared multiple strategies for data preparation [19]. These strategies make the data, which can be noisy and inconsistent, ready for mining. Joshi A.P. et al. also explain the activities for preprocessing of data [6]. All these steps help in creating high-quality data. We used these techniques in one way or another to process our dataset as well.

Understanding log data becomes another challenge. Z Shen et al. describes a two-tier visual analysis tool – TrailExplorer2 [7]. This tool can extract user activity and the time period when the user was doing the concerned activity. Such a tool acts as a reference for our project for the goal of creating an automatic visual summary generation model. H. Lam et al. describe the Session Viewer tool for the analysis of web session logs [8]. Their type of data is similar to the dataset we are considering, both having time-stamped data. Statistical methods are discussed in that paper that can be put to use in our future work. In the paper by Padmaja and Dr. Ananthi, they gave an explanation of how to divide the large dataset into smaller datasets and then apply visualization methods [11].

Next, we tried to understand the features and benefits of word clouds and the ways to automatically show visual summaries. Also, we looked into specifically log mining using NLP by C. Bertero et al. [17]. They utilized Google's word2vec method, a technique for learning words with rich vector representations. This way they were able to consolidate a log file into a single point. This point still consists of information for further classification. This method was something interesting to understand as we developed our own design model. Han Guo et al. share a case study that discusses methodologies quite suited to our own project [20]. The algorithms that are mentioned can be used to extract frequent words and arrive at some hypothesis to gain insights. F. Heimerl et al. talk about Word Cloud Explorer. It is a tool that enables users by providing an advance natural language processing framework [5]. Implementing such a framework will boost our implementation, and thus we plan to integrate it in the future. Show Me functionality as discussed by J Mackinlay et al. automates the presentation of commercial analyst systems such as Tableau [9]. Since we decided to use Tableau for our exploration of data and the creation of an interactive dashboard, this helped us to understand the working of one of the features of Tableau.

Coming to the design aspect of the project, we studied the guidelines that are given for creating various types of visualizations, before we came up with our own visualization design. We first looked at the paper by Moore and Purchase, where they talk about three requirements of complex visualization [1]. These include utility, soundness, and attractiveness. They then move on to give the attractiveness element of a design more spotlight. Jeffrey Heer et al. introduce Prefuse [3]. It is a user interface for interactive visualizations. Both structured and unstructured data can be used with it. This opened our eyes to the possible libraries out there that we can leverage in our design. One of the most important papers associated with the design study was by Michelle et al. [4]. It covers 8 guidelines that govern the use of multiple views, as we planned to have in our framework. These 8 guidelines include attention management, consistency, self-evidence, space-time resource optimization, decomposition, parsimony, complementarity, and diversity. As we were deciding on having Treemaps as our primary view, we read through a few papers related to the applications of Treemaps, which are worth mentioning in the references [13] [14] [15] [16]. They inspired us to include Treemaps in our design, by showing how versatile they can be in representing different kinds of data.

## 2 APPROACH TO VISUALIZATION DESIGN

Since we have based our design process on the given dataset and a usage scenario, let's first understand that scenario to understand the context of our design development.

### 2.1 Scenario

#### 2.1.1 Understanding the human analysis process

Mark is an intelligence analyst. He is given the task of analyzing a number of documents. He is required to report back to his supervisor if he finds anything out of ordinary. It could be a specific entry in the document that catches his eye or it could be a general pattern that he identifies among all those documents. To achieve this rather tedious task he goes through all the documents and highlights words in them. He also searches for some words in the document and changes the topics as he shifts his focus to new things. It is a very time-consuming iterative process and the last thing Mark wants to do is to have to jot down his thought process every now and then. It disturbs the flow of work and honestly is just downright boring. Now when he does come across an interesting finding, he is in a predicament. How is he supposed to present and justify this to his supervisor without the help of proper notes? His supervisor cannot go through all his scribbled notes, if there are any, to understand and validate how Mark came to the conclusion. Without verifying the authenticity, they cannot take action on it. It is very difficult for Mark to explain his strategies because how a person explores the data is complicated, and also very unique to an individual. This problem can also occur when he asked to collaborate with another analyst. Without any means for Mark to explain the way he explored and analyzed the dataset to the other analyst, it is next to impossible for them to work together on the same dataset as there could be conflict on what is being investigated, or there could just be redundancy or inconsistency in the findings.

### 2.2 Design Rationale Summary

The goal of the project is to analyze analysts' interactions and attempt to map their behavior to determine their actions. The datasets include user interactions with three separate document directories for eight different analysts. An analyst conducts operations like opening a document, reading a paper, searching for and highlighting a keyword, and changing the subject. The entire interaction was separated into multiple chunks of varying lengths to better understand how the user behaves. This segmentation of data adds randomness to the model and prevents bias.

The model was created by loading all the libraries and datasets into the system. Then, an iterative strategy was adopted to skim through the documents and extract all the necessary data based on the interaction type with the help of a function. The "Documents read" column in our initial log data table was also changed to determine what a user was attempting to read in those documents. To accomplish this, the entire document directory was cleaned by eliminating any special characters, turning the text to lowercase, replacing all URLs and images, deleting stop words, and finally lemmatizing and tokenizing all keywords. The 30 highest occurring words were extracted to make sense of what a user was reading. After performing all such actions, the initial table was updated and finally exported as an excel file to be linked to Tableau for visualization. The final table contains all the interactions performed by an analyst in a certain time segment throughout their analysis; this data assists us in identifying a pattern of how an analyst thinks and behaves.

After linking the excel file with the filtered data to Tableau, we undergo further data exploration. We experiment with visualization charts to find what represents the dataset most appropriately. Since the data is highly text-based, we followed Schneiderman's mantra and decided to give the overview first and then zoom in to provide

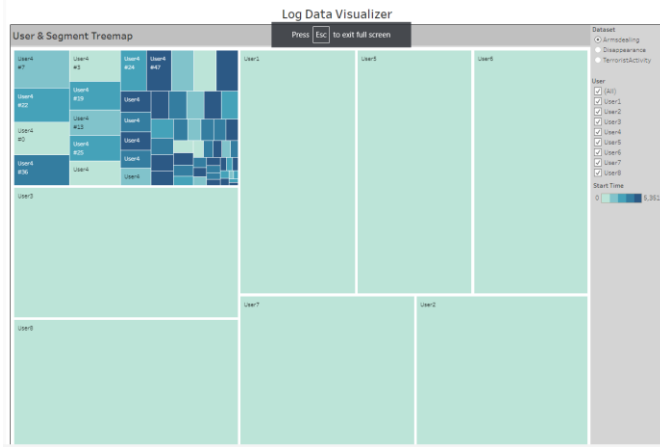


Fig. 1. Upper Level of Log Data Visualizer

filtered content and then at the end give the user the control to see details on demand for a particular activity or time segment. Therefore we picked Treemap as our primary visualization as that seemed to best present all users at once and the segments of each of their log data in an overview, and you can then click on a user to zoom into it, and see all the segments. Clicking over a segment gives you the tooltip that again gives a small overview of some of the parameters. It also gives you the option to go a level down and get more details about that segment. At the lower level, where you will find the details, you are welcomed with a dashboard divided into 4 views. Each view shows different aspects of the data.

### 2.3 The Interface

We constructed a Treemap, at the upper level of the design (Fig. 1), to visualize log summaries of 8 Analysts that are interacting with multiple documents within 3 datasets, namely Arms dealing, Terrorist Activity, and Disappearance. Since it is temporal data, the conventional approach for visualization would be a timeline representation but we have prioritized the duration of the segment over the chronological order in which the interaction events occurred. When we were trying to understand the segmented log data, we found that the length of segments also holds significance. We believe that if an analyst spent a longer time in a segment, he has most likely done more activities. Looking into it first would allow the person trying to understand the analysts' rationale, to grasp the primary interactions of the analyst. This can save time if they are able to identify any pattern by studying only the top 5 or 10 longest segments. The smaller segments will have fewer activities done within them and thus can only be visited if trying to get a much more clear idea about a certain pattern or fact. For a quick dirty analysis, visiting the longer segments should be helpful.

This leads us to choose Treemaps, which allows us to show a descriptive outline view of all analysts working on a particular dataset. Treemaps are a great way of showing information using area encoding and we have used the area to show the time duration of both the analysts and the segments. Therefore if in the overview having selected all the analysts, the rectangles are uneven then that means we have different amounts of log data for each analyst. If we have the log data of the same length of duration for each analyst, we will have rectangles of the same size. Thus, in one glance, with the help of the pre-attentive processing of our brain, we can determine how much log data we have for each analyst and if we require more. The boxed nested visual of the Treemap also helps us perceive the hierarchal nature of the given log data where there is an analyst's data, and within that data there are segments, and within those segments are interactions that need to be studied. The first two levels can be shown within a Treemap, and for the third level, we drill

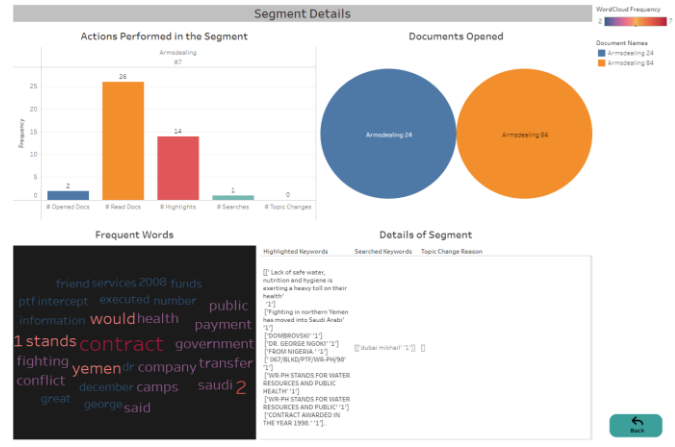


Fig. 2. Lower Level of Log Data Visualizer

down to another dashboard where we see detailed information on the analyst's activity during that segment.

Albeit we prioritized the length of duration, it does not mean we disregarded the time-stamped component of the log data. We have incorporated color encoding within the rectangles of the Treemap to show the chronological order in which the interactions occurred. The color gradient goes from light to dark based on the start time of the segment. The lighter shades denote the events that happened at the beginning and as the shades get darker, they denote the events that happened later within the total given duration of the analyst's data. This gives the broader context of when a particular action took place, so we can also investigate the segments either preceding or following it, to know more about the cause or effect of that action, respectively.

In our application, we have used the general name 'User' instead of 'Analyst', within Treemaps, to emphasize the fact that this can be used as a general framework to visualize summaries of other similar log data, even though our tool at present is only showing visual summaries of 3 datasets, that include system event logs from analysis tools which the analysts (such as Mark) use. For the purpose of distinguishing the 'User using our visualization' from the 'User of the analysis tool', we will continue to denote the latter as 'Analyst' throughout this paper.

The first step, while navigating the application, would be to choose a dataset among the three datasets in the right-side panel. Then we can either select to view all the analysts' data, of that dataset, at once or we can select only one or a few analysts, from the list in the right-side panel. The latter will show all the segments of only the selected analyst(s). Then we can click on a segment to be greeted with a tooltip that gives us the overview points of that particular segment. This can be quickly accessed as well by just hovering your pointer on the segment. When you click a segment you get the option to view the segment details. Click that and you will find yourself looking at a new dashboard (Fig. 2) divided into 4 parts. All these parts are simple visual representations that most users are familiar with or won't have difficulty understanding it. This is to remove the visual complexity at this lower level as the textual log data is itself detailed.

The Top-Left part contains the bar chart that shows the frequency of the actions that are performed in that segment. These actions include Opened Documents, Read Documents, Highlights, Searches, and Topic Changes. Looking at this view lets you know at a glance which activity was performed the most and least. If that activity is something that you want to examine in that segment you can continue to study other charts or you can go back to the overview and check out the other segment. The Top-Right part contains a

bubble chart showing the documents that were opened in this segment. The size of the bubbles correlates to the frequency of the document opening. This helps the user of the visualization know at once which one or two documents were most opened and analyzed in this segment. If they want to know more about what the analyst did with those specific documents, they can stay at this level and explore. Smaller bubbles may not have the name of a document written in them, therefore we have provided a legend on the right-side panel. You can also hover over all the bubbles to see the name of the document along with the number of times that document was opened.

In the Bottom-Left corner, you will find a Word Cloud. This shows the frequency of some words that are present in the documents of this segment. These are the Top 30 words that were filtered out in our data processing and can hold significance in identifying the general topics that are common in those documents. This chart can also be used to detect any pattern or trend, by observing the highest frequency words among different segments of a particular analyst. You can also find the legend, in the right-side panel, for the color encoding done in the Word Cloud. Hovering interactivity is also present in this chart where hovering over a keyword shows the tooltip containing the exact frequency of the keyword. Lastly, in the Bottom-Right corner, there is a table that shows the most detailed view. In this, you will find the actual textual log data which can be investigated to verify facts after having been done with the quick analysis using all the other charts. This table consists of columns such as Highlighted Keywords, Searched Keywords, and Topic Change Reasons. Each column lists the important keywords that are related to that activity and can be read in detail to understand the full context of the analysis done in that segment. You can hover over a cell of the table to show a tooltip that shows the same data in a more readable format. Due to space constraints of the current design, sometimes not all of the data is visible in the table cell. Therefore we have added this workaround for now. In future iterations of the design, we plan to add more levels, so as to resolve the space issues. The Back button at the bottom of the page leads you back to the overview page containing the Treemap.

### 3 EVALUATION

A necessary evaluation of our visualization is to determine the effectiveness of the information conveyed by it, and how easily and accurately was it received by the viewer. We wanted to make sure we have kept the visualization easy to understand and navigate but at the same time, it does not lose showcasing critical information in the process. To perform such an evaluation of our visual design, we conducted user studies where we showed the visualizations to a group of people – 5 males and 3 females with ages ranging from 22 to 26. This was a peer review, so the participant pool was all non-analyst. We are yet to take an expert review from a professional analyst, at the time of writing this paper. Considering the participants of the user study were all non-analysts, we kept the tasks quite straightforward. We did provide them with the background of what the data and visualizations are all about, before giving them the control to explore the data by themselves within the visualization, to achieve the analysis tasks.

We tested them with 4 tasks: 2 of those tasks required them to find out a specific value of data, by navigating through the visualization, and the other 2 tasks required them to find a pattern or trend within the dataset. This way we measured if they were able to infer the information in the correct context and to what degree was it helpful. We also measured the time taken to complete each of these tasks to understand whether it was difficult to navigate through the visualization. Each participant was provided with the same set of tasks, in order for us to compare the time taken for each task at the end.

The tasks to find specific data values were:

1. For the dataset of Arms Dealing, which keyword(s) was given the most importance by Analyst 4 within the top 3 longest segments?
2. For the dataset of Disappearance, which document(s) was of most interest to Analyst 2 within the top 3 longest segments?

The tasks to find a pattern or trend were:

1. How would you categorize the most significant keywords that came up in the documents read by Analyst 1 in the top 5 longest segments, from the Disappearance dataset?
2. On average, how many documents were opened, in the top 5 longest duration segments, by Analyst 3 from the Terrorist Activity dataset?

We also conducted a short questionnaire after the tasks to ask relevant questions so as to get additional feedback. This included questions such as:

1. Were you able to grasp the conveyed background information in the given amount of time?
2. How easy was it to navigate through the visualization?
3. Which visual elements stood out the most and helped infer from the visualization?
4. Which visual elements made it difficult to infer the visualization?

### 4 DISCUSSION AND FUTURE WORK

As we progressed with our implementation of Treemaps, we came across quite a few limitations that we overlooked initially. We found out that it is only possible to incorporate a limited number of analysts at once in the treemap or else it will make the overview of all analysts in the Treemap crowded and look visually very complex and unappealing. The same goes for the segments view inside the rectangles. If there are a lot of segments then it will be very difficult to hover or click on the segments plus it would be difficult to make sense of it all.

We also came to know that with our current design, it will not be possible to provide the option of comparing 2 or more analysts' behavior. Our initial use case scenario (as described in section 2.1.1) did not have the requirement to compare the behaviors of multiple analysts. This is not a shortcoming, as our application serves as the solution to the problem statement we designed it for, but this is more like a feature that we can integrate as we expand and improve our design.

Another shortcoming that we acknowledge is that we did not evaluate our design with a larger group of participants. This would have helped us get more feedback and helped us improve and optimize our design. We also were not able to take feedback from analysts to know what more they expect from such an application. Also, we are required to pre-process the data before providing it as input to our application. The application does not take raw log data as input to both pre-process and visualize it. We plan to make the pre-processing of the data automatic and integrated into the application in the future.

The feedback from the user study was very useful. The participants gave constructive criticism where they pointed out areas of design that they struggled with. One of them said, "Every time I had to go into a segment to get any sort of information. The overview design (Treemap) does not convey any useful information associated with the understanding of human data exploration. And there is no side-by-side view of segments. No option to toggle between segments.

Each time I need to go back to the overview and then into the other segment.”

Another one mentioned that “It is difficult to exactly point out the chronological order of the segments through just the color gradient. For instance, if I want to find what happened in the subsequent segment I had to hover over all the segments with very similar shades of color and check the segment numbers in order to find the next segment. The use of color gradient within each rectangle showing the chronological order is good for quick overview but time-consuming when doing a detailed analysis.”

We noticed that the first task usually took more time to do for each participant, and subsequent tasks took comparatively less time. When asked about this in the questionnaire, their reply was unanimous that it took them a while to get hang of the visualization. And without the initial orientation where we gave them the background knowledge of what the design is about, they would have struggled even more. So we came to know that the design is not very intuitive initially. It becomes intuitive with subsequent usage. If we add help links or annotations in the future, then maybe that would guide the viewer and make the navigation smoother.

The participants also pointed out the advantages of our design. Two of them appreciated how we have divided the log data into different view levels of complexity. This way the viewer does not get overwhelmed at once. Another one appreciated that since we have converted most of the textual log data into visual designs – it was less time consuming to navigate through the data and make inferences from it.

## 5 CONCLUSION

We designed an application that helps viewers to do an exploratory data analysis via a visual summary of textual log data consisting of various interactions, to understand the thought process and behavior of the analysts. Our design contains two levels – upper level and lower level. The upper level provides the overview with zoom and filter functionality. The lower level provides the details which can be accessed on demand. Thus our visualization follows Ben Schneiderman’s mantra for visual information seeking. The upper level contains Treemap and the lower level contains 4 visuals – a bar chart, bubble chart, word cloud, and table with details of actions. Treemap shows analysts and the segments of their log data. The size encoding within the treemap correlates to the length of duration of the data, and the color encoding is based on the start times of the segments. The bar chart shows the frequency of actions performed within the selected segment. The bubble chart presents the documents that were opened in the given segment and their frequency. Word cloud emphasizes the most common words in the documents and the table shows the filtered data of interactions in text format.

Having gone through this whole design process, we found out both good and bad things about Treemaps. Since we were committed to the design halfway through, we tried our best to leverage all the good points of it. For the bad points, we have decided to create an elaborate design of which the current design will be a part of so we can still take advantage of the benefits of it while having the limitations compensated by other visual representations. In future work, we plan to have multiple levels and have the overview contain more relevant information. We will also have side by side view for comparison of segments and we will have annotations to guide navigation. We are hoping that this future elaborate design can then widen the scope of the application. This visual summaries tool could show the interaction log data of, for instance, an in-house software in a company. Knowing how the users have interacted with the

software can help the developers optimize their software to provide more intuitive access to the information within it.

## 6 ADDITIONAL LINKS

### 6.1 Link to the Live Demo

<https://info-viz-project.w3spaces.com/>

### 6.2 Link to the Code Repository

[https://drive.google.com/drive/folders/1NxWe7\\_WWmsNdu\\_r4J6dYFL\\_JDf9QoxHwX?usp=sharing](https://drive.google.com/drive/folders/1NxWe7_WWmsNdu_r4J6dYFL_JDf9QoxHwX?usp=sharing)

### 6.3 Link to the Demo Reel

[https://drive.google.com/file/d/1Qj5IV\\_JmMJ47Kh4YCK8-Eg99OfdPLYVT/view?usp=sharing](https://drive.google.com/file/d/1Qj5IV_JmMJ47Kh4YCK8-Eg99OfdPLYVT/view?usp=sharing)

## ACKNOWLEDGMENTS

We would like to thank Dr. Eric Ragan and Ferby Cremer for their helpful feedback and support throughout the designing of this visual application.

## REFERENCES

- [1] Moore, A. V., & Purchase, H. (2011). On the role of design in information visualization. *Information Visualization*, 10(4), 356–371.
- [2] S. K. Card and J. Mackinlay, "The structure of the information visualization design space," *Proceedings of VIZ '97: Visualization Conference, Information Visualization Symposium and Parallel Rendering Symposium*, 1997, pp. 92-99,
- [3] Jeffrey Heer, Stuart K. Card, and James A. Landay. 2005. Prefuse: a toolkit for interactive information visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*. Association for Computing Machinery, New York, NY, USA, 421–430.
- [4] Michelle Q. Wang Baldonado, Allison Woodruff, and Allan Kuchinsky. 2000. Guidelines for using multiple views in information visualization. In *Proceedings of the working conference on Advanced visual interfaces (AVI '00)*. Association for Computing Machinery, New York, NY, USA, 110–119.
- [5] F. Heimerl, S. Lohmann, S. Lange and T. Ertl, "Word Cloud Explorer: Text Analytics Based on Word Clouds," 2014 47th Hawaii International Conference on System Sciences, 2014, pp. 1833-1842
- [6] Joshi A. P, Patel B. V. Data Preprocessing: The Techniques for Preparing Clean and Quality Data for Data Analytics Process. *Orient.J. Comp. Sci. and Technol*; 13(2,3).
- [7] Z. Shen, J. Wei, N. Sundaresan and K. -L. Ma, "Visual analysis of massive web session data," *IEEE Symposium on Large Data Analysis and Visualization (LDAV)*, 2012, pp. 65-72
- [8] H. Lam, D. Russell, D. Tang and T. Munzner, "Session Viewer: Visual Exploratory Analysis of Web Session Logs," 2007 IEEE Symposium on Visual Analytics Science and Technology, 2007, pp. 147- 154
- [9] J. Mackinlay, P. Hanrahan and C. Stolte, "Show Me: Automatic Presentation for Visual Analysis," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1137-1144, Nov.-Dec. 2007
- [10] Y. Guo, et al., "Survey on Visual Analysis of Event Sequence Data" in *IEEE Transactions on Visualization & Computer Graphics*, vol. 28, no. 12, pp. 5091-5112, 2022.
- [11] Padmaja, S. and Dr. Ananthi Sheshasaayee. "Clustering of User Behaviour based on Web Log data using Improved K-Means Clustering Algorithm." (2016).
- [12] Kalmukov, Yordan. (2021). Using word clouds for fast identification of papers' subject domain and reviewers' competences.

- [13] Toshiyuki Asahi, David Turo, Ben Shneiderman, (1995) Using Treemaps to Visualize the Analytic Hierarchy Process. *Information Systems Research* 6(4):357-375.
- [14] Shneiderman, Ben, and Catherine Plaisant. "Treemaps for space-constrained visualization of hierarchies." (1998).
- [15] Y. Tu and H. -W. Shen, "Visualizing Changes of Hierarchical Data using Treemaps," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1286-1293, Nov.-Dec. 2007
- [16] J. J. Van Wijk and H. Van de Wetering, "Cushion treemaps: visualization of hierarchical information," *Proceedings 1999 IEEE Symposium on Information Visualization (InfoVis'99)*, 1999, pp. 73-78, doi: 10.1109/INFVIS.1999.801860.
- [17] C. Bertero, M. Roy, C. Sauvnaud and G. Tredan, "Experience Report: Log Mining Using Natural Language Processing and Application to Anomaly Detection," 2017 IEEE (Institute of Electrical and Electronics Engineers) 28th International Symposium on Software Reliability Engineering (ISSRE), 2017, pp. 351-360
- [18] M. Gray, A. Badre and M. Guzdial, "Visualizing usability log data," *Proceedings IEEE (Institute of Electrical and Electronics Engineers) Symposium on Information Visualization '96*, 1996, pp. 93-98
- [19] Bhaya, Wesam. (2017). Review of Data Preprocessing Techniques in Data Mining. *Journal of Engineering and Applied Sciences*. 12. 4102-4107. 10.3923/jeasci.2017.4102.4107
- [20] H. Guo, S. R. Gomez, C. Ziemkiewicz and D. H. Laidlaw, "A Case Study Using Visualization Interaction Logs and Insight Metrics to Understand How Analysts Arrive at Insights," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 51-60, 31 Jan. 2016