TextVista: NLP-Enriched Time-Series Text Data Visualizations

Fateme Rajabiyazdi
Carleton University
Bruyere Research Institute
Ottawa, Canada
fateme.rajabiyazdi@carleton.ca

Shri Harini Ramesh
Carleton University
Ottawa, Canada
shrihariniramesh@cmail.carleton.ca

Beck Langstone
Carleton University
Ottawa, Canada
becklangstone@cmail.carleton.ca

Daniil Kulik
Carleton University
Ottawa, Canada
daniilkulik@cmail.carleton.ca

Justin Pontalba
Novacene.ai
Ottawa, Canada
jpontalba@novacene.ai

Abstract

There is a vast amount of unstructured text data generated every day analyzing and making sense of these text-based datasets is a complex, cumbersome task. The existing visualization tools that analyze text data, leveraging Natural Language Processing (NLP) techniques, are often tailored for structured text-based data. They also fail to support reading, a crucial analysis task to validate the output of NLP techniques. We designed and developed TextVista, an NLP-enriched visualization tool that supports analysts during their analysis of unstructured text with temporal references. Our tool combines techniques including clustering, sentiment analysis, and threat detection with three views that visualize high-level patterns in the data to encourage reading. We report on TextVista's iterative design process, which included a focus group to distill design requirements, a think-aloud interview study with data analysts to understand their impressions of the tool, and a diary study to assess its long-term usage. Through this process, we identified how TextVista supported the analysis of unstructured text with temporal references using NLP techniques and fostered methods to promote reading in situ. TextVista also encouraged serendipity when analyzing data via its question-focused overviews and flexible avenues to explore data.

Keywords: Data Analysis, Reasoning, Problem Solving, Decision Making, Qualitative Evaluation, Text/Document Data

ACM Reference Format:
Fateme Rajabiyazdi, Shri Harini Ramesh, Beck Langstone, Daniil Kulik, and Justin Pontalba. 2024. TextVista: NLP-Enriched Time-Series Text Data Visualizations. In *Proceedings of Graphics Interface (GI '24)*. ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

GI '24, June 03–06, 2024, Halifax, NS

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

There has been a surge in text-based data created daily, much of which originates from sources like medical records, news articles, and social media, all with a temporal nature. Analyzing such data is crucial to understanding the thematic and temporal trends within the data, unleashing new insights, and making informed decisions; however, as this text data is unstructured, it is impractical to manually process and analyze such data [34]. While machine-learning approaches have been successful in modeling human languages, they often lack the syntactic and semantic understanding, as well as the domain expertise needed to interpret unstructured data [36]. Natural Language Processing (NLP) techniques can transform unstructured text into structured data, addressing language ambiguity challenges and enhancing the utility of data for applications such as text analytics.

Although we can employ NLP techniques to examine unstructured text data, the complexities of data can pose difficulties when identifying unexplored questions or questions we haven't realized are essential to ask. Data visualization is a promising solution that can support analysts in uncovering *unknown unknowns*, i.e., questions that are only uncovered while visually browsing data, finding underlying patterns and relationships that otherwise would have remained hidden, and revealing patterns and discovering valuable insights that might have otherwise remained concealed [8, 28]. Data analysts often search for answers to specific questions about the relationships between segments of their data, the social context of language usage, and so on. While NLP techniques can suggest potential patterns in data and help analysts choose which data to consult further, it is up to the analyst to navigate the data, read its source, and develop deeper insights. Visually displaying the NLP outputs can support analysts in validating the NLP technique outputs and allow analysts to develop a deeper sense of trust in the system. Thus, we need to develop visualization tools that provide explicit visual links between an NLP technique's output and the source data used to derive said output, tools that can support data analysts in quickly and easily determining the audience, context, and purpose of the underlying source data.

Many visualization tools have leveraged NLP techniques to support the exploration of unstructured text data [59, 60, 70, 74], while few have integrated text with other structured fields such as time [10, 18, 22, 61]. However, these tools often cater exclusively to specific goals for domain experts, such as medical chart review for healthcare providers [60, 61]; social media research for social scientists [10]; reading, storytelling, fact-checking for journalists [22, 59, 70]. These tools also do not support data analysts in analyzing their text data beyond keyword extraction, summarizing, and finding patterns or clusters [18, 57, 74]. Most unstructured text-based data have temporal references. However, to our knowledge, there is a lack of tools for data analysts to visually analyze unstructured text with temporal references, aiding in finding patterns, relationships, entities, phrases, sentiments, toxicity, and anomalies while having direct access to the raw data.

We designed, developed, and evaluated *TextVista*, an NLP-enriched text data visualization tool that supports analyzing unstructured text with temporal references by visualizing source data and high-level NLP patterns. Inspired by prior research [59, 60, 62], TextVista supports transitions between high-level summaries by visually displaying the patterns alongside the source data to promote reading. To understand how NLP techniques and visualizations could support the analysis of unstructured text with temporal references, we conducted focus groups with data analysts who had different analysis goals. The focus group transcripts were distilled into challenges, analysis-based questions, and tasks that informed TextVista's design. We then evaluated TextVista using a think-aloud interview study to gain insights into analysts' thought processes as they interacted with TextVista. After making aesthetic changes to the interface suggested by interview study participants, we conducted a diary study with two analysts who used TextVista to analyze their own datasets. These two evaluations demonstrated that combining multiple NLP techniques with visualizations can facilitate the analysis of unstructured text-based data. By supporting transitions from high-level patterns to the source text, TextVista not only supported reading, a critical step during text analysis, but also contextualized the patterns identified by the NLP techniques. Our contributions are:

1. A set of design requirements for visualizing unstructured text data with temporal references enriched by NLP.
2. *TextVista*, a data visualization tool for analyzing unstructured text with temporal references, supporting data analysts in finding patterns, relationships, entities, phrases, sentiments, and anomalies with direct access to the data.
3. A longitudinal diary study to evaluate how data analysts can integrate TextVista into their daily workflow outside of a controlled study environment.
4. A set of guidelines on how to best design NLP-enriched text-based data visualizations.

2 Related Work

We review visualizations that support the analysis of unstructured text, as well as visualizations that facilitate reading while sensemaking and support serendipitous text analysis.

2.1 Visualizations Analyzing Unstructured Text

Text analysis is a common task that analysts from many fields perform; however, it can be challenging to visualize text-based data due to its lack of structure. To address this lack of structure, some visualizations have integrated structured fields with unstructured text [41]. For example, some visualizations leverage temporal references [11, 22, 53, 63, 71]. TIARA [71] helps visualize text documents by summarizing them into a set of topics and showing the evolution of these topics over time. Timelinecurator [22] is a timeline creation tool that journalists can use to automatically extract event data from unstructured text documents with temporal references. Parallel tag clouds [11] visualize the differences among subsets of the data over time using word clouds. Other researchers have also used temporal references in unstructured data to visualize themes, entities, and frequencies of certain words over time [42, 53, 63].

Other researchers have leveraged multiple structured fields such as dates, locations, or author names [18, 57] to visually display text data. ViTA-SSD is a system that allows users to choose and visualize the distribution of various structured fields based on a color-coded matrix coupled with histograms [57]. TextTile is a text visualization tool that applies operations performed on categorical, ordinal, and quantitative structured data to unstructured data using tile-like visualization segments [18]. However, these tools do not support analysts in analyzing their text data beyond keyword extraction, summarizing, and finding patterns or clusters.

Additionally, some visualizations have leveraged NLP techniques such as Named Entity Recognition (NER), clustering, labelling, and sentiment analysis to support the exploration of unstructured text data [21, 50, 58]. Jigsaw supported analysis through interactive, modular visualizations that were powered by NER [58]. Although Jigsaw supported the exploration of the text through various views, it did not provide overviews of the themes in the underlying text, so users had to have a 'target' in mind before exploring their data. ConceptVector [50] used clustering to visualize clusters in semantic space. While ConceptVector enabled users to directly manipulate parts of a visualized cluster to refine concepts to their needs, it offered limited support for sensemaking or reading source text within a visualization. While NLP techniques can support the analysis of unstructured text, they will not replace the human judgement needed for fine-grained, qualitative forms of analysis [4, 21]. Thus, to our knowledge, there is a lack of NLP-enriched tools for data analysts with diverse goals to visually analyze unstructured text only with temporal references.

2.2 Visualizations for Reading While Sensemaking and Serendipitous Text Analysis

While analyzing text, analysts search for insights to answer questions or hypotheses they have. During this process, reading (or skimming) is a key part of an analyst's workflow [27, 37, 59, 73]. Several tools have been designed to support reading while interacting with visualizations. For example, Storifier, a text-based visualization tool for journalists [59], enabled users to drill down from a corpus-level overview to the original source text. The ability to explore a large corpus of text-based data at macro and micro scales provided users with a supportive interface to efficiently explore narratives. As another example, to support the reviewing of medical notes, Chartwalk [61] used NLP techniques to extract an overview of a patient's chart and create a timeline of events. While some of these visualization systems supported reading, they were not built to encourage high-level patterns and serendipitous discoveries while reading.

Previous research has investigated models and visualizations designed to support serendipity [3, 7, 44, 46, 64]. McCay-Peet and Toms proposed a model of serendipity in knowledge work [44]. During this process, users search for something interesting [7], of value, or that sparks curiosity [46]. While searching, an unexpected trigger is encountered that causes a bisociation (i.e., an association between two previously unconnected pieces of data) and leads to an unexpected solution to a problem that was not initially intended to be solved [16, 44]. Although these 'serendipitous triggers' can play a powerful role in the process of analysis, little is known about what makes an effective serendipitous trigger.

Research has also examined the precipitating conditions of serendipity, i.e., analytical active learning [19, 46]. By ensuring that the user is at the center of interaction when using exploratory visualization systems and encouraging them to make discoveries in different ways, one can create an environment that promotes serendipity [7]. Few visualization tools have explored designing for serendipity. The Bohemian Bookshelf [64] was a visualization tool that helped readers discover books from a library's e-book collection by playfully exploring cover art colour, a timeline view, author name, and book thickness. Providing views with different foci enabled users to change their focus and thus enhanced serendipity. Serendip [3] was a re-orderable matrix text visualization tool that used topic modelling techniques to identify the topics. The matrix order could be set by the distribution of a user-selected topic, similarity, or query. Both Serendip and Bohemian Bookshelf used visualizations to create environments conducive to serendipity; however, these systems do not support reading and analysis of unstructured text.

3 Design Requirements and Prototype

To understand how data analysts investigate unstructured text datasets and gather a list of questions for their investigation, we employed an iterative, user-centered design process [24, 47] and conducted a focus group.

3.1 Participants

We recruited 4 data analysts who analyzed text-based datasets as part of their work: P1 - a 47-year-old male CEO of an AI Analytics company with 2 years of experience; P2 - a 51-year-old male Data Engineer and Analyst with 5 years of experience; P3 - a 44-year-old male Analytics Engineering Consultant with 7 years of experience; and P4 - a 41-year-old male Professor and Data Analyst with 6 years of experience

3.2 Methodology

Following the nested model for visualization design and validation [45] and a framework for requirement gathering focus groups [33, 35], we conducted a focus group with two sessions. The outcomes from the first session informed the topics and low-fidelity prototypes discussed in the second session. The focus group was chosen for its ability to foster idea exchange among participants [17, 68]. Each session, conducted remotely in Microsoft Teams, included a demographic survey and lasted 3 hours, audio/video recorded.

During the first focus group session, we asked questions and facilitated discussions about participants' current approaches and tools for analyzing text-based data and their goals and challenges. Following this, we conducted grounded theory analysis of the transcripts using open coding in NVivo software. Each transcript was independently coded by two researchers, both are master's students in human-computer interaction. These researchers underwent qualitative research training from an experienced qualitative analyst. Researchers discussed and developed a codebook, resolving discrepancies during the coding. Emerging themes and patterns from the codes were identified and discussed. From this analysis, we distilled our initial set of design requirements and analysis tasks, which informed the creation of a series of low-fidelity paper prototypes (See Supplementary Materials).

In the second focus group session, we presented the design requirements and visualization tasks and conducted a walkthrough of our low-fidelity paper prototypes, which helped avoid threats to the validity of the prototypes [45]. We used a Miro board to share prototypes and gather feedback. Then, we analyzed the transcripts from the second focus group session using grounded theory following a similar approach to the analysis conducted in the first focus group session [12].

3.3 Results

The results from the two sessions led to the identification and validation of the design requirements and the tasks that data analysts need to conduct to analyze their text datasets.

3.3.1 Transitions to Support Reading.
All participants spoke about how once they had searched for or identified a pattern or trend, they would drill down to read the associated unstructured text, *"The more I read through all of that [text], I'm starting to get an understanding of the structure."* (P2). Participants also desired assistance in identifying new ways to look at their data,*"a semantic helper that pre-processes it and kind of tells me, what would be the different angles you could look at that data from? What … different entities you could extract from that data?"* (P2). Thus, it is necessary for our tool to **support transitions between different levels of detail to facilitate reading within views (DR1)**.

3.3.2 Timescale Explorations.
Participants showed interest in identifying changes in data over time, especially in analyzing relationships between people or groups, as well as between people and topics, and understanding the sentiment's role in these relationships. While they employ NLP techniques to quantify sentiment in text segments [29, 69], participants struggled to understand how and why sentiment changed over time. e.g., *"Am I going to take a snapshot at different points in time of the same dataset, or am I going to merge it all together? ... How will I measure it to help me understand how it has evolved since the last time? It becomes complicated."* (P3). Additionally, participants noted challenges in working with datasets of varying sizes, granularity levels, and time scales. They emphasized the need for dynamically adjusting and revealing different data dimensions over time. e.g., *"I really think of [flexible timescales] as a microscope ... If you have a microscope that allows you to see the atom, great, the atom won't change because of the microscope ... it's the microscope that changes the object of research itself"* (P4). Thus, our tool needs to **use flexible timescales to support exploring the temporal dimension of data (DR2)**.

3.3.3 Triggering Serendipitous Discovery.
Participants usually began their analysis motivated by a research goal or question, such as the dynamics between people or the relationships between people and the topics discussed in the data. They mentioned being triggered by unexpected information, like a person or place, or discovering a pattern that led to new questions about their data. e.g., *"we think we're going after something and then we all of a sudden get, you know, punched in the face with something that we weren't expecting so I think that there's an element of, how do we capture that?"* (P1). Participants also discussed the challenge of finding 'unknown unknowns' or unexpected things that they were not aware of and did not understand [54]. For example, P1 shared an instance of discovering an unknown unknown, stating *"You know, like when [my team was] doing analysis on Ukraine and Russia a month ago, we saw that Syria popped up and we're like, oh, wait a minute, we had no idea. So you follow the node, and then you start to understand."* These discussions highlighted how topics, entities, sentiments, and unexpected words could serve as triggers for serendipitous insights. Thus, our tool should use NLP techniques to **identify data that could be used to create and visualize potential triggers for serendipity (DR3)**.

View Tasks \ Corresponding Low-Level Tasks	Sort	Filter	Retrieve Value	Find Anomalies	Characterize Distribution
Text Frequency View Tasks					
T1. What terms are most frequently used? When?	X	X	X	X	X
T2. What term combinations are frequently used? When?	X	X	X	X	X
T3. What is the text data source?	X	X	X		
Sentiment and Semantic Clustering View Tasks					
T4. What are the topics being discussed? When?	X	X	X	X	
T5. How are topic clusters distributed over time?	X	X		X	X
T6. What is the sentiment of the text? When?	X	X		X	
T7. What is the sentiment towards a topic? When?	X	X		X	
T8. What is the toxicity of the text? When?	X	X		X	
T9. What is the toxicity score towards a topic? When?	X	X		X	
T10. What is the text data source?	X	X	X		
Entity Dial View Tasks					
T11. What entities are being discussed? When?	X	X	X	X	
T12. What are the topics being discussed? When?	X	X	X	X	
T13. What is the sentiment of the text? When	X	X	X	X	
T14. What is the toxicity of the text? When?	X	X	X	X	
T15. What is the text source?	X	X	X		

Table 1. The low-level tasks supported by Text Frequency, Sentiment and Semantic Clustering, and Entity Dial views.

3.3.4 Exploration via Personalization.
Participants were from various fields with diverse goals and questions about their data, but they all sought insight. They were eager to uncover the story hidden within the intricate layers of information. They mentioned, *"I'm sure that, like, I have all the data that I need to tell a really good story...it's really what's hidden behind, like a thick layer of the media biases, all of the opinion pieces … that hides the skeleton of what the story is"*. This poses a challenge when determining what to show within a visualization or view. Participants expressed frustration with the lack of suitable tools or methods to aid their exploration. *"It's just annoying because I just don't have the methodology or whatever the helper to help me discover"*. The user should be encouraged to quickly and flexibly explore different aspects of their data by personalizing the view in ways that are meaningful to their analysis [7]. Thus, our tool should **support exploration via personalization (DR4)**.

3.4 Visualization Tasks

Though our participants' goals differed, they sought answers to similar questions during their analyses. We identified 15 questions that were common across participants and their corresponding tasks to answer these questions. To determine the best way to develop a visualization system to address the needs and tasks, we aligned these questions, we aligned these questions to Amar, Eagan, and Stasko's taxonomy [5]. We grouped these questions into three categories (views) based on their similarities, as shown in Table 1. The first view corresponds to tasks that explore terms used in the text and their frequency (T1, T2). The second view is for exploring the sentiment (T6) and toxicity score (T8) of text entries. This also includes topics being discussed (T4), how the topic clusters change over time (T5), and the sentiment (T7) and toxicity (T9) of these topics and how they change over time. The third view is for identifying entities (e.g., players, organizations) (T11). It includes tasks such as exploring topics in relation to entities (T12) and analyzing the sentiment (T13) and toxicity (T14) of the texts in relation to the identified entities and topics. All these three views include direct access to the source of data to read the input text (T3, T10, T15).

In order to meet DR1, we ensured all views began with an overview. Participants can click on parts of the visualization that pique their interest in the overview, which will then bring the corresponding text entry. This allows participants to read the source text easily. To fulfill DR2, we introduced two ways to filter time - by time range (selecting the start and end points) and by aggregation interval (years, quarters, months, weeks, days, etc.). To meet DR3, we added playfulness to our design, allowing users to interact with data by topics, entities, and sentiments across all three views. We supported exploratory search, providing analysts with agency to determine what should be visualized. Finally, DR4 was fulfilled by providing personalization options, such as choosing encodings (color, shape, line type) and the utility of adding or removing columns in the tabular reading pane.

4 TextVista

TextVista [1] was implemented using React.js and D3.js, inputting timestamped, text-based, CSV-formatted files that were processed via NLP techniques. In response to the three categories of tasks, we designed and developed TextVista with three views (i.e., the *Text Frequency View*, the *Sentiment and Semantic Clustering View*, and the *Entity Dial View*). All views had consistent filtering options, interactions, layouts, and colourblind-friendly colour schemes [65]. Below, we describe the NLP techniques and processing performed and provide a system example workflow. [2]

4.1 NLP Techniques

TextVista can visualize any data that contains unstructured text with temporal references. Our data input contained artifacts such as hashtags and emojis; thus, before the NLP techniques were applied, the data was cleaned by removing all artifacts and translating them into English if necessary. Based on the tasks identified during focus group sessions, we then employed the below NLP techniques to process the data to respond to the tasks (Table 1). We also identified the text data source (T3, T10, T15) by identifying the corresponding variable before applying NLP techniques.

- **Term Frequency:** To address (T1, T2), the Term Frequency technique [32] was used to output the frequency and indexed location of extracted unique terms and n-grams (i.e., phrases).
- **Clustering and Cluster Labelling:** To enable users to filter by topics (T4, T12), the text data was transformed into a vector embedding space using Sentence Bidirectional Encoder Representations from Transformers (SBERT) [51] and the vectors were organized into clusters using Hierarchical Agglomerative Clustering (HAC) [13]. Initially, each text entry formed its own cluster, and clusters were merged iteratively based on similarity. The central topic was identified as the node or cluster member that achieved the highest importance score in a graph network using the PageRank algorithm [49]. Clusters were then given descriptive, human-readable labels.
- **Principal Component Analysis (PCA):** To address (T5), PCA was performed to simplify the vector representations into two dimensions to enable the visualization of the semantic space of the text [32, 40]. We experimented with 3 different dimensionality reduction methods: PCA, t-distributed Stochastic Neighbor Embedding (t-SNE), and Uniform Manifold Approximation and Projection (UMAP). When we plotted the reduced data in a 2D space, the PCA gave us the most visually appropriate results.
- **Sentiment Analysis:** To assist users in answering questions about the sentiment of their text (T6, T7, T13), pre-trained BERT models [15, 43] categorized each cluster as either *positive* or *negative*. We only considered positive and negative sentiments as the BERT model performs better in binary sentiment classification [25, 72].
- **Toxicity score:** To address (T8, T9, T14), a toxicity score was used. One source of unstructured text is online sources, where threats and toxicity are rampant, Google Perspective API [3] [39] was used to classify the cluster (topics) and text entries as *threats* or *not threats*. We visualized toxicity scores and sentiment analysis together because varying sentiments may contribute to increased conversation toxicity [6, 52]. Combining them may reveal more patterns and potentially trigger serendipitous discoveries.
- **Named Entity Recognition (NER):** To answer (T11), NER was performed using contextual string embeddings combined with a bi-directional Long Short-Term Memory network with a Conditional Random Fields layer (Bi-LSTM-CRF) [2]. Similar to the Stanford NLP four-class model of NER [20], the entities were labelled as person (PER), location (LOC), organization (ORG), or other (MISC).

4.2 Example Workflow

To illustrate how TextVista can be used to visually explore and analyze unstructured text data with temporal references, we describe an example workflow with Frank. Frank is a fictional policy analyst who is responsible for making suggestions to Canadian public health agencies about how to improve the efficacy of their policies. He is monitoring social media (i.e., Twitter) to identify trends in the public's response

[1] TextVista: https://witty-meadow-01b0d9610.2.azurestaticapps.net/.

[2] The workflow uses that final design of TextVista, though it went through multiple iterations before reaching this point. For example, the Filter Menu was moved based on participant comments in our evaluations.

[3] As this API was primarily trained to detect 'threatening language' in social media text, it is limited in that the subjective nature of threatening language can lead to biased or misclassifications.

Figure 1. The Sentiment and Semantic Clustering View. (A) The stacked histogram in this example depicts data over 30-minute intervals. When hovering over a section, a tooltip provides more information about the sentiment distribution. (B) The cluster view displays the data in semantic space, with (C) clustered outliers in the bottom left-hand corner indicating a bot. (D) The Reading Pane shows the underlying data source. (E) The Filter Menu allows toggling between clusters and sentiment.

to public health policies and to determine the benefits and drawbacks of current policies. He is analyzing tweets collected on May 10, 2022, about COVID-19 that may relate to Canada (i.e., with shared locations in Canada, mentions of Canadian health policies and Canadian public figures, etc.). His dataset contains 482 rows of CSV data and is 1 MB.

Frank imports his dataset into TextVista and begins exploring using the Text Frequency View (Figure 2). In this view, he sees the most frequent terms and term combinations (n-grams) plotted over 15-minute intervals (Figure 2A; T1, T2). In the Filter Menu (Figure 2B), he sees that "COVID-19" and "CANADA" occurred in the dataset 142 and 108 times, respectively. Immediately, Frank is intrigued by the n-gram "SAY NO TO" that occurred 36 times, so he clicks on its triangle in the Filter Menu. He looks at the Reading Pane for more details and notices that tweets with "SAY NO TO" were all created by a single author who was advertising fake vaccine QR codes (Figure 2D). As there is currently a policy about vaccine QR codes, Frank makes a note to investigate this further and report this to his team.

To understand the nature of the discourse in the data, Frank moves to the Sentiment and Semantic Clustering View. In this view, his data topics (clusters) are shown in a stacked histogram (Figure 1A; T4) and a cluster-based semantic space visualization (Figure 1B; T5). To avoid confusion with the Text Frequency View, a different set of colors was used to encode clusters. Frank also looks into the source of the text entries extracted from the data source (Figure 1D; T10). In the semantic visualization (Figure 1B), Frank notices outliers clustered in the bottom left corner of the semantic space (Figure 1C). He clicks on each dot, which highlights the data in the stacked histogram and places the source data entry at the top of the Reading Pane. Reading the source data reveals that Frank has a bot that is repeatedly tweeting *"Fauci (G.o.F.)+Ralph Baric/North Carolina"*; (Figure 1D).

Frank then hovers his mouse over the stacked histogram, and a tooltip reveals a summary of the sentiment in that histogram section (T7). Frank notices that one cluster has more negative sentiments than others. To further investigate this cluster's sentiment patterns over time (T7) and the toxicity of the topics (T9), Frank toggles the toggle switch to "Select Sentiment" (Figure 1E) to switch the focus of stacked bar to sentiment (Figure 3). Frank quickly sees that this cluster is very negative, with a few toxic sections.

Frank moves to the Entity Dial View, where entities were visualized as bubbles in the center of a radial stacked bar chart, with their icons indicating their entity type and their size indicating how frequently they occurred during the selected time period (Figure 4A). While using this view, Frank moved the dial around and noticed an *NHL* entity bubble in

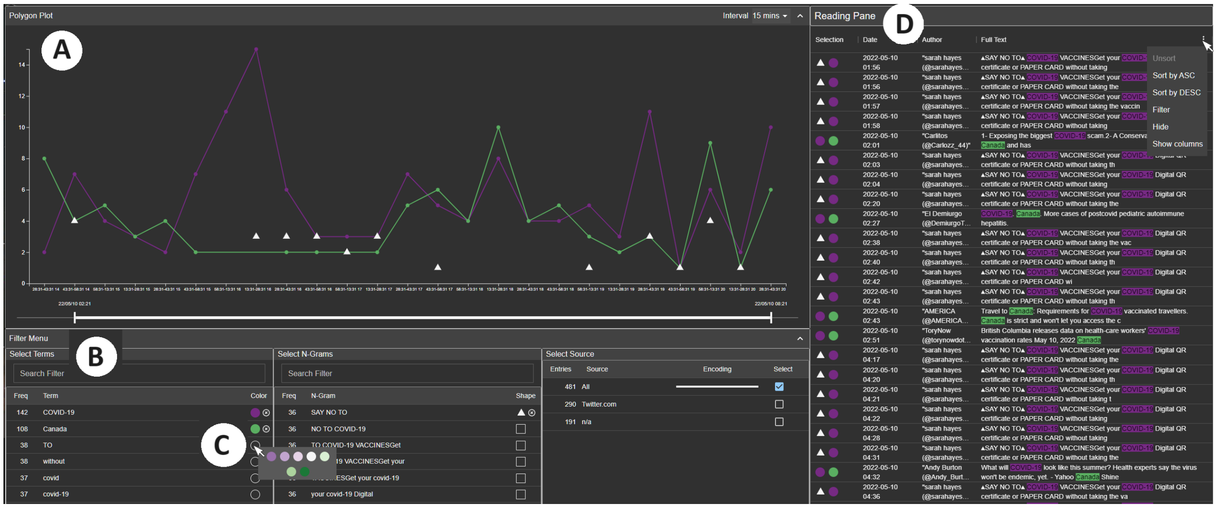

Figure 2. The Text Frequency View. (A) A line graph displaying term frequency over time. (B) The Filter Menu for term or n-gram (i.e., phrase) filtering and (C) Flexible encoding options using color or shape. (D) The Reading Pane shows dates, authors, sources, and source text for each term or n-gram, with the option to hide columns as needed.

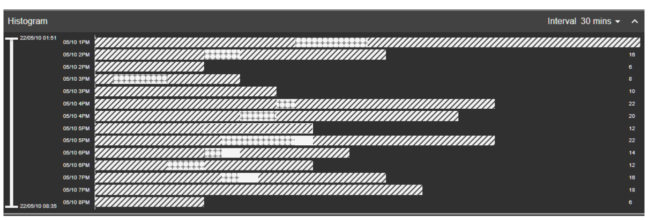

Figure 3. The Sentiment and Semantic Clustering View with the focus toggled to "sentiment." Negative sentiment is represented by diagonal lines, positive sentiment by crosses, and threatening sentiment by solid white color.

the centre of the view (T11). After clicking on it and reading the corresponding source tweets in the Reading Pane (Figure 4E), he sees tweets *"NHL teams setting their own itineraries to avoid COVID-19 testing requirements at US-Canada border"*. Frank makes a note to recommend standardizing testing practices at land borders to policymakers.

5 Think-Aloud Interview Study

After implementing TextVista, we wanted to evaluate how well it met the design requirements and needs of data analysts. We conducted a think-aloud interview study to get feedback and understand participants' thought processes using TextVista. We did not conduct a comparison study with a baseline system, as the goal of our system was to help analysts find insight into their data. In this case, a comparison or usability study is not appropriate to assess if analysts found insight [26]. Instead, we chose to interview participants while observing them using our system and thinking out loud. We focused on probing what participants learned about the data and how their insights were obtained (e.g., asking *"Did you have any 'aha' moments? Did you find any unexpected discoveries or connected some knowledge?"*) [48].

5.1 Participants and Method

We recruited 8 participants with experience in data analysis through our professional networks and snowball sampling. They all had experience in analyzing various forms of data, including text. P2, P3, and P4 were part of the initial focus group. The remaining five participants were new, including P5, a 34-year-old agender senior data visualization specialist with 2 years of experience; P6, a 37-year-old female data visualization specialist with less than a year of experience; P7, a 26-year-old male foresight analyst with 2 years of experience; P8, a 59-year-old female senior research manager and analyst with 7 years of experience; and P9, a 27-year-old male data scientist with 2 years of experience.

Participants completed the think-aloud interview study individually, each session lasting approximately one hour. Interviews took place via Microsoft Teams, with participants remotely accessing the prototype via their web browser and sharing their screens. Sessions were audio, video, and screen recorded, transcribed, and analyzed using grounded theory [12]. Each transcript was independently coded by two researchers, both of whom are master's students in human-computer interaction. These researchers underwent training in qualitative research from an experienced qualitative analyst. Researchers discussed and developed a codebook, resolving discrepancies during the coding process. Emerging themes and patterns from the codes were identified and discussed. The dataset used during the study comprised Reddit posts, TikTok video captions, and tweets about cryptocurrency that were posted between May 3, 2009, and August 4,

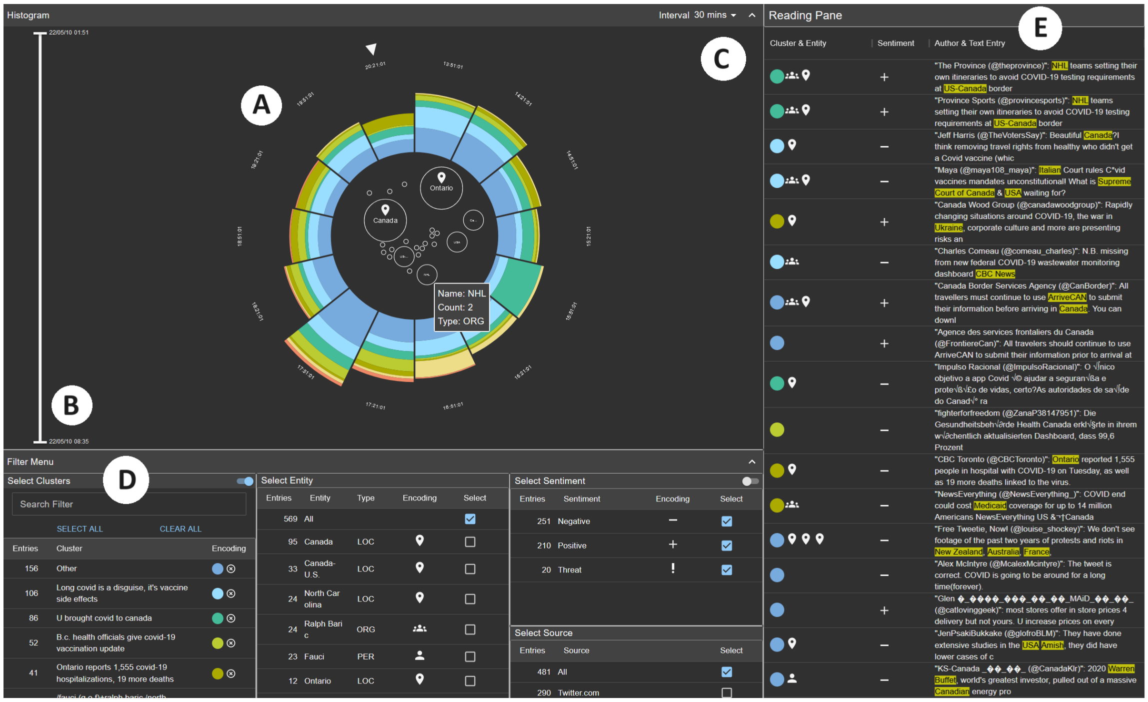

Figure 4. The Entity Dial View. (A) A radial stacked bar chart shows entities over time, with entities displayed as bubbles in the dial's center, their size indicating frequency. (B) The slider time range filter. (C) A dropdown time interval filter for selecting aggregation intervals. (D) The Filter Menu allows users to encode clusters by color and filter by entity, sentiment, and text source. (E) The Reading Pane, where users could read the source text.

2022. This dataset had over 20,400 rows, and each row was 25-70 words long. After completing a demographic survey, participants were given a 15-minute walkthrough of TextVista (~5 minutes/view). Participants then used TextVista to explore the dataset while thinking aloud. After, semi-structured interviews were conducted to understand what participants learned about the dataset, the potential use of TextVista in their data analysis workflows, and to gather feedback.

5.2 Results

The study revealed several insights about TextVista.

5.2.1 Supporting Reading. Participants appreciated the ability to review the source data corresponding to patterns that caught their interest. For instance, P9 mentioned *"One of the first things I wanted to do was click on these and find out what this [line] entailed, and so I do love the fact that that updates down here [in the Reading Pane]"*. P6 commented *"I really appreciate[d] the fact that the text can be accessed in its entirety."* P6 and P8 also commented that the Reading Pane was their favourite aspect of the visualization; P8 said *"I have to say that my favourite was in the text itself. I loved playing there...I love trying to manipulate [data] to see what I can get it to show me."* The flexibility of the Reading Pane was also found to be useful, with P2 and P3 commenting on the utility of adding or removing columns in the tabular Reading Pane as they thought about what they would like to have displayed while analyzing their own data.

5.2.2 Personalizing Explorations. TextVista had two ways to filter time, i.e., by time range and by aggregation interval. The first enabled participants to select the starting and ending shown in the visualization, while the second enabled them to choose the aggregation time interval, i.e., years, quarters, months, etc. P8 commented that they *"really liked the idea of being able to change the time element to see if I look at it by days, by weeks, by months– how does my visualization of it change? And then how do I think about how the data works differently because of that? That was an interesting aspect."* The interval-based time scaling did help participants uncover nuances in trends over time. TextVista's timescale was controlled by user selections, future iterations could add server-side architecture to support the automatic selection of the best aggregation interval based on a dataset's time range and a user's screen dimensions [31, 67]. Participants appreciated how choosing their own encodings supported their exploration of the data in different ways. P4 commented, *"I like the fact that you can choose your colors easily. It targets what you're looking at that I liked that a lot. This is something that whether they're different shapes and colors, they were intuitive and that was really a strength of the platform."*

5.2.3 Finding Patterns and Serendipitous Discoveries.

Participants commented that patterns quickly jumped out at them while using all three views, with half reporting having had 'Aha!' or serendipitous moments that encouraged them to dig deeper. In the Entity Dial View, bubbles piqued participants' interest. P2, P3, P5, P8 and P9 noticed the entity 'Elon Musk' and clicked on it. As P3 noted, *"There's an interesting moment here. This means where Elon Musk and Cryptomarket are both mentioned, Grimes is mentioned too."* Serendipitous discoveries were also made while using the Text Frequency View. Seeing the data by frequency piqued participants' curiosity and led them to further discoveries of a pattern (e.g., bots were identified, *"giveaway crypto cryptocurrency"*).

P8 was particularly interested in understanding the sentiment of the data, so they used the Sentiment and Semantic Clustering View to compare the amount of positive, negative, or threatening (toxic) content. Participants found TextVista engaging. P7 said, *"For me, [it] was just a really new and exciting way to look at data."* P9 commented, *"It's fun to interact with; it kind of makes data fun, creative, and cool."*. P2 highlighted the enjoyment of exploration, *"The greatest utility is being able to just mess around, play with data, and just explore what could come out because you never know. That's where you're going to potentially find really interesting insights."*

5.2.4 Offering Various Data Views.

During the study, participants shared their preferences for various features and views. These comments highlighted how providing multiple access points to the data not only encouraged exploration but also helped them address critical questions in their analysis. P3 preferred the Sentiment and Semantic Clustering View, saying *"That's where I would have probably spent most of my time because it answers the types of questions that I'm wondering when I'm looking at my dataset ... how are those elements meeting in the cluster of meaning?"*. P9 preferred the Text Frequency View, commenting, *"It's clear, it's concise, it tells me exactly what I want to know."*. P5 preferred the Entity Dial View, saying *"You can take some time and pause and analyze this, it's breathable. It's not overwhelming. ... it was really clear to see patterns, to see days or weeks with the highest amount of entries and the entities ... it's also fun and interesting to explore."* Each TextVista's view helped different analysts answer different types of questions they had in mind.

6 Diary Study

During the Think-Aloud Interview Study, we gathered insights into analysts' thought processes. However, participants' unfamiliarity with the dataset limited their ability to dig deeper into the data. Thus, evaluations of visualization systems to support data analysis need to capture a holistic and realistic perspective, which necessitates studying the systems in their intended environment with domain experts and realistic tasks [38]. need to capture a holistic and realistic perspective, which needs to be done by studying the systems used in their intended environment with realistic tasks and domain experts [38]. To understand the use of TextVista in the real world, we conducted a two-week diary study where participants used their own datasets with TextVista. We selected diary studies as they can provide insights about usage patterns, pain points, and potential improvements in a real-world setting [30, 66]. A two-week period allowed participants ample time with TextVista and 'incubation' periods for extended reflection and discovery.

6.1 Participants and Methods

We recruited 2 participants via our professional networks and snowball sampling. They both had experience in analyzing various forms of data, including text.P2 was part of the initial focus group and interview study, while P10 is a 49-year-old female analytics consultant with one year of experience. Participants used their own datasets during the study. Both participants collected and formatted their datasets in advance, providing them with a level of awareness of their data. P2 analyzed a dataset on global protest movements collected in May 2023, comprising 370 rows of text data sourced from online media. Each row was about 500 words long. P10 analyzed a dataset on election promises of Canadian politicians from 1994 to 2019, consisting of over 1,400 rows with each row containing 25-100 words.

We first applied NLP techniques (refer to Section 4.1) to preprocess the datasets and upload them to private, password-protected instances of TextVista accessible to participants for two weeks. Participants maintained a diary documenting their discoveries, analysis methods, time spent using TextVista, and ideas for improvements (See Supplementary Materials). After the two-week period, we conducted 30-minute semi-structured interviews via Microsoft Teams which were audio, video, and screen recorded. Transcripts and diaries were analyzed using grounded theory [12]. Each transcript was independently coded by two researchers, both of whom are master's students in human-computer interaction. These researchers underwent training in qualitative research from an experienced qualitative analyst. Researchers discussed and developed a codebook, resolving discrepancies during the coding process. Emerging themes and patterns from the codes were identified and discussed.

6.2 Results

By exploring data with TextVista, participants were able to find new insights. P2 wrote 7 journal entries, analyzing a global protest movement dataset to understand active movements, involved entities, and media narratives. P10, with 1 journal entry, examined a political promise dataset to uncover prevalent words, themes, and political rhetoric changes over time. Here, we reflect on the analysis of these journals.

6.2.1 Analysis of Global Protest Data (P2).

P2 "had a hunch" about the protest movement patterns across the globe when he began his analysis and found that the Entity Dial view was the most useful for his goals. He described the aggregate view that Entity Dial provided him "immediate value" by showing him a useful summary of his data. As he explored his data, he formulated new questions to answer, i.e., *"there were questions that arose from looking at that [Entity Dial] graph and so that's where I spent most of my time."* P2 highlighted how the Entity Dial view aided his understanding of frequently mentioned entities in news articles about protest movements. He utilized the flexible time scale to focus on specific events and accessed the Reading Pane by clicking on entities, saying *"The dial is a great way to select relevant entries in the Reading Pane. I can trace back the most important stories that happened."* This approach deepened P2's understanding of relationships and dynamics among the entities of interest, revealing differing media narratives.

P2 was intrigued by the Sentiment and Semantic Clustering view but found it difficult to understand the visual representation of the semantic space. While he could detect outliers, P2 found it difficult to understand why certain clusters were outliers or why they were grouped together, e.g., *"I know that it's based on an algorithm that joins things together, clusters things together, but how did it come to those results? if I could understand...why are those [entries] close together? That would be useful."* Visualizing the decision process of the clustering algorithm could help users understand how the process has classified their data in a way that only visualizing the final result cannot. P2 was able to find new insights using the Entity Dial view, via its flexible time scale and Reading Pane. The patterns revealed in the view encouraged him to ask new questions about his data, which he could answer by filtering and accessing the Reading Pane.

6.2.2 Analysis of Political Rhetoric Data (P10).

P10 began his process by exploring each view. When he encountered a pattern that he didn't understand, he wrote down his questions about the data and returned to them later, e.g., *"Some of the results were immediately clear when I clicked on the patterns [but] in some cases when I was analyzing it and being like, OK, I don't understand what this is, then just writing it, it brought me other questions and when I went back to the app, I was actually able to figure it out."* This strategy of asking new questions and stepping away from the analysis for a time served as an incubation period during his analysis.

As P10 was primarily concerned with topics and terms used, he reviewed the terms by exploring the frequent terms and n-grams in the Text View and focused on particular time periods. This led to insights into trends in political parties' rhetorics over the years. The Entity Dial view confirmed P10's ideas about the entities discussed at different times. Similar to P2, P10 found the semantic space in the Sentiment and Semantic Clustering was difficult to understand.

6.3 Summary

Both P2 and P10 found insights in their data and confirmed that TextVista was useful for their analysis. Certain features were particularly helpful for both participants, namely the quick transitions between the visualizations and the Reading Pane and the flexible encodings. Both of these features enabled participants to identify patterns and gain an understanding of what these patterns meant. The temporal dimension of the views helped both participants understand how their data evolved over time. While the stacked histogram in the Sentiment and Semantic Clustering view made sense to participants, they found the semantic space confusing and difficult to understand. They expressed a desire for more understandable cluster labels and transparency about how the clustering algorithm created the clusters.

7 Discussion

The development and evaluation of TextVista revealed many insights about the design of future time-series text-based data visualization systems that leverage NLP and support reading, as well as opportunities for future NLP techniques to support visualization. Based on these insights, we provide guidelines for designing NLP-enriched text-based data visualizations.

Supporting Reading during Analysis Workflows: All participants highlighted the importance of reading in their analysis workflows to understand the nuances of relationships within their datasets. The in-situ support for reading enabled participants to have more fluid, intuitive workflows across TextVista's multiple views. These findings are similar to those of other text-focused visualization systems such as ChartWalk [61] and Doccurate [62] and echo beliefs that despite advances in NLP, systems should not use NLP to replace fine-grained human judgment [21]. During our focus group, participants wanted a visualization that would help them understand the relationships between entities and topics in their data. They found it challenging to grasp these relationships due to the dynamic nature of perspectives and changes over time within different data segments. While TextVista's views showed trends, analyzing deeper relationships between entities required close reading (e.g., P2 read the source text to understand how entities in different protests related to each other). Using NLP techniques to identify or quantify such relationships would be useful to explore, as would the development of in-text visualization techniques to support the reading of multiple data sources at once [59]. . Deep learning text summarization techniques may also be useful, but the level of summarization and trustworthiness of summaries for users reading the source data directly needs to be determined [56].

Facilitating Discovery; Read to Deepen Understanding: During our focus groups, we identified how participants were often triggered by topics, entities, sentiments, and unexpected words while discovering *unknown unknowns*. It is important to note that the *context* of these triggers influenced the discoveries that were made. TextVista supported such discoveries in two ways. First, question-focused overviews are designed to prompt participants to ask new questions about their data and guide them in determining what to explore next. These views facilitated exploratory searches, giving analysts the autonomy to decide what should be visualized. Second, by incorporating source text within every view, TextVista allowed fluid transitions between visualizations and source text. This allowed participants to utilize reading to uncover insights beyond what the visualizations displayed. The fusion of these techniques enabled participants uncover more insights about their data and deepen their understanding of these insights as they moved between views and source data, providing a more comprehensive analysis experience. There is a need for further research [3, 64] to discover what makes an effective serendipitous trigger and how to contextualize these triggers using visualization.

Transparency of Visualized Data: The Sentiment and Semantic Clustering view visualized the underlying data to provide insights into its semantic nature. In some cases, this enabled the detection of outliers and bots, but this was not always appreciated. For example, during the Diary Study, P2 detected outliers but couldn't understand why these data points were outliers. In the field of AI, there has been an increased interest in explainable AI (XAI) and techniques that can provide more transparency about the data and methods used within AI and machine learning algorithms [9]. While the Sentiment and Semantic Clustering view was found to be useful, future work should explore which techniques or UI elements could be integrated into systems similar to TextVista to provide increased transparency about the NLP techniques and data used within visualizations.

Opportunities for (Semi-) Guided Exploration: Although TextVista's views were designed to address underlying questions and tasks, they did not explicitly guide participants to specific questions about their data or suggest areas of focus within the views. Participants were left to freely explore the interface and focus on what was important to *them*; however, some users may benefit from the inclusion of semi-guided exploration techniques. By incorporating techniques to compute data interestingness [23] or detect outliers [1], or by utilizing one's interaction history, systems could highlight unusual or interesting entities, topics, or outliers that a user may want to explore. Such functionality should be customizable, allowing users to turn it on/off as needed or perhaps automatically activate when a user seems fixated on a particular view or topic in their data. In scenarios where multiple users have consulted a dataset, systems could also provide awareness about dataset segments or views that received abundant or little attention. However, which UI components or techniques should be used to draw attention to or away from such content and how much or little guidance should be provided to users remains to be explored.

Limitations: This research adopted a user-centered design approach to develop TextVista; however, it should be noted that user-centered design can sometimes lead to an over-reliance on user feedback, potentially stifling innovation as designers might only iterate on existing ideas rather than explore radical new ones [47]. Challenges can also arise when this system is scaled up and delivered to a wider set of users and large or more unique datasets are employed. Future work should focus on assessing TextVista's performance and scalability with larger datasets. As such, the findings should be carefully considered when applying them to structured text visualizations or data without time-based information. The number of participants in the design process was small, which may not represent all potential future users, so the findings and TextVista's implementation may have unintentionally perpetuated biases or overlooked the needs of future users [14]. Additionally, our diary study showed adherence problems [55]. However, we believe sharing the results of such deep, sustained engagement in the system is beneficial in choosing when and with whom to conduct a diary study.

8 Conclusion

Given the increasing volume of text-based data in today's digital landscape, relying solely on manual techniques for analysis is impractical. This work sought to design and develop a visualization system that integrated NLP techniques to support the analysis of unstructured text with temporal references. We began by conducting a focus group with data analysts to understand text-based analysis processes and needs. The focus group discussions were distilled into design requirements and visualization tasks to guide the development of TextVista. We conducted two evaluation studies: a think-aloud interview study and a diary study. From the think-aloud interview study, we gained insight into how well TextVista addressed our design requirements and into our users' thought processes as they used the visualization. The diary study provided deeper insights into how TextVista was able to support long-term text analysis. Across both evaluations, participants found TextVista to be a useful tool that supported them in determining what to explore in their data and prompted them to ask new questions about their data.

9 Acknowledgments

We thank our participants whose invaluable contributions made this research possible. Additionally, we extend our appreciation to the Natural Sciences and Engineering Research Council of Canada (NSERC), Ontario Center of Innovation, and NovaceneAI, for their generous support.

References

[1] Charu C. Aggarwal. 2017. *Outlier Detection in Categorical, Text, and Mixed Attribute Data.* Springer International Publishing, Cham, 249–272. https://doi.org/10.1007/978-3-319-47578-3_8

[2] Alan Akbik, Duncan Blythe, and Roland Vollgraf. 2018. Contextual String Embeddings for Sequence Labeling. In *Proceedings of the 27th International Conference on Computational Linguistics*. Association for Computational Linguistics, 1638–1649. https://aclanthology.org/C18-1139

[3] Eric Alexander, Joe Kohlmann, Robin Valenza, Michael Witmore, and Michael Gleicher. 2014. Serendip: Topic model-driven visual exploration of text corpora. In *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)* (Paris, France). IEEE, 173–182. https://doi.org/10.1109/VAST.2014.7042493

[4] Mohammad Alharbi and Robert S. Laramee. 2019. SoS TextVis: An Extended Survey of Surveys on Text Visualization. *Computers* 8, 1 (2019). https://doi.org/10.3390/computers8010017

[5] R. Amar, J. Eagan, and J. Stasko. 2005. Low-level components of analytic activity in information visualization. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.* 111–117. https://doi.org/10.1109/INFVIS.2005.1532136

[6] Michele Avalle, Niccolò Di Marco, Gabriele Etta, Emanuele Sangiorgio, Shayan Alipour, Anita Bonetti, Lorenzo Alvisi, Antonio Scala, Andrea Baronchelli, Matteo Cinelli, and Walter Quattrociocchi. 2024. Persistent interaction patterns across social media platforms and over time. *Nature* (03 2024), 1–8. https://doi.org/10.1038/s41586-024-07229-y

[7] Russell Beale. 2007. Supporting serendipity: Using ambient intelligence to augment user exploration for data mining and web browsing. *International Journal of Human-Computer Studies* 65, 5 (2007), 421–433. https://doi.org/10.1016/j.ijhcs.2006.11.012 Ambient intelligence: From interaction to insight.

[8] Enrico Bertini and Denis Lalanne. 2010. Investigating and Reflecting on the Integration of Automatic Data Analysis and Visualization in Knowledge Discovery. *SIGKDD Explor. Newsl.* 11, 2 (2010), 9–18. https://doi.org/10.1145/1809400.1809404

[9] Saša Brdnik. 2023. GUI Design Patterns for Improving the HCI in Explainable Artificial Intelligence. In *Companion Proceedings of the 28th International Conference on Intelligent User Interfaces* (Sydney, NSW, Australia) *(IUI'23 Companion)*. Association for Computing Machinery, 240–242. https://doi.org/10.1145/3581754.3584114

[10] Nan-Chen Chen, Michael Brooks, Rafal Kocielnik, Sungsoo Ray Hong, Jeff Smith, Sanny Lin, Zening Qu, and Cecilia Aragon. 2017. Lariat: A Visual Analytics Tool for Social Media Researchers to Explore Twitter Datasets. https://doi.org/10.24251/HICSS.2017.228

[11] Christopher Collins, Fernanda B. Viegas, and Martin Wattenberg. 2009. Parallel Tag Clouds to explore and analyze faceted text corpora. In *2009 IEEE Symposium on Visual Analytics Science and Technology*. 91–98. https://doi.org/10.1109/VAST.2009.5333443

[12] Juliet M Corbin and Anselm Strauss. 1990. Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative sociology* 13, 1 (1990), 3–21.

[13] Shainen Davidson, Vaibhav Kesarwani, and Kenton White. 2022. Forecasting and Understanding the 2021 Canadian Federal Election Using Twitter Conversations. *Proceedings of the Canadian Conference on Artificial Intelligence* (2022). https://caiac.pubpub.org/pub/ewuzz3aj

[14] Nicola Dell, Vidya Vaidyanathan, Indrani Medhi, Edward Cutrell, and William Thies. 2012. "Yours is Better!": Participant Response Bias in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI'12)*. Association for Computing Machinery, 1321–1330. https://doi.org/10.1145/2207676.2208589

[15] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *CoRR* abs/1810.04805 (2018). https://doi.org/10.48550/arXiv.1810.04805 arXiv:1810.04805

[16] Miguel Pina e Cunha et al. 2005. *Serendipity: Why some organizations are luckier than others.* Technical Report. Universidade Nova de Lisboa, Nova School of Business and Economics.

[17] James K Esser. 1998. Alive and well after 25 years: A review of groupthink research. *Organizational behavior and human decision processes* 73, 2-3 (1998), 116–141. https://doi.org/10.1006/obhd.1998.2758

[18] Cristian Felix, Anshul Vikram Pandey, and Enrico Bertini. 2017. TextTile: An Interactive Visualization Tool for Seamless Exploratory Analysis of Structured Data and Unstructured Text. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 161–170. https://doi.org/10.1109/TVCG.2016.2598447

[19] Gary Alan Fine and James G Deegan. 1996. Three principles of serendip: insight, chance, and discovery in qualitative research. *International Journal of Qualitative Studies in Education* 9, 4 (1996), 434–447. https://doi.org/10.1080/0951839960090405

[20] Jenny Rose Finkel, Trond Grenager, and Christopher Manning. 2005. Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling. In *Proceedings of the 43nd Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, pp. 363–370. http://nlp.stanford.edu/~manning/papers/gibbscrf3.pdf

[21] Ilas Flaounas, Omar Ali, Thomas Lansdall-Welfare, Tijl De Bie, Nick Mosdell, Justin Lewis, and Nello Cristianini. 2013. Research Methods in the Age of Digital Journalism. *Digital Journalism* 1:1 (2013), 102–115. https://doi.org/10.1080/21670811.2012.714928

[22] Johanna Fulda, Matthew Brehmer, and Tamara Munzner. 2016. TimeLineCurator: Interactive Authoring of Visual Timelines from Unstructured Text. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2016), 300–309. https://doi.org/10.1109/TVCG.2015.2467531

[23] Liqiang Geng and Howard J. Hamilton. 2006. Interestingness Measures for Data Mining: A Survey. *ACM Comput. Surv.* 38, 3 (2006), 9–es. https://doi.org/10.1145/1132960.1132963

[24] Sarah Gibbons. 2016. Design Thinking. https://www.nngroup.com/articles/design-thinking/

[25] Anna Glazkova, Maksim Glazkov, and Timofey Trifonov. 2020. g2tmn at Constraint@AAAI2021: Exploiting CT-BERT and Ensembling Learning for COVID-19 Fake News Detection. *CoRR* abs/2012.11967 (2020). arXiv:2012.11967 https://arxiv.org/abs/2012.11967

[26] Saul Greenberg and Bill Buxton. 2008. Usability Evaluation Considered Harmful (Some of the Time). In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI'08)*. Association for Computing Machinery, 111–120. https://doi.org/10.1145/1357054.1357074

[27] Abram Handler and Brendan O'Connor. 2017. Rookie: A unique approach for exploring news archives. *CoRR* abs/1708.01944 (2017). http://arxiv.org/abs/1708.01944

[28] Jeffrey Heer, Frank van Ham, Sheelagh Carpendale, Chris Weaver, and Petra Isenberg. 2008. *Creation and Collaboration: Engaging New Audiences for Information Visualization*. Springer, 92–133. https://doi.org/10.1007/978-3-540-70956-5_5

[29] C. Hutto and Eric Gilbert. 2014. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media* 8, 1 (2014), 216–225. https://doi.org/10.1609/icwsm.v8i1.14550

[30] E.H. Janssens, K.A.M. andv Bos and J.G.M. et al. Rosmalen. 2018. A qualitative approach to guide choices for designing a diary study. *BMC Med Res Methodology* 18, 140 (2018). https://doi.org/10.1186/s12874-018-0579-6

[31] Uwe Jugel, Zbigniew Jerzak, Gregor Hackenbroich, and Volker Markl. 2014. M4: A Visualization-Oriented Time Series Data Aggregation. *Proceedings of the VLDB Endowment* 7 (2014), 797 – 808. https://doi.org/10.14778/2732951.2732953

[32] Daniel Jurafsky and James H. Martin. 2023. *Speech and Language Processing, An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.* Prentice Hall. 131 pages.

[33] Ethan Kerzner, Sarah Goodwin, Jason Dykes, Sara Jones, and Miriah Meyer. 2019. A Framework for Creative Visualization-Opportunities Workshops. *IEEE Transactions on Visualization and Computer Graphics*

25, 1 (2019), 748–758. https://doi.org/10.1109/TVCG.2018.2865241
[34] Cornelia Kiefer. 2016. Assessing the Quality of Unstructured Data: An Initial Overview.. In *LWDA*. 62–73.
[35] Christian Knoll, Asil Çetin, Torsten Möller, and Miriah Meyer. 2020. Extending Recommendations for Creative Visualization-Opportunities Workshops. In *2020 IEEE Workshop on Evaluation and Beyond - Methodological Approaches to Visualization (BELIV)*. IEEE, 81–88. https://doi.org/10.1109/BELIV51497.2020.00017
[36] Kostiantyn Kucher and Andreas Kerren. 2015. Text visualization techniques: Taxonomy, visual survey, and community insights. In *2015 IEEE Pacific Visualization Symposium (PacificVis)*. 117–121. https://doi.org/10.1109/PACIFICVIS.2015.7156366
[37] Philippe Laban and Marti Hearst. 2017. newsLens: building and visualizing long-ranging news stories. In *Proceedings of the Events and Stories in the News Workshop*, Tommaso Caselli, Ben Miller, Marieke van Erp, Piek Vossen, Martha Palmer, Eduard Hovy, Teruko Mitamura, and David Caswell (Eds.). Association for Computational Linguistics, Vancouver, Canada, 1–9. https://doi.org/10.18653/v1/W17-2701
[38] Heidi Lam, Enrico Bertini, Petra Isenberg, Catherine Plaisant, and Sheelagh Carpendale. 2011. *Seven Guiding Scenarios for Information Visualization Evaluation*. Research Report 2011-992-04. https://inria.hal.science/hal-00723057 Superseded by and improved in a follow-up journal article..
[39] Alyssa Lees, Vinh Q. Tran, Yi Tay, Jeffrey Sorensen, Jai Gupta, Donald Metzler, and Lucy Vasserman. 2022. A New Generation of Perspective API: Efficient Multilingual Character-Level Transformers. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (Washington DC, USA) *(KDD'22)*. Association for Computing Machinery, 3197–3207. https://doi.org/10.1145/3534678.3539147
[40] Tal Linzen. 2016. Issues in evaluating semantic spaces using word analogies. In *Proceedings of the 1st Workshop on Evaluating Vector-Space Representations for NLP*. Association for Computational Linguistics, 13–18. https://doi.org/10.18653/v1/W16-2503
[41] Shixia Liu, Xiting Wang, Christopher Collins, Wenwen Dou, Fangxin Ouyang, Mennatallah El-Assady, Liu Jiang, and Daniel A. Keim. 2019. Bridging Text Visualization and Mining: A Task-Driven Survey. *IEEE Transactions on Visualization and Computer Graphics* 25, 7 (2019), 2482–2504. https://doi.org/10.1109/TVCG.2018.2834341
[42] Dongning Luo, Jing Yang, Miloš Krstajić, William Ribarsky, and Daniel Keim. 2012. EventRiver: Visually Exploring Text Collections with Temporal References. *IEEE transactions on visualization and computer graphics* 18 (01 2012), 93–105. https://doi.org/10.1109/TVCG.2010.225
[43] Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. 2011. Learning Word Vectors for Sentiment Analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1* (Portland, Oregon) *(HLT'11)*. Association for Computational Linguistics, 142–150. https://doi.org/doi/10.5555/2002472.2002491
[44] Lori McCay-Peet and Elaine G. Toms. 2010. The Process of Serendipity in Knowledge Work. In *Proceedings of the Third Symposium on Information Interaction in Context* (New Brunswick, New Jersey, USA) *(IIiX'10)*. Association for Computing Machinery, 377–382. https://doi.org/10.1145/1840784.1840842
[45] Tamara Munzner. 2009. A Nested Model for Visualization Design and Validation. *IEEE Transactions on Visualization and Computer Graphics* 15, 6 (2009), 921–928. https://doi.org/10.1109/TVCG.2009.111
[46] Xi Niu and Fakhri Abbas. 2017. A Framework for Computational Serendipity. In *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization* (Bratislava, Slovakia) *(UMAP'17)*. Association for Computing Machinery, 360–363. https://doi.org/10.1145/3099023.3099097
[47] Don Norman. 2013. *The design of everyday things: Revised and expanded edition*. Basic books.
[48] C. North. 2006. Toward measuring visualization insight. *IEEE Computer Graphics and Applications* 26, 3 (2006), 6–9. https://doi.org/10.1109/MCG.2006.70
[49] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. The PageRank Citation Ranking : Bringing Order to the Web. In *The Web Conference*. https://api.semanticscholar.org/CorpusID:1508503
[50] Deokgun Park, Seungyeon Kim, Jurim Lee, Jaegul Choo, Nicholas Diakopoulos, and Niklas Elmqvist. 2018. ConceptVector: Text Visual Analytics via Interactive Lexicon Building Using Word Embedding. *IEEE Transactions on Visualization and Computer Graphics* 24, 1 (2018), 361–370. https://doi.org/10.1109/TVCG.2017.2744478
[51] Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084* (2019). https://doi.org/10.18653/v1/d19-1410
[52] Guilherme H. Resende, Luiz F. Nery, Fabrício Benevenuto, Savvas Zannettou, and Flavio Figueiredo. 2024. A Comprehensive View of the Biases of Toxicity and Sentiment Analysis Methods Towards Utterances with African American English Expressions. arXiv:2401.12720 [cs.CL]
[53] Christian Rohrdantz, Ming C. Hao, Umeshwar Dayal, Lars-Erik Haug, and Daniel A. Keim. 2012. Feature-Based Visual Sentiment Analysis of Text Document Streams. *ACM Trans. Intell. Syst. Technol.* 3, 2, Article 26 (feb 2012), 25 pages. https://doi.org/10.1145/2089094.2089102
[54] Donald Rumsfeld. 2011. *Known and unknown: a memoir*. Penguin.
[55] Anjeli Singh and Sareeka Malhotra. 2013. A Researcher's Guide to Running Diary Studies. In *Proceedings of the 11th Asia Pacific Conference on Computer Human Interaction* (Bangalore, India) *(APCHI'13)*. Association for Computing Machinery, 296–300. https://doi.org/10.1145/2525194.2525261
[56] Shengli Song, Haitao Huang, and Tongxiao Ruan. 2019. Abstractive text summarization using LSTM-CNN based deep learning. *Multimedia Tools and Applications* 78 (2019), 857–875.
[57] Axel J. Soto, Ryan Kiros, Vlado Kešelj, and Evangelos Milios. 2015. Exploratory Visual Analysis and Interactive Pattern Extraction from Semi-Structured Data. *ACM Trans. Interact. Intell. Syst.* 5, 3, Article 16 (sep 2015), 36 pages. https://doi.org/10.1145/2812115
[58] John Stasko, Carsten Gorg, Zhicheng Liu, and Kanupriya Singhal. 2007. Jigsaw: Supporting Investigative Analysis through Interactive Visualization. In *2007 IEEE Symposium on Visual Analytics Science and Technology*. 131–138. https://doi.org/10.1109/VAST.2007.4389006
[59] Nicole Sultanum, Anastasia Bezerianos, and Fanny Chevalier. 2021. Text Visualization and Close Reading for Journalism with Storifier. In *2021 IEEE Visualization Conference (VIS)*. IEEE, 186–190. https://doi.org/10.1109/VIS49827.2021.9623264
[60] Nicole Sultanum, Michael Brudno, Daniel Wigdor, and Fanny Chevalier. 2018. More Text Please! Understanding and Supporting the Use of Visualization for Clinical Text Overview. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI'18)*. Association for Computing Machinery, 1–13. https://doi.org/10.1145/3173574.3173996
[61] Nicole Sultanum, Farooq Naeem, Michael Brudno, and Fanny Chevalier. 2023. ChartWalk: Navigating large collections of text notes in electronic health records for clinical chart review. *IEEE Transactions on Visualization and Computer Graphics* 29, 1 (2023), 1244–1254. https://doi.org/10.1109/TVCG.2022.3209444
[62] Nicole Sultanum, Devin Singh, Michael Brudno, and Fanny Chevalier. 2019. Doccurate: A Curation-Based Approach for Clinical Text Visualization. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2019), 142–151. https://doi.org/10.1109/TVCG.2018.2864905
[63] Guodao Sun, Yingcai Wu, Shixia Liu, Tai-Quan Peng, Jonathan J. H. Zhu, and Ronghua Liang. 2014. EvoRiver: Visual Analysis of Topic Coopetition on Social Media. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 1753–1762. https://doi.org/10.1109/TVCG.2014.2346919
[64] Alice Thudt, Uta Hinrichs, and Sheelagh Carpendale. 2012. The Bohemian Bookshelf: Supporting Serendipitous Book Discoveries

through Information Visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI'12)*. Association for Computing Machinery, 1461–1470. https://doi.org/10.1145/2207676.2208607

[65] Paul Tol. 2021. Colour Schemes. Issue 3.2. https://personal.sron.nl/~pault/data/colourschemes.pdf

[66] Martin Tomitsch, Nikash Singh, and Ghazaleh Javadian. 2010. Using Diaries for Evaluating Interactive Products: The Relevance of Form and Context. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction* (Brisbane, Australia) *(OZCHI'10)*. Association for Computing Machinery, 204–207. https://doi.org/10.1145/1952222.1952266

[67] Fujee Tsung and Kaibo Wang. 2010. Adaptive charting techniques: Literature review and extensions. *Frontiers in Statistical Quality Control 9* (2010), 19–35.

[68] Marlene E Turner and Anthony R Pratkanis. 1998. Twenty-five years of groupthink theory and research: Lessons from the evaluation of a theory. *Organizational behavior and human decision processes* 73, 2-3 (1998), 105–115. https://doi.org/10.1006/obhd.1998.2756

[69] Peter D. Turney. 2002. Thumbs up or Thumbs down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics* (Philadelphia, Pennsylvania) *(ACL'02)*. Association for Computational Linguistics, 417–424. https://doi.org/10.3115/1073083.1073153

[70] Md Main Uddin Rony, Enamul Hoque, and Naeemul Hassan. 2020. ClaimViz: Visual Analytics for Identifying and Verifying Factual Claims. In *2020 IEEE Visualization Conference (VIS)*. 246–250. https://doi.org/10.1109/VIS47514.2020.00056

[71] Furu Wei, Shixia Liu, Yangqiu Song, Shimei Pan, Michelle X. Zhou, Weihong Qian, Lei Shi, Li Tan, and Qiang Zhang. 2010. TIARA: a visual exploratory text analytic system. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Washington, DC, USA) *(KDD '10)*. Association for Computing Machinery, New York, NY, USA, 153–162. https://doi.org/10.1145/1835804.1835827

[72] Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2009. Articles: Recognizing Contextual Polarity: An Exploration of Features for Phrase-Level Sentiment Analysis. *Computational Linguistics* 35 (2009), 399–433. https://api.semanticscholar.org/CorpusID:9423000

[73] Seid Muhie Yimam, Heiner Ulrich, Tatiana von Landesberger, Marcel Rosenbach, Michaela Regneri, Alexander Panchenko, Franziska Lehmann, Uli Fahrer, Chris Biemann, and Kathrin Ballweg. 2016. new/s/leak – Information Extraction and Visualization for Investigative Data Journalists. In *Proceedings of ACL-2016 System Demonstrations*. Association for Computational Linguistics, 163–168. https://doi.org/10.18653/v1/P16-4028

[74] Xiaoyu Zhang, Xiwei Xuan, Alden Dima, Thurston Sexton, and Kwan-Liu Ma. 2023. LabelVizier: Interactive Validation and Relabeling for Technical Text Annotations. In *2023 IEEE 16th Pacific Visualization Symposium (PacificVis)*. 167–176. https://doi.org/10.1109/PacificVis56936.2023.00026