FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Narrative Visualization of News Stories

Mariana Filipa da Costa

Mestrado Integrado em Engenharia Informática e Computação

Supervisor: Prof. Sérgio Nunes, PhD

July 19, 2021

Narrative Visualization of News Stories

Mariana Filipa da Costa

Mestrado Integrado em Engenharia Informática e Computação

Approved in oral examination by the committee:

Chair: Prof. Alexandre Valle de Carvalho, PhD
External Examiner: Prof. Eduarda Mendes Rodrigues, PhD
Supervisor: Prof. Sérgio Sobral Nunes, PhD

July 19, 2021

Abstract

The issue of information overload has become all too prevalent in our daily lives. Too many sources of varying degrees of credibility have made their way into our social network feeds, our notifications, our televisions; there is too much of everything, everywhere we look. The hopeful promise brought about by the Internet, of an interconnected world in which information flows freely and uninterruptedly, has been supplanted by the dreadful realization of our inability to handle such an overwhelming amount of information.

While the visualization of news content has been a fertile field of research, the focus tends to be on a large-scale analysis, frequently through the use of topics to describe extensive collections of news articles. Moreover, approaches that bring about the issue of narrative in journalistic content are quite rare. This dissertation proposes the use of visualization as a method to analyse the underlying narrative in news articles and to showcase it in a concise and memorable manner. The goal is to present the reader with an information hub for a journalistic narrative that highlights its actors, locations, and events in a time aware manner.

We began by conducting interviews with five journalists in order to explore the problem space. The journalists agreed that visualizations have the potential to be useful in helping the average reader understand a news story, if well designed. However, sub-par visualizations distract from the original story and only serve to add noise. Situations where the time frame and order of events is important to the comprehension of the story are some of the contexts in which visualizations can excel. The participants recognized that visualizations can work with both past occurrences and recent events; nevertheless, they point out that using them for stories still in development poses additional challenges such as the need to update the visualization frequently and the difficulty in keeping the consumer engaged. Arguments were made in favour of high-level, abstract visualizations focused on topics and more narrative-driven ones, with more journalists leaning slightly towards the latter. Finally, we must also take into consideration the learning curve that might be associated with more intricate visualizations, as well as the possibility that not everyone's reading style might accommodate for heavily visual stories.

We then designed a prototype aimed at exploring the use of storyline visualizations with news stories. Having an open-source framework for creating narrative charts as our foundation, we have built a publicly available prototype that highlights the events and participants in a story that is given as input to the system. This visualization allows for a number of interactions, including zooming in on a particular event, filtering events, entities, and time frames, and editing the placement of elements in the visualization. This prototype was then evaluated through user tests and a survey. The user tests were conducted with five journalists, some who also participated in the initial interviews. As for the survey, it was sent out to the academic community of the University of Porto, from which 178 responses were registered. By contacting journalists and the general public, we were able to gather insight into how both the hypothetical creators and consumers perceive the application. The results demonstrate that participants found the prototype interesting, albeit with usability problems. We also collected a substantial amount of suggestions for improve-

ment, which could serve as the starting point for future development cycles of the prototype. From the feedback provided during both evaluation methods, we conclude that storyline visualizations can be successfully applied to journalistic narratives, thus proving to be a viable research path for future endeavors.

Keywords: News visualization, Text visualization, Human-computer interaction, Visualization of narratives

Resumo

A questão da sobrecarga de informação tem-se tornado uma constante no nosso dia-a-dia. Demasiadas fontes de informação com um grau de credibilidade variado surgem nas nossas redes sociais, nas nossas notificações, nas nossas televisões; há demasiado de tudo, onde quer que olhemos. A promessa feita pela Internet de um mundo interligado no qual a informação se desloca livre e ininterruptamente foi suplantada pela constatação da nossa incapacidade de lidar com a quantidade avassaladora de conteúdo ao qual somos expostos.

Ainda que a visualização de conteúdo noticioso seja uma área de investigação prolífica, o foco tende a ser na análise em larga escala, frequentemente através do uso de tópicos por forma a descrever extensas coleções de artigos noticiosos. Adicionalmente, estudos que abordem a questão da narrativa no conteúdo jornalístico são raros. Esta dissertação propõe o uso da visualização como método de análise e exposição da narrativa subjacente a artigos noticiosos, de uma forma concisa e memorável. O objetivo é apresentar ao leitor um centro de informação sobre a narrativa jornalística que saliente os participantes, localizações e eventos, seguindo uma escala temporal.

Começámos por conduzir entrevistas com cinco jornalistas com o intuito de explorar o espaço do problema. Os jornalistas concordaram que as visualizações, se desenhadas adequadamente, têm o potencial para serem úteis na compreensão da notícia por parte do leitor comum. Contudo, visualizações fracas distraem da história e apenas acrescentam ruído. Situações em que o enquadramento temporal e a ordem dos eventos são importantes para a compreensão da história são alguns dos contextos nos quais a visualização pode sobressair. Os participantes reconheceram que as visualizações são apropriadas tanto para ocorrências históricas como para eventos recentes; mesmo assim, eles mencionaram que o uso de visualizações para histórias em desenvolvimento apresenta desafios adicionais, como a necessidade de atualizar a visualização frequentemente e a dificuldade em manter os leitores cativados. Foram dados argumentos tanto a favor de visualizações mais abstratas, de alto nível, bem como a favor de visualizações mais voltadas para a narrativa, sendo que os jornalistas se mostraram ligeiramente inclinados para as visualizações narrativas. Por fim, temos de considerar a curva de aprendizagem que pode estar associada com visualizações mais intrincadas, bem como a possibilidade de que nem todos podem ter o estilo de leitura necessário para acomodar histórias altamente visuais.

Posteriormente, desenvolvemos um protótipo com o intuito de explorar o uso de visualizações de cadeias narrativas para artigos noticiosos. Tendo como ponto de partida uma *framework* disponível em código aberto para a criação de gráficos narrativos, construímos um protótipo público que destaca os eventos e participantes numa história. Esta visualização permite um número de interações, incluindo o *zoom* num evento em particular, a filtragem de eventos, entidades, e intervalos temporais, e a edição do posicionamento de elementos na visualização. Este protótipo foi avaliado através de testes com utilizadores e um inquérito. Os testes foram realizados com cinco jornalistas, alguns dos quais participaram também nas entrevistas iniciais. Relativamente ao inquérito, este foi enviado para a comunidade académica da Universidade do Porto, tendo registado 178 respostas. Através do contacto com os jornalistas e com o público em geral, fomos capazes de

recolher informação sobre como os potenciais criadores e consumidores percecionam a aplicação. Os resultados demonstram que os participantes consideram o protótipo interessante, não obstante alguns problemas de usabilidade. Também recolhemos uma quantidade considerável de sugestões de melhorias que poderão servir de ponto de partida para futuros ciclos de desenvolvimento do protótipo. Através das respostas aos dois métodos de avaliação, concluimos que visualizações de cadeias narrativas podem ser aplicadas com sucesso a narrativas jornalísticas, mostrando-se como uma possibilidade viável para investigações futuras.

Keywords: Visualização de notícias, Visualização de texto, Interação humano-computador, Visualização de narrativas

Acknowledgements

Thank you to the team behind Text2Story for giving me the opportunity to work on this project and to the journalists who participated in the interviews and tests we conducted.

Thank you to Professor Sérgio Nunes for the invaluable guidance provided throughout this work and for putting up with my many questions. You were a glimmer of hope in a system in which I have been steadily losing faith.

Thank you to my family, without whom I would not be here today, and to all the people who stood by my side over these last few years. To my parents, my love for you is unwavering. Words cannot express the full extent of my gratitude, so I hope that through my actions, past, present, and future, I can show those around me how much they mean to me.

This work was developed in the context of the research project financed by the ERDF – European Regional Development Fund through the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 and by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia within project PTDC/CCI-COM/31857/2017 (NORTE-01-0145-FEDER-03185).

"Apathy's a tragedy and boredom is a crime
Anything and everything
All of the time"

Bo Burnham *in* Welcome to the Internet

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In today's world, we are subject to an uninterrupted stream of information on a daily basis. Part of this information consists of news articles of varying degrees of quality and relevance. The relentless pace at which news articles are communicated to the public often leads to an inability to process but a small fraction of them. Furthermore, not all news articles are unique nor relevant. This combination of excess, repetition, and questionable relevance places further stress on the reader, who must allocate a considerable amount of time to not only interpret the narratives within each news articles but to also carefully choose which articles are of interest. While the reader struggles to consume the news, the abundance of news ferociously consumes the attention of the reader, bringing about a need to allocate this attention efficiently [54].

The duality of what we are offered and what we are capable of absorbing can, in turn, translate into disinterest in trying to keep up with current events, as one may feel this cannot be accomplished without spending a considerable amount of time and effort (which one may not afford). The combination of a disinterested audience and noisy communication channels between the journalistic community and the general public can undermine the foundations of a democratic society, as uninformed citizens are more susceptible to the actions of malicious actors with little regard for democracy.

1.1 Motivation

We believe that one way to tackle this generalized disinterest revolves around minimizing the difficulties faced during the consumption of news. These difficulties can stem from two distinct stages of this process: the selection of news articles and their interpretation. In regards to the former, the sheer quantity of journalistic material and the frequent redundancy of news articles account for some of the major obstacles. As for the latter, not only must the reader determine whether or not a particular article is relevant, they must also make sense of the underlying narrative and contextualize it (i.e. understand its impact and its place in the web of overarching topics that

compose the news sphere). Conducting this process on a regular basis can be time-consuming, sometimes prohibitively so.

Some tactics are already used to reduce the cognitive burden often placed on the reader. High-impact news stories often receive specialized coverage that aims to summarize what is known thus far through various means (e.g. visualizations and video essays). This summary provides the reader with a centralized hub containing the most relevant information regarding a particular narrative. However, this approach has two clear disadvantages: it is a heavily manual process, and it tends to be reserved for influential stories. Smaller news articles local to a particular city or about which information is scarce are two examples of situations that may not justify such an ordeal.

We believe that there is value in the study of journalistic narratives through semi or fully-automated means that can be applied to news stories regardless of their perceived impact or their origin. The possibility of using an application to easily explore news stories through representations other than pure text can open up new doors in terms of how the reader consumes the news.

1.2 Goals

It is with this in mind that we propose the use of visualization to present the underlying narrative of a set of related news stories in a memorable and accessible manner. Taking into consideration a collection of news that are in one way or another connected to each other, the objective is to produce a self-contained visualization that succinctly brings forward narrative elements such as actors, events, and locations in a temporally ordered sequence. The work described in this document rests upon the following hypothesis:

Visualization contributes to a better understanding of the narratives behind news articles

1.3 Approach

The first step was to collect and review 17 information visualization papers related to news visualizations. This analysis allowed us to conclude that the narrative surrounding news stories is rarely explored in detail. Then, we conducted a set of interviews with experts in the field of Journalism in order to gather some insight into the role of visualization in understanding news content, as well as into the possible benefits and downsides of using visualizations with a more narrative-driven perspective. During these interviews, the journalists agreed that, if designed correctly, visualizations can be useful to the reader. Nevertheless, poorly-constructed ones may only add more noise to the news reading process. While not all news topics lend themselves to narrative-driven visualizations, it is possible to find news stories that resemble traditional narratives, thus enabling this type of visualizations. From these interviews, we moved on to the implementation of a prototype that focuses on storyline visualizations such as the one from Munroe [46]. Having structured data as input, the application produces a storyline which showcases the narrative's events and participants, allowing the user to interact with the visualization in a number of ways. The prototype

was then evaluated through user tests with five journalists and surveys sent out to the academic community of the University of Porto. In both the user tests and the surveys, participants showed interest in the application as well as concern in regards to its usability. The suggestions collected from these tests provide us with a foundation upon which to consider future work. Overall, the positive feedback gathered corroborates the potential which this type of visualization has to help news readers during the news consumption process.

1.4 Structure

The structure of the document is as follows. In Chapter 2, we conduct a literature review in order to understand how previous work has tackled the topic of news visualizations. Afterwards, Chapter 3 introduces our proposed solution to the visualization of journalistic content from a narrative-driven perspective. In Chapter 4, we analyse the content from interviews conducted with five journalists prior to the implementation of the prototype, described in Chapter 5. Finally, Chapter 6 goes over the results of the user tests and survey, with Chapter 7 providing an overview of the work described in this document as well as an exploration of potential paths for future work.

Chapter 2

News Visualization

News as the subject of visualizations is not a novel topic. The overwhelming growth of and ease of access to information has fueled research efforts that aim to help people make sense of it. While these efforts are by no means exclusive to the news domain, a considerable amount of the tools developed are generic enough to allow for their usage in this context. To understand what has been done in the field of news visualization, we have collected 17 papers that we believe to be a representative subset of the works that share a similar goal to ours: to aid in the comprehension of the news. Before diving into the selected works, we will briefly discuss the concept of visualization and some of its sub-fields as a means to contextualize the specific task of visualizing the news. Figures for the works not illustrated visually in this section can be found in Appendix D.

2.1 Visualization and sub-fields

Visualization is the act of creating a mental model in one's mind [55]. It is an activity intrinsic to the human species, and one which allows us to make sense of our surroundings. Moving over to the computational space, visualization tools try to elicit this human activity and promote the acquisition of insight. They structure complex and large amounts of data into palatable nuggets of insight. Since visualizing is a human activity, any mechanisms that attempt to seize the potential of our visualization capabilities are also constrained by the perceptual and cognitive quirks and limitations of the human mind [55]. From here on out, *visualization* and *visualization tool* will be used interchangeably, as we will navigate the computational field exclusively.

Turning our attention to the abstract nature of the information which we wish to visualize, one major challenge is made clear: the visual encoding of information which "does not have any obvious spatial mapping" [11]. Text in general, and news in particular, are relevant examples of abstract information without immediate mappings to the visual realm. A number of questions arise when we are faced with the task of visualizing text: which section of the text do we wish to

visualize? Do we want analyze the structure or the content? Do we want to understand the overall sentiment of the text or the narrative structure that guides it?

2.1.1 Text visualization

Text visualization is the sub-field of information visualization that aims to provide the necessary techniques that help process this textual information [10]. The most common text formats analysed in text visualization are documents, corpus (a collection of documents) and streams (continuous streams of textual information, such as Twitter feeds) [10]. More, various semantic units can represent a document; for example, we may consider paragraphs, sentences or the full text as distinct definitions of *document*. The documents can be fed to the visualization in their original form, or after going through some sort of text processing [36] (e.g. named entity recognition or TF-IDF weighting). One further characteristic of the input is its domain. The possibility for data sources are quite extensive, in part due to the sheer amount of information available in this era of information overload.

Text visualization techniques can be placed under four different categories, according to the subject of the main underlying analysis task: document similarity, text content, sentiments and emotions, and exploration [10]. Techniques for visualizing document similarity represent documents as points in a 2D or 3D visualization plane, with the distance between two points encoding their similarity (the closer, the more similar). The analysis of textual content encompasses a number of tasks such as summarization, word-level analysis, topic visualization, event visualization, and storyline visualization. The last three tasks are particularly common in the papers reviewed and will be discussed in greater detail in later sections. Sentiment and emotional visual analysis techniques can be seen in the work of Wanner et al. [61], in which news feeds related to the U.S presidential election are automatically processed to extract positive and negative opinion words. Finally, document exploration techniques attempt to offer the user a set of tools that allows them to conduct an efficient analysis of the data. For example, ParallelTopics [15] (illustrated in Figure 2.1) offers multiple coordinated views to this effect.

2.1.2 Narrative visualization

Ware [62] introduces two distinct uses of visualization: as a means to explore data and as a means to explain data. Visualizations can be used not only to discover patterns, but to also expose those patterns to other people and persuading them into accepting your interpretation. When presenting results, the author of the visualization should be "in control of the cognitive thread" [62]. For this to happen, the author presents the audience with a sequence of visual patterns and words. The audience's attention will be controlled by the presentation, "and any attention controlling sequence of information is a form of narrative" [62]. While *narrative visualization* is not mentioned, the idea of a visualization tool which aims to guide the reader and convince them to accept the author's interpretation of the underlying information as valid ties in with subsequent discussions surrounding this field.

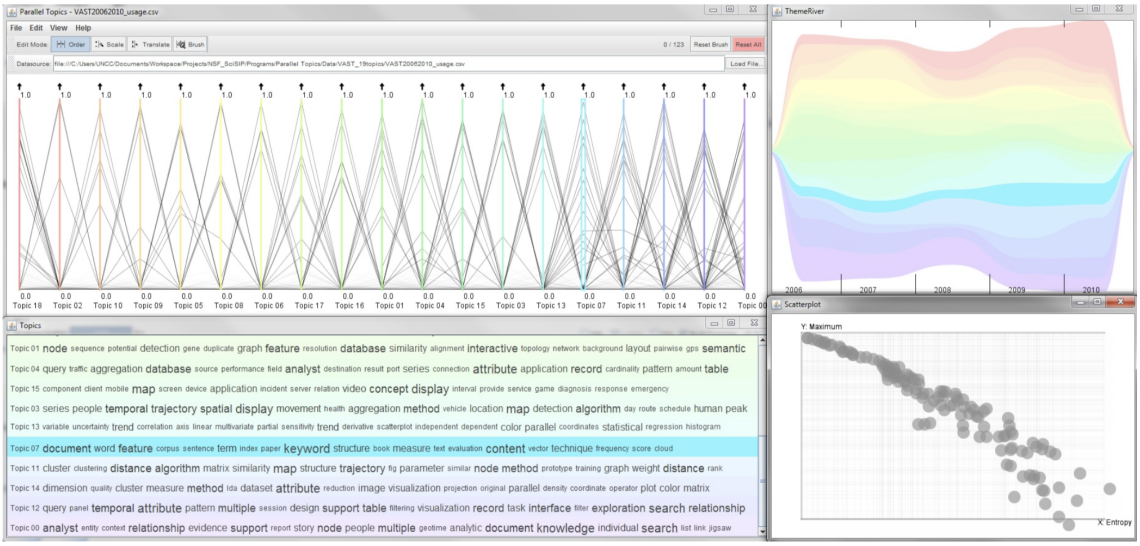


Figure 2.1: Multiple coordinated views in ParallelTopics [15].

Gershon and Page [23], in one of the earlier works related to the inclusion of narrative in information visualization, reflect on how storytelling can facilitate the capture of information encoded in a visualization. They define the value of storytelling in information visualization in terms of its ability to compel the audience and, ultimately, speed up their understanding of the often complex data behind said visualizations. Segel and Heer [50], in one of the seminal works of narrative visualization, review the design space of "data stories", visualizations that present a story while still allowing for interaction and alternative explanations. Hullman and Diakopoulos [29] echo the definition of narrative visualization proposed by Segel and Heer [50]: that of a visualization that combines communication and exploration as a way to convey a story and to guide the end-user in the process of obtaining insight. Dove and Jones [18] build upon the works of Segel and Heer [50] and Hullman and Diakopoulos [29] to further reflect on the role of narrative visualization and what makes it stand out from other forms of visualization. Narrative visualization is characterized as the incorporation of interactive data visualization techniques in storytelling. The authors propose that one of the distinguishing features of this type of visualization is its ability to both communicate insight as well as the experience leading to it, allowing the reader to put themselves in the shoes of the author and understand the process through which insight was built without going through the mental strain behind its creation. It is the combination of exploration and narrative that enables "authors to communicate discoveries more successfully by reproducing something of the experience of finding them" [18]. Figueiras [20] also stands behind the idea of narrative visualization as a way through which the end-user can make sense of the overwhelming amount of data to which we are exposed. It is through the combination of text and images provided by narrative visualizations that we can overcome the limitations of its individual parts. Once more, storytelling is seen as a way to facilitate understanding of complex topics.

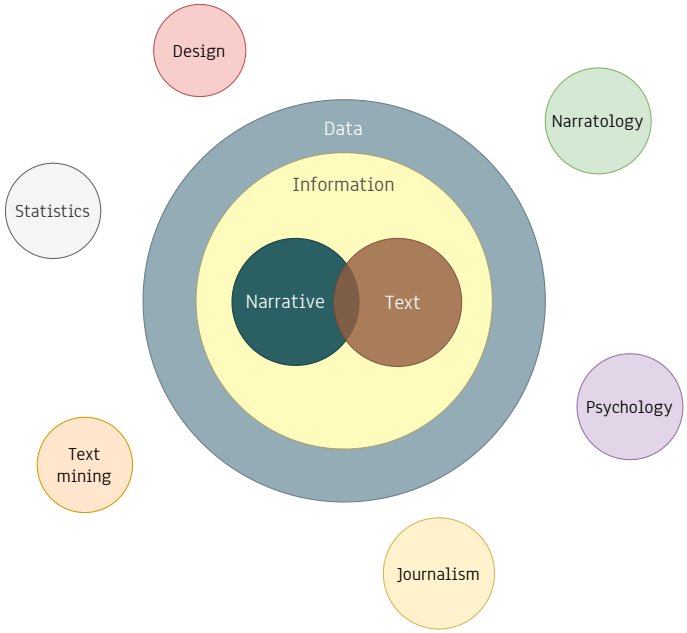


Figure 2.2: Some of the domains directly and indirectly tied to visualization.

2.2 Related areas

Although the focus of the works we will analyze falls under the umbrella of text visualization, we would be remiss if none of the multitudes of concepts surrounding visualization were to be mentioned. One of the fascinating characteristics of the field is the need for cooperation. To build a useful visualization tool, we must incorporate techniques from a vast range of knowledge areas. It is inherently interdisciplinary since it is often necessary to understand both the underlying content of the visualization, as well as the context in which the visualization tool will be used and the factors that affect it. Figure 2.2 aims to showcase a small fraction of this diversity.

On the one hand, the concept of visualization is vast, with multiple sub-domains tackling a wide array of occasionally overlapping topics (e.g. narrative and text visualization). On the other hand, other fields, which may not be directly related to visualization, have a tremendous impact on the design of a visualization tool (e.g. text mining and journalism). While this document's scope does not allow us to cover these fields extensively, they are crucial in contextualizing our work, inspiring discussion, and consolidating the diversity and multidisciplinarity of the craft and science of visualization. Thus, we have chosen to highlight some which came about during our research.

Starting with an extensively mentioned initiative in the selected papers, Topic Detection and Tracking (TDT) was a research program started in 1998 that addressed the ”event-based organization of broadcast news” [1]. The TDT initiative's initial motivation was to provide a system that would allow for the monitoring of broadcast news and the subsequent discovery of latent events taking place worldwide. It is one of the earliest efforts in the study of large news corpora and a cornerstone of text mining applied to the news domain.

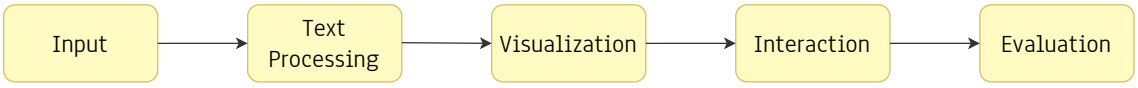


Figure 2.3: Pipeline of the development of a visualization tool..

From text mining to the study of narrative structures, Vossen et al. [60] adopt the narratology framework of Bal and Van Boheemen [2] as a way to describe a story, a key concept upon which their system is based. This is a distinct approach from the majority of the papers studied, which tend to set aside less abstract and semantically richer concepts tied to narrative structures in favor of a wider and more generic analysis of the news space, typically through topics and their evolution. We believe that it would be interesting to see more research efforts towards understanding how narrative structure can be applied to the news and how visualization can adequately convey this more focused and concrete representation.

Finally, the connection between news visualization and journalism, albeit not articulated often in the works reviewed, provides an important opportunity for collaboration. As we will discuss in Chapter 4, journalists are well-suited to provide first-hand reports of how a journalistic narrative is structured and which elements may or may not be relevant to a visualization. Acquiring insight into the inner-workings of news creation can help us plan an implementation more in tune with the needs of real-world readers.

2.3 Methodology

The papers were collected through the use of Google Scholar using the following terms: “text”, “visualization”, “narrative”, “storyline”, “news”, “topic”, and “analysis”. This querying process resulted in a large body of papers, which we then narrowed down. The criteria for this refinement process were as follows: the paper should partly or exclusively focus on visualization methods, the paper should partly or exclusively focus on the news domain, and only one paper per on-going project was selected (i.e. an overarching project such as TIARA [42] would be represented by only one paper). In addition to this broader search, we also performed a recursive exploration through the papers that appeared to be more in line with our objectives (e.g. LeadLine [16] and Story Tracker [35]).

The 17 selected works cover a wide range of publications. However, two venues are extensively represented in this survey, accounting for 8 of those 17 papers: IEEE Transactions on Visualization and Computer Graphics (5 papers), and IEEE Conference on Visual Analytics Science and Technology (3 papers). Some of the venues more closely tied to the news, such as the Workshop on Computing News Storylines and the Events and Stories in the News Workshop, encompassed few works tied to visualization. In regard to the temporal dispersion of the papers, the range covered allows us to both historically contextualize the need to understand large amounts of information, as well as to understand the evolution of specific techniques.

The visualization tools studied are the result of a pipeline composed of the following steps: input (i.e. what do we want to understand?), text processing (i.e. how can we transform the input into a visualization-friendly format?), visualization (i.e. which visual encoding is the most appropriate for our goals?), interaction (i.e. how can the user communicate with the visualization?), and evaluation (i.e. does our system perform as expected? Does it accomplish our goals?). While the papers' coverage of each topic may differ, all papers describe at least one of them. Understanding how previous works approached this process, their limitations and shortcomings, as well as their contributions, allows us to build a strong theoretical foundation upon which to base our work. We will cover all sections of this pipeline; however, for our purposes, the last three tasks are the most relevant (as Figure 2.3 suggests).

2.4 Studies

A common metaphor in the analysis of topical trends is that of a river's flow, introduced by Havre et al. [27] in the early 2000s and used extensively since (e.g in ParallelTopics [15], LeadLine [16], and TextFlow [13]). The ThemeRiver visualization allows for the analysis of thematic trends evolving over time, having as input a large collection of documents. Themes are represented as color-coded currents in the river, whose flow from left to right indicates the passage of time. The vertical width of each current represents that particular theme's strength at a given timestamp; consequently, the river's vertical width is the collective strength of all themes. An example of the river visualization used in ThemeRiver is available in Figure 2.4. In addition to the river flow, the visualization also contains event annotations, providing more context to the thematic changes. This type of visualization eases the detection of patterns and trends in the data. The authors propose that visual metaphors leverage the "perceptual and cognitive abilities" of the user when presenting data to ease the process of obtaining insight. They also justify their choice of metaphor by transferring psychological knowledge to their domain. Humans tend to favor the big picture as opposed to low-level details, ultimately forming objects from individual parts. Several factors influence these objects' construction, such as "similarity, continuity, symmetry, proximity, and closure." ThemeRiver's design leverages these features as a way to promote the communication of patterns in the data.

As ThemeRiver has suggested, the need to make sense of large amounts of information is not a recent phenomenon. Another example of some earlier work in this domain is Contexter, a visualization tool designed to aid in the analysis of extensive collections of news stories [26]. The system extracts three representations from the input documents: a full-text representation, a bag-of-words representation, and named entity sets. Subsequently, the visualization system provides several distinct views (simultaneously displayed in a grid-like structure): the set of all extracted entities (from which the user can select one), a graph representing the local context of a selected named entity (i.e. other named entities which appear in the same documents as the selected ones), a set of characteristic words (i.e. with higher TF-IDF values) from the documents which contain

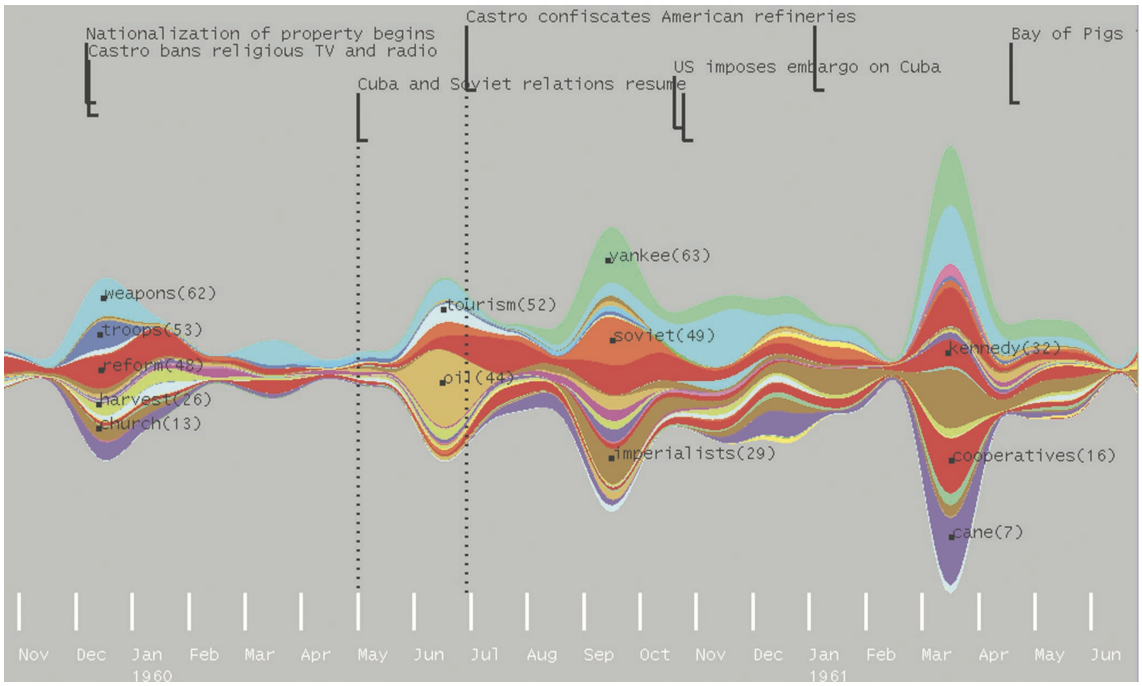


Figure 2.4: River visualization of ThemeRiver [27].

the selected named entity, and a set of the keywords most frequently associated with the selected named entity.

Fisher et al. [21] aim to let the users explore how concepts in large text streams evolve and how they relate to each other in their visualization tool, Narratives. This relation can be established by studying the topical keywords associated with news events. The work focuses on blog entries that mention news stories, so there is an indirect connection with the news domain. While the term narrative is frequently used throughout the paper, the underlying information does not resemble a narrative in the word's ordinary sense. Instead, the system studies concepts in the news sphere across a particular time frame and the most commonly associated keywords. Narratives uses a line graph to depict the evolution of keywords' popularity. The user can specify these keywords through the use of a search bar. The horizontal axis represents the time frame, and the vertical axis represents the percentage of events in the database. Each point in the graph is the percentage of events mentioning a particular keyword on a particular day. Clicking on a point displays a list of the most common keywords to the term the user searched for on that particular day; keywords are sized according to relative frequency. Furthermore, peaks in the graph are annotated with the most popular keywords. The user can also explore the correlation between the terms for which they searched. Searching for pairs of keywords tells the system only to consider times where the terms co-occurred. In addition to the several ways correlation is studied, the user can also set up alarms when events containing a particular keyword arrive at the system, which triggers sound and visual cues. A dropdown list of the top 20 terms at multiple time intervals is also available. Lines for related terms can be plotted on the same axes or in parallel. The authors briefly present two

possible use cases and a similarly brief comparison with another tool designed for the visualization of topics in news articles, IN-SPIRE [28].

TIARA is an interactive visual analytics tool aimed towards exploring large document collections [42]. From the features produced by a Latent Dirichlet Allocation [4] model, the visualization tool builds a summary of the underlying information. The authors refer to the adopted visualization technique as a stacked graph. However, the actual implementation closely resembles a river visualization (e.g. ThemeRiver [27]). As such, for the purposes of describing this work, we will use the terms "river" and "stacked graph" interchangeably. The river visualization is accompanied by keyword clouds that encode the topic and content evolution over time. Several interaction techniques are available to the user should they want to drill-down to obtain more information. Three possible interaction categories are introduced: topic details on demand (e.g. using a "magic lens" to expand the number of keywords displayed in a particular section), text information on demand (e.g. contextualizing a keyword with text snippets of relevant documents), and coordinated multi-view analysis (e.g. visualizing relevant meta-data); the latter is explored in a domain-specific manner, namely in the context of email and company reputation analysis. TIARA is evaluated both through the study of use cases (more specifically, with an email analysis scenario), as well as through user tests which measured both objective (i.e. answer completion rate, answer time, and error rate) and subjective (i.e. usefulness, usability, and system satisfaction) measures. Ten participants were recruited for the aforementioned test.

Luo et al. [44] take a more event-driven approach and propose EventRiver, a system that allows the user to detect, track and explore a text collection with temporal references through the integration of text analysis and visualization. Figure 2.5 provides an overview of the system's interface. To extract these events, the authors assume that events are the catalysts of document generation and that documents reporting on the same event will be conceptually similar and close in time. Clusters of documents are thus characteristic of events. These clusters can be further grouped to reveal long-term stories of related events that are not necessarily close in time to one another. The events are encoded as bubbles in a river, whose width represents its duration and the height its relevance measured in terms of document quantity. Events belonging to the same long-term stories have the same color. These event bubbles are laid out in a horizontal axis depicting time, and the vertical position of the bubbles indicate different stories. Events in a particular story are vertically adjacent. The height of a story denotes its importance (which can be determined using a set of user defined criteria). In addition to the bubbles, each event has two textual annotations placed in parallel that contextualize it within an overarching story and highlight its characteristic keywords. Furthermore, the user can also trigger an additional contextual window that provides a few sentences that further characterize the event. On the topic of interactions, the authors split the interactions according to user intent: browsing, search, and investigation. While browsing, the user can filter out events according to their influence, remove/include events based on selection and time frame, sort according to relevance (determined by multiple criteria), and manually change the events' position. When searching, the user can look for a keyword or select an event, highlight other events with similar content. Finally, when investigating an event, the user can click

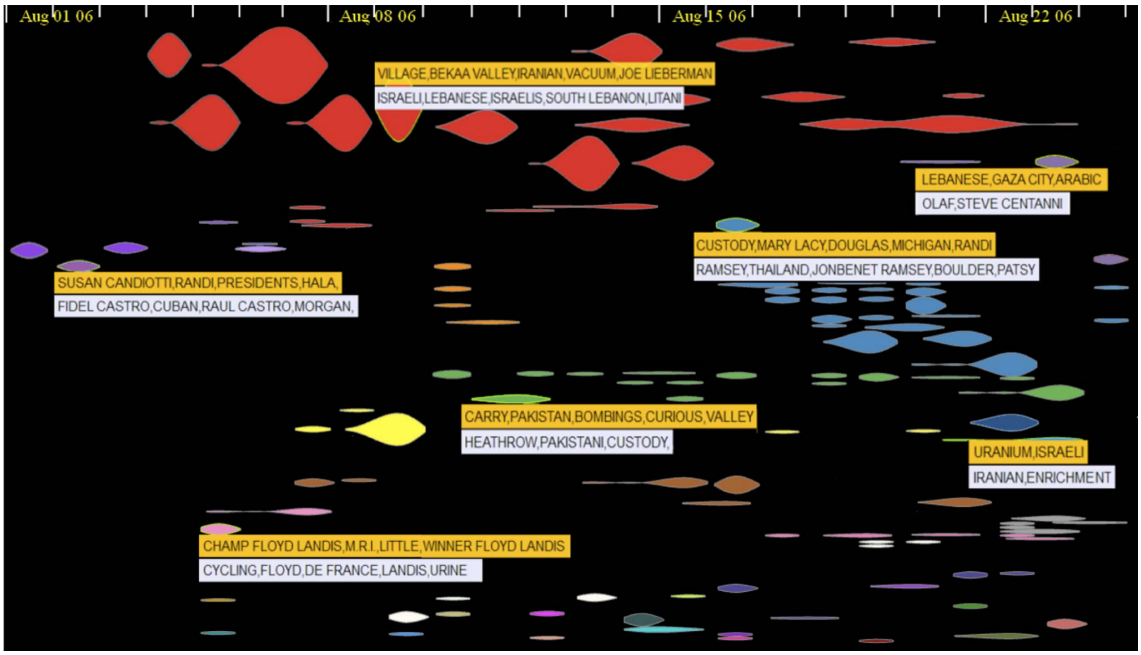


Figure 2.5: Interface of EventRiver [44].

on it and explore an additional view that provides added contextual information, including snippets and full text of the articles which originated the events. The system was evaluated through multiple techniques: case studies, user studies, and experiments to test the efficiency and efficacy of the algorithm (as well as to compare the performance with other tools, namely IN-SPIRE [28] and LensRiver [25]).

Krstajić et al. [33] turn their attention to the visualization of news streams. One of the main obstacles of news streams is the real-time nature of the data, which inevitably translates to uncapped growth of the information displayed. This growth forces some information to be excluded to avoid cluttering. The authors propose the use of article threads that retain only the most critical information overtime on a particular subject. The evolution of a thread depends on the number of articles, its duration, and its age. Threads that are "old, short, and sparsely populated" are deleted. Groups of threads belonging to a particular topic are displayed vertically and parallel to each other. Each thread is color-coded according to its number of articles, following a scale that goes from grey to pure red. The only textual information available is that of the topic labels beside each group of related threads. The authors did not conduct any study to evaluate the system.

While on the topic of streams, CloudLines is presented as a solution to handle large quantities of dynamic, time-series event data [34]. The authors utilize timeline distortion techniques to prioritize recent data and interaction with individual events that are closer in time. A Focus+Context approach is adopted to access events in densely-populated sections. The system allows the user to extract an overview of the data while at the same time making an atomic analysis of events possible. More recent events, which are deemed more important, are given more real-estate; older events are demoted but remain present to contextualize recent information. Similar to the article

threads [33] approach, related events are grouped together, vertically parallel to each other. A circle with a minimum size represents each event to allow for interaction and visibility. These circles can overlap and are positioned horizontally according to their time-stamp. The user can zoom in on densely populated sections to analyze individual items. Clicking on an item will display more information (e.g. in the case of a news article, the URL of said article, its title, and its image). How this information is displayed is not explained. The authors propose two use cases and discuss them in some detail.

ParallelTopics is a system designed to explore large document collections and their topical distribution using interactive visualizations and probabilistic topic modelling [15]. One of the leading sets of questions that the authors aim to answer concerns the "relationship between topics and documents". The topics extracted using Latent Dirichlet Allocation are then fed to a system composed of four coordinated views: a tag cloud of topics (in which each line depicts a topic and the associated keywords, scaled according to frequency and ordered from left to right according to relevance), a parallel coordinates graph that represents the "probabilistic distribution of documents across topics", a scatterplot that allows the user to identify multi and single-topic documents more efficiently, and a river visualization based on ThemeRiver that presents topic evolution over time. There is a comprehensive set of interactions available. The views are coordinated in the sense that a selection in one of them translates into a selection of the corresponding data in others. Two case studies were conducted to explore two distinct text corpora related to scientific publications. While the paper does not explicitly cover the news domain (thus violating one of the criteria for selecting publications), the coordinated views approach seemed particularly relevant.

Cui et al. [13] propose TextFlow, a topic mining and visualization system that facilitates the understanding of topic evolution in text data. Three sets of features are extracted from the data: topic evolution trends, critical events, and keyword correlations. The visualization process encodes this information using topic flows, glyphs, and threads, encoding information at different granularities. Figure 2.6 presents an example of a TextFlow visualization in the context of the news. Interacting with the system helps "refine the analysis results and gain insight into the data progressively". In addition to the previously mentioned visualization techniques, TextFlow also provides tag clouds and timelines as complements, intending to provide detailed textual information. The topic mining model obtains topic merging/splitting patterns through an incremental Hierarchical Dirichlet Process (HDP) and subsequently extracts critical events and keyword correlations. The varying flow-height corresponds to the number of documents in a particular topic. Topic flow can be split or merged. Topics are aligned based on timestamp and ordered by a flow algorithm developed by the authors. Each topic is color-coded (merged topics are colored by combining the parents' colors). Threads are lines between two time-points where a keyword is used; weaving threads denote the co-occurrence of keywords in wave bundles. The amplitude of these wave bundles encodes the number of occurrences of the keywords in them. Two interaction types are provided to the user: hovering (provides simplified information on the element in question) and selecting (additional information is made available). Performing a selection also allows the back-end model to recommend appropriate topic flows to highlight and draw co-occurrence keyword patterns related to the

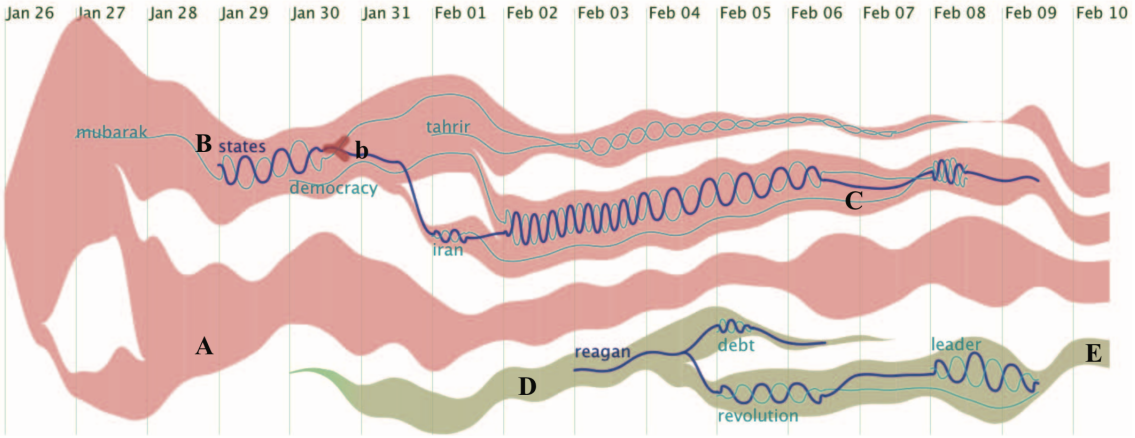


Figure 2.6: Example of a TextFlow visualization of Bing news data [13].

selected keyword. This visualization system is complete and, as one of the case studies appears to indicate, suited for the analysis of news stories. The river-flow metaphor positively contributes to the visual appeal. However, that same case study also points to an abundance of short-lived topics that do not lend themselves to this visualization type. Besides, topic merging/splitting, while present, is not common.

Dou et al. [16] explore two distinct sources of text data (online news and Twitter feeds) from an event-centered perspective, "identifying meaningful events in the news and social media data" [16]. One of the fundamental ideas behind this work is the belief that "bursts" of information have corresponding triggering events. There are three main components to LeadLine: topic identification using Latent Dirichlet Annotation (LDA), Early Event Detection "to identify the temporal scale for events", and named entity recognition to extract people and locations related to an event. These methods output the attributes which characterize an event: who, what, when, and where. In addition to this characterization, LeadLine also provides a visual interface that allows for the exploration of these attributes and the adjustment of algorithm-specific properties that alter the output of the topic and event extraction systems. Figure 2.7 showcases some of the multiple views available in LeadLine.

Dou et al. [17] explore the relationship between topics with HierarchicalTopics, a visual analytics system that allows for the representation and navigation of a large number of topics. The system creates a hierarchy using the Topic Rose Tree algorithm proposed by the authors (based on the Bayesian Rose Tree [6]). The output of this process is then represented visually through the use of two techniques: a river and a topic tree (in which the leaves are the tag clouds of keywords in each topic). This combination of visual representations brings forward the topics' content and hierarchy, as well as their temporal evolution. Moreover, the two views are coordinated, thus facilitating the exploration of the interface. A number of interactions are made available to the user, from the more common zoom and pan operations, to a magnifier (to enlarge words in the tag cloud within a certain radius of the mouse cursor) and a highlighter (to allow for multiple instances of a keyword in the various tag clouds to be detected). The user can also modify and update the topic

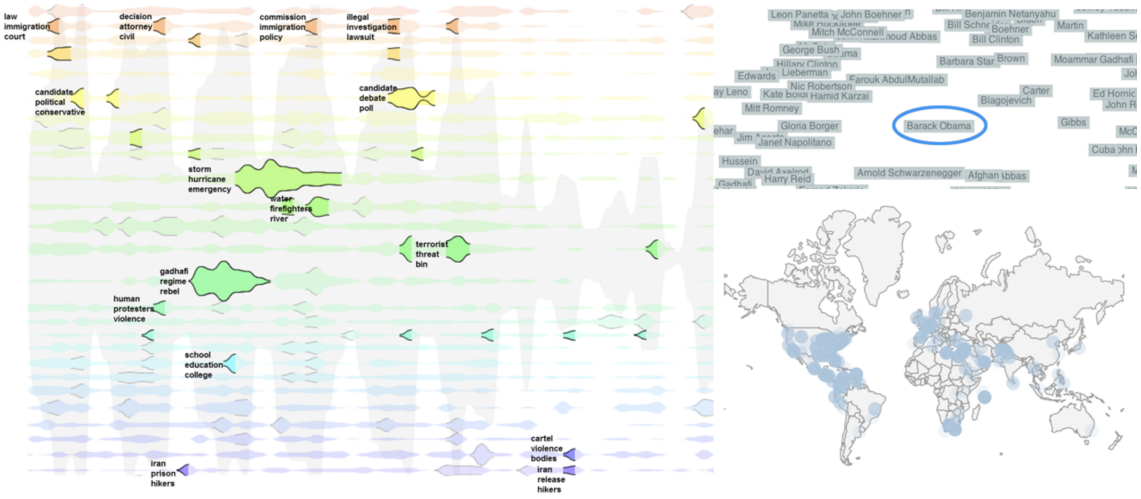


Figure 2.7: Multiple visual representations provided by LeadLine [16].

tree in accordance to their mental model of the topics. Furthermore, the information can also be filtered temporally, and the underlying textual content is also made available according to the selection made in the river view. The authors conducted a case study (with one expert with research experience in the fields of Human-Computer interaction and Information Retrieval), and a user study with eighteen participants, in which quantitative measures were taken and the opinions of the participants were registered.

Shahaf et al. [51] propose zoomable metro maps as a way to understand the information landscape. Like other works, they cite the need to make sense of complex stories "in an era of information overload" as the driving force. Metro maps are "structured summaries of information" which aim to expose connections between documents in a collection. The different levels of zoom provide users with distinct granularities with differing levels of detail and interaction. The authors expand upon their previous notion of metro maps of information: metro lines which follow distinct, coherent narrative threads, connecting metro stops (clusters of words that frequently coexist in a given period). Zooming in or out influences the scope of the metro stop. The proposed map structure accommodates different story structures (i.e., it is not exclusive to linear stories). Finally, the implemented algorithm for generating these metro maps is scalable since it depends linearly on the document set's size. The authors compare their implementation with other systems. Twenty-seven users were randomly assigned two summarization systems out of four to learn two stories. After using the systems, the participants were then asked to write brief paragraphs about what they learned from those two stories. Amazon Mechanical Turk was used to evaluate the paragraphs; the workers were also asked to justify their choice.

Krstajić et al. [35] leverage the versatility of parallel coordinate graphs in Story Tracker, a visual analytics system that allows for observing the temporal evolution of news stories in dynamic information streams. One of the reasons behind the project is the pace at which the news cycle renews itself, making it difficult for readers to establish a connection between current and past

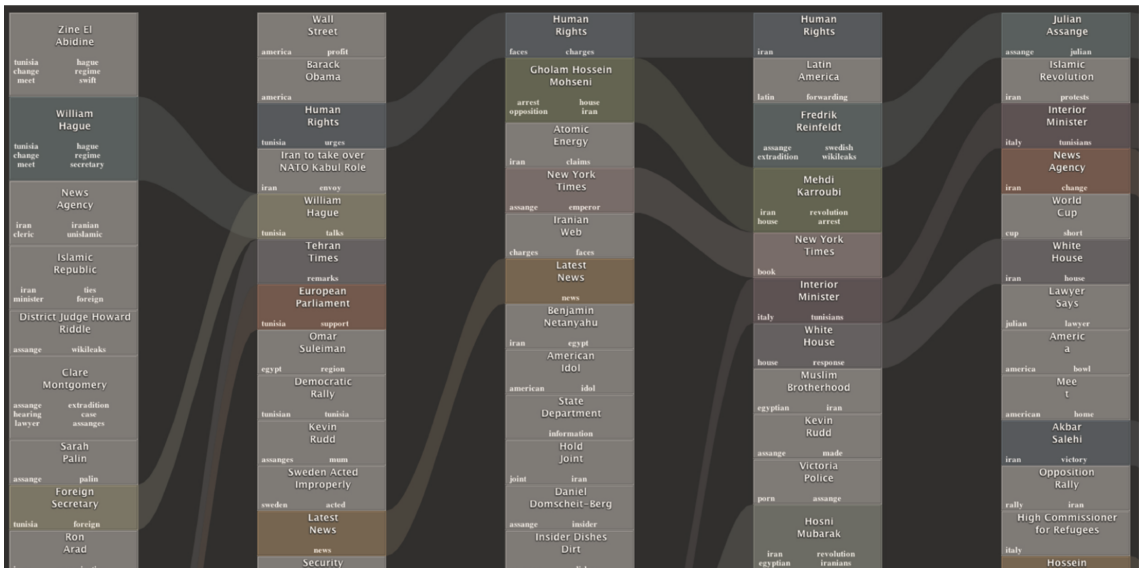


Figure 2.8: The main view of the Story Tracker visualization tool [35].

events. The authors use text mining and interactive visualization techniques to accomplish their goals. The visualization provides three distinct levels of zoom, with increasing levels of detail. First, the user is presented with an overview of all stories extracted and the connections between them; at this level, no textual information is provided. This main view is encoded in Figure 2.8. Next, the main view zooms in on a particular time frame, adding textual information to the stories (e.g. title and most important keywords). The height of each story's rectangles (blocks) corresponds to the number of articles in that story. The third zoom level, the zoomed view, allows the user to explore a particular story and its evolution. Finally, the article view provides detailed information on a particular article (e.g. full text, related entities, and images). The authors use the news stream provided by the European Media Monitor, focusing exclusively on news articles written in English. The clustering process relies on two algorithms for clustering search results: Lingo (extract frequent phrases) and Suffix Tree Clustering ("assumes similar documents share identical phrases"). The documents are clustered in 24-hour windows. Topic overlap, merging, and splitting is addressed through the comparison of clusters in adjacent positions. Interestingly, the authors justify the small time frame of comparison (of only one day previous to the current one) partly because "in most cases, stories appear on consecutive days, without long breaks between them". Other authors try to accommodate for these long breaks (e.g. [60]). The visualization interface allows the user to change various parameters of the pipeline, such as the algorithm (Lingo or STC), the input (metadata types), sorting criteria (cluster strength or the number of articles), and cluster labels (algorithm labeling or most important document). Additionally, they can filter data based on keywords and time frames. Regarding testing methodology, a user study with eight expert participants was conducted.

Vossen et al. [60] borrow some concepts from the literary world and present a formal model for encoding storylines extracted from massive streams of news. Based on the narratology frame-

work of Bal and Van Boheemen [2], the authors construct their model around the concept of a *fabula*—"a sequence of chronologically ordered and logically connected events involving one or more actors". Three distinct moments characterize a *fabula*: the exposition, the predicament, and the extrication. While the authors claim that these moments can be identified in every narrative text, they also warn that news articles are not a typical narrative. Several reasons are attributed for this: the focus on climactic events, the need to publish frequently and quickly (often with little information available), the speculation inherent to the gaps in the narrative and successive additions of information as the story unfolds throughout multiple articles. The authors utilize a Natural Language Processing pipeline which generates a semantic representation of the content according to the Simple Event Model (SEM) [58]. Participants and events are bound to instances mentioned in the document through coreference relations. The events are anchored either to a time reference in the document or to the document publication time. The result is a timeline of events, from which a story is subsequently extracted. This extraction consists of selecting the most relevant events. The events are ranked according to prominence (based on the number of mentions); the most prominent events become the climax. Next, the events are placed before and after the climax according to the bridging relations' strength (the stronger they are to the climax event, the closer they'll be). The process repeats itself for the events that are not connected to the climax.

Tannier and Vernier [57] propose an authoring and visualization tool for thematic timelines. Keywords and time frames are transformed into a timeline graph that encapsulates chronological information and the importance of the events that occurred on a particular day. The pipeline follows this simplified structure: the extraction of documents that conform to the user's query, the extraction of dates from within those documents, date ranking (determining their importance), the association of a textual description to an event, and the visualization of the extracted events. While the paper's topic certainly pertains to what we are trying to accomplish, it has considerable limitations: only one event can be detected per day, and events can only last one day. Any sense of story continuity is thus lost, and we end up with a series of disconnected events from which we cannot extract an overarching story.

Van Meersbergen et al. [59] aim to tackle rich text interpretations with StoryTeller, a visual analytics tool for complex multi-dimensional textual data. One example of this would be news stories, which, according to the authors, "contain about 200 event mentions on average". StoryTeller adopts an event-centric approach, grouping events with multiple participants into storylines. The visualization provides three main views: participant-centric (co-participation graph based on the movie narrative charts of Munroe [46]), event-centric (temporally ordered events in rows), and data-centric (text snippets used to extract the events). We can observe the co-participation graph in Figure 2.9. The StoryTeller system was designed to follow the seven tasks of the "visual information seeking mantra" designed by Shneiderman [53]: overview, zoom, filter, details-on-demand, relate, history, and extract.

Finally, Laban and Hearst [37] developed newsLens, a system to collect, extract, and visualize news stories. Their contribution is two-fold: the aggregation and organization of a large, multi-source dataset of news articles into a collection of major stories and the naming and visualization

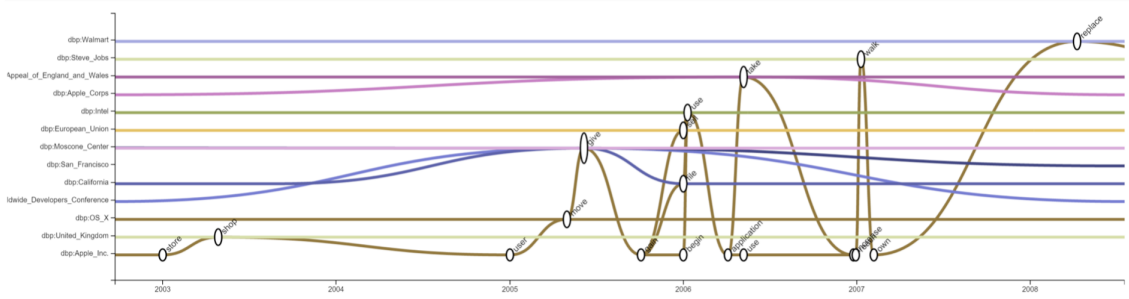


Figure 2.9: The Co-Participation graph of the Storyteller visualization tool [59].

of said stories. The visual interface consists of a series of lanes, organized temporally, where each lane represents a story and is annotated with relevant information. The authors aim to help the reader consume complex, long-lasting stories in a more efficient way. In order to generate the topics, the pipeline extracts keywords from the articles. Each document is represented in a bag-of-words vector, and a TF-IDF transform is applied. A keyword is selected if its TF-IDF score is higher than a threshold. Then, the articles are clustered into local groups called topics. Two articles are in the same topic "if they share several keywords and are published in a close range of time". The topics are then merged into long-ranging stories. The authors point out three important phenomena in story generation from topics: linking ("assigning a topic from a preceding graph to a topic in the current graph"), splitting (divide one topic into two distinct ones), and merging (joining two distinct topics if a current cluster contains articles previously assigned to two distinct old topics). A story is the product of successive linking and merging operations. These stories are named by selecting the highest-scoring noun phrase from all headlines; this score is calculated using a series of features determined by the authors (such as the presence of common or proper nouns).

2.5 Discussion

As we have previously mentioned, all of the papers reviewed cover one or more of the topics that compose the visualization pipeline proposed. We will briefly cover each section, exploring some of the most common choices, what is or is not in line with our goals, and what are some of the gaps in the study of news visualization.

Concerning input, all works deal with large collections of documents (be it news articles or not). Table 2.1 summarizes these findings. Although there are some differences between static and dynamic (i.e. streams) information sources, the scale at which they are consumed by the visualization pipeline is undeniably large. On the one hand, this bulk analysis can provide contextual information that is difficult, if not impossible, to obtain through individual documents (e.g. trends in news topics and the popularity of certain themes). On the other hand, introducing the possibility of a fine-grained exploration of the news (e.g. relationships between the entities mentioned in a news article) may allow the audience to capture relevant knowledge that would otherwise slip by

unnoticed. Providing the user with more perspectives on the underlying information allows for a complete understanding of the domain.

	Input		**Domain**	
	Collection	**Stream**	**General**	**News**
ThemeRiver [27]	✓		✓	✓
Contexter [26]	✓			✓
Narratives [21]		✓	✓	✓
TIARA [42]	✓		✓	
EventRiver [44]	✓		✓	✓
Krstajić et al. [33]		✓	✓	✓
CloudLines [34]		✓		✓
ParallelTopics [15]	✓		✓	
TextFlow [13]	✓		✓	✓
LeadLine [16]	✓		✓	✓
HierarchicalTopics [17]	✓		✓	✓
Metro lines [51]	✓		✓	✓
Story Tracker [35]		✓		✓
Vossen et al. [60]		✓		✓
Tannier and Vernier [57]	✓			✓
Storyteller [59]	✓		✓	✓
newsLens [37]	✓			✓

Table 2.1: Input and domain of the papers reviewed.

Regardless of the scale of the input, it frequently undergoes some processing to reduce its dimensionality and extract the desired features. One common thread to the aforementioned papers is the indifference towards richer semantic features of the underlying text. The techniques used to extract what will ultimately be visualized vary (e.g. topic modelling and correlation); however, they mostly revolve around relatively simple representations of the input document (e.g. bag-of-words). For the purposes of a more abstract analysis of the data (e.g. a topic-centric approach), this representation often suffices. By contrast, should we want to look deeper into the document's semantics (e.g. to explore its narrative structure), a more complex representation is necessary.

Having processed the documents, we now move on to the heart of the visualization pipeline: the visual representation. An overview of the visualization techniques used can be found in Table 2.2. One of the characteristics that nearly all of the works reviewed take into consideration is time. The passage of time is vital in understanding the evolution of a text corpus, particularly the news. For example, knowing the number of news articles related to a particular topic that were published during a specific time window allows us to infer when that topic was more or less relevant. However, we can analyze time from another perspective, one not tied to the publication or creation moment, but to the story within an individual document. This view opens up new possibilities for the fine-grained exploration mentioned previously, and is something to keep in mind when proposing new ways to visualize the news.

One would assume that papers that propose to handle events would operate at a finer level of granularity than those which deal in topics. Intuitively, topics tend to define a broader web of concepts that tie a set of documents together along a particular theme. Indeed, sometimes that is the case, such as with the work of van Meersbergen et al. [59]. The authors defend that the use of clustering techniques such as topic modeling do not lend themselves to describe stories in detail [59]. To accommodate for these details, they use a storyline chart in a "participant-centric view". However, in other visualization tools, such as EventRiver [44], events are just as abstract as topics, relying solely on keywords and neglecting a more story-like structure. Once more, this type of approach is conducive to a higher-level analysis of the underlying information and not to a more thorough examination.

We cannot pinpoint one dominant visualization technique, although the use of a river metaphor is present in a considerable portion of the papers. Timelines are also a popular choice, along with pixel-based techniques. One interesting takeaway is the use of visualizations that rely on textual content (e.g. tag clouds) as complements to the primary representation (e.g. in HierarchicalTopics' [17] topic cloud, and TextFlow's [13] tag cloud). While these visualization techniques fall short in conveying temporal information, they are valuable tools in understanding the gist of a collection of documents (or a single document). To explore a body of text more thoroughly, they may not be the best fit.

From simple line graphs (e.g. Narratives [21]) to complex rivers that split and merge (e.g. the work of Cui et al. [13]), the techniques used to achieve the shared goal of "making sense of large amounts of information" cover a broad spectrum of complexity. This diversity can be attributed not only to the representation chosen but also to the number of views provided to the user. Works such as that of Dou et al. [15] present various coordinated views into the underlying data, contrasting with the somewhat frugal pixel-based technique of Krstajić et al. [33]. However, more perspectives do not always translate into a complete understanding of the information. Efficiently transmitting insight requires a balance between simplicity and completeness.

The catalog of visualization techniques collected, from parallel coordinates graphs (e.g. StoryTracker [35]) to metro lines (e.g. the work of Shahaf et al. [51]), and everything in between, is by no means extensive. Nonetheless, it provides us with some insight as to how a large-scale exploration can be conducted, and it raises one important question: how might these techniques perform if given different, semantically richer intermediate representations? Furthermore, will they still efficiently convey information at a smaller scale (e.g. at the level of a single document)?

Interaction is similarly at the core of visualization, playing a role equally as important as that of the visual representation itself. In fact, Card et al. [11] defines interaction as a necessary element to information visualization. It is the dialog between the system and the user [14], the features through which said user is allowed to manipulate and interpret the representation, directly or indirectly [63]. We have considered seven distinct categories, as per the taxonomy of Yi et al. [63]: *select* (i.e. "mark something as interesting"), *explore* (i.e. "show me something else"), *reconfigure* (i.e. "show me a different arrangement"), *encode* (i.e. "show me a different representation"), *abstract/elaborate* (i.e. "show me more or less detail"), *filter* (i.e. "show me something condition-

	Line graph	Scatterplot	Node-link	Parallel coordinates	Tree	Timeline	Pixel-based/Area	Tag cloud	River	Map	Metro map	Storyline
ThemeRiver [27]									✓			
Contexter [26]			✓									
Narratives [21]	✓					✓						
TIARA [42]			✓						✓			
EventRiver [44]							✓					
Krstajić et al. [33]							✓					
CloudLines [34]							✓					
ParallelTopics [15]		✓		✓				✓	✓			
TextFlow [13]							✓	✓	✓			
LeadLine [16]							✓	✓	✓	✓		
HierarchicalTopics [17]					✓			✓	✓			
Metro lines [51]											✓	
Story Tracker [35]				✓					✓			
Vossen et al. [60]		✓					✓					
Tannier and Vernier [57]						✓	✓					
Storyteller [59]						✓	✓					✓
newsLens [37]						✓	✓					

Table 2.2: Visualization techniques provided by the papers reviewed.

ally"), and *connect* (i.e "show me related items"). To understand how these categories can be used to classify a particular visualization tool, we will briefly analyse one of the works which apply techniques covering this spectrum of interaction, LeadLine [16]. Alongside this analysis, we also provide an overview of the distribution of aforementioned techniques throughout the works in Table 2.3.

First, LeadLine provides the user with several representations of the input data, including a tag cloud, a map with locations mentioned in the articles analyzed, a stacked graph overlaying a river visualization of the topics in the news collection, and an entity graph (*encode*). When hovering over an element in one view (e.g. a topic in the topic stream), the corresponding elements in the other views are highlighted (*connect*). Clicking on a section of the topic stream allows the user to access the documents represented by that area (*abstract/elaborate*). Selecting named entities will highlight events from different topics (*select*). Through the use of a slider, the user can determine the level of "sensitivity" in the event detection process, thus controlling the events' granularity (e.g. only consider big spikes in topic activity as an event) (*reconfigure*). Finally, the selection made in one view determines what is or is not visualized in another (*filter*), and the exploration is accomplished through the dynamic addition and removal of entities from the entity graph according to the selection in other views (*explore*).

Unlike with other sections of the pipeline, it is difficult to determine whether or not specific techniques are conducive to high-level analysis. The categories we have considered appear to be scale agnostic, in the sense that all seven of the enumerated tasks are relevant both when working with large-scale, abstract representations, as well as with smaller, more semantically complex structures. Just as with visual representations, the choice of interactions that are made available to the user must be dictated by where we wish to be in the author-driven and user-driven spectrum (a dichotomy thoroughly studied by Segel and Heer [50]).

	Select	**Explore**	**Reconfigure**	**Encode**	**Abstract/Elaborate**	**Filter**	**Connect**
ThemeRiver [27]				✓			
Contexter [26]	✓			✓			✓
Narratives [21]					✓	✓	✓
TIARA [42]	✓	✓		✓	✓	✓	✓
EventRiver [44]	✓	✓	✓		✓	✓	
Krstajić et al. [33]						✓	
CloudLines [34]		✓			✓		
ParallelTopics [15]	✓			✓	✓	✓	✓
TextFlow [13]	✓			✓	✓	✓	✓
LeadLine [16]	✓	✓	✓	✓	✓	✓	✓
HierarchicalTopics [17]	✓	✓	✓	✓	✓		✓
Metro lines [51]		✓			✓		
Story Tracker [35]		✓	✓		✓		
Vossen et al. [60]	✓				✓		
Tannier and Vernier [57]		✓	✓				
Storyteller [59]	✓	✓		✓	✓	✓	✓
newsLens [37]		✓			✓		

Table 2.3: Interactions provided by the papers reviewed.

Finally, we collected the evaluation methods used in the papers reviewed and classified them according to the categories present in Table 2.4. Unlike the survey of Wanner et al. [61], case studies encompass both evaluations conducted solely by the researchers and those where domain experts or users were also involved. Some works do not conduct any evaluation process or do not provide enough evidence of the proposed method's efficacy (Wanner et al. [61] classify these works under "Anecdotal evaluation"). Case studies appear to be the most popular evaluation method; however, the methodology behind these case studies differs between projects. For example, while Dou et al. [16] recruited users to explore the system and provide feedback, Cui et al. [13] conducted the case studies internally. Algorithm comparison was used in three of the research

papers; the lack of adherence to this method may be explained by the difficulty in comparing visualization methods. Often, the proposed methods are either: not similar enough to other methods, or there are no directly comparable methods. In the case of Narratives [21], the authors explicitly say that the visualization tool is not directly comparable with other methods; despite this, they perform a brief side-by-side exploration with a similar tool. Algorithm performance and ground truth comparison were the least common methods. For the purposes of this discussion, we consider "algorithm performance" evaluations to include information on computation performance, as well as algorithm correctness (e.g. whether a pre-determined set of themes was discovered in a data set). A focus on usability metrics over raw performance data may be behind the low adoption rate for these tests. A considerable number of papers (five) do not conduct any evaluation method (according to the criteria set above). For the evaluations that involved external participants, Dou et al. [17] was the one with the highest adherence, at 27 participants (the participants were all associated with Stanford University and compensated for their time).

	Method					**Participants**	
	Case study	**User test**	**Algorithm comparison**	**Algorithm performance**	**Ground truth comparison**	**Number of participants**	**With experts**
ThemeRiver [27]		✓				2	
Contexter [26]							
Narratives [21]	✓		✓				
TIARA [42]	✓	✓	✓			10	
EventRiver [44]	✓	✓	✓	✓	✓	12	
Krstajić et al. [33]							
CloudLines [34]	✓						
ParallelTopics [15]	✓	✓				5	Yes
TextFlow [13]	✓						
LeadLine [16]	✓					8	
HierarchicalTopics [17]	✓	✓	✓			18	
Metro lines [51]	✓	✓				27	
Story Tracker [35]	✓					8	Yes
Vossen et al. [60]							
Tannier and Vernier [57]							
Storyteller [59]	✓						
newsLens [37]							

Table 2.4: Evaluation methodology of the papers reviewed.

This overview of the selected papers allows us to extract some possibilities for further research.

As mentioned before, the papers favour a "big-picture" approach as opposed to a more detailed perspective. The news are analyzed in bulk, losing their individuality and corresponding fine-grained characteristics to allow for a more abstract portrait of the underlying information. This zoomed-out view can undoubtedly help us glean useful insights into the news content; however, we may want to explore news articles in detail occasionally. Richer intermediate representations that are more semantically-aware than techniques such as bag-of-words can open up new doors towards this goal. One particularly promising framework is that of Discourse Representation Theory (DRT), which relies on the assumption that a mental representation is built iteratively as discourse is processed [24]. Exploring how to visual encode the Discourse Representation Structures (DRS), which are the output of this process, can shine a new light onto the content of the news.

The term "news story" is frequently used in the literature; nevertheless, story-like characteristics and elements are rarely taken into consideration (the works of Vossen et al. [60] and van Meersbergen et al. [59] are some of the exceptions). While news articles have distinct characteristics from other bodies of text produced in a literary context, they report on real-life events, communicating to their readers some story through a journalistic narrative. Moreover, these individual reports, when approaching a particular topic over time, can form an overarching narrative that is only made clear when sufficient information has been gathered. Visualizing these narratives irrespective of scale is a possibility that has so far gone relatively unexplored.

Despite not being included in the aforementioned visualization pipeline, defining the requirements of our visualization tool is just as important as the implementation itself. If we claim our goal is to help the public understand the news more efficiently, then we must study what the public needs. What do people value more? Exploring the big picture or analyzing the details? What do journalists believe is the best way to approach this? Research techniques such as surveys and expert interviews are ways to obtain answers to these questions and certainly showcase some potential for possible contributions to this field.

2.6 Visualizing narratives

Few of the works reviewed take into account the journalistic narrative, the written manifestation of the news story. Elements such as events, actors, interactions, and physical settings often go unexplored and their potential untapped. Although relatively obscure in the news domain, the visualization of narratives has been a steady presence in other fields. Gershon and Page [23] reflect on how storytelling can ease the absorption of information present in a visualization. Many other since have studied the power of narrative in information visualization (e.g. Segel and Heer [50], Hullman and Diakopoulos [29], Dove and Jones [18], and Figueiras [20]).

Narrative visualization, a term popularized by Segel and Heer [50], tackles this interplay between narrative and visualization. However, as discussed by Padia [48], introducing storytelling

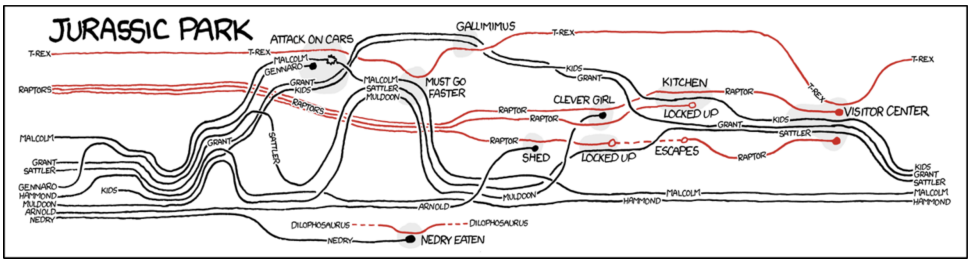


Figure 2.10: Narrative chart of the movie “Jurassic Park” produced by Robert Monroe [46].

elements in visualization is different from visualizing the narratives themselves. Whereas narrative visualization has been the subject of a considerable amount of academic projects, efforts into the study of these visual narratives are scarcer [48].

One technique originated by Munroe [46] (showcased in Figure 2.10), the storyline visualization, attempts to visually encode narrative elements in a way that succinctly encapsulates the flow of a story through the manifestation of the interactions between its intervenients and the major events in which they partake. Entities are represented by lines that evolve from left to right, horizontal and parallel to each other. Adjacent lines indicate an interaction between the corresponding entities and events are encoded by nodes. This technique has been used extensively, both in more general applications [43, 56], as well as in specific scenarios such as meeting summarization [52], the evolving interactions between developers in software projects [47], and the analysis of plots in literary works [31]. Despite the potential, these techniques usually simplify the problem domain by imposing application-specific constraints and cannot accommodate for issues such as nonlinearity, single-participant events, and multiple timelines [48].

Other approaches to the topic of visual narratives attempt to tackle these issues; one such example is that of Story Curves, “a visualization technique for exploring and communicating nonlinear narratives in movies” [32]. The authors point out some of the shortcomings of storyline visualizations, namely their inability to encode a character’s absence (since the entity lines are continuous). The definition of ‘story’ and ‘narrative’ is also discussed. While the two terms are often used interchangeably (as is the case with this document), their meanings differ significantly. One the one hand, *story* refers to “what is told”; on the other hand, *narrative* refers to the way that content is told, to “how it is told”. While the chronological order of the events is referred to as ‘story time’, the order in which these events are narrated belongs to the ‘narrative time’. Transposing these concepts to the news domain, we can conclude that often the narrative order does not follow the story order. For example, a news articles begins by reporting an event and is then followed by other articles which expose events that lead up to the initial occurrence. The impact of non-linearity in comprehension an visualization of news stories is not something in which we will focus; it is, nonetheless, a relevant characteristic of this medium. Furthermore, while we acknowledge the formal differences between narrative and story, we will continue to use the term ‘news story’ when referring to journalistic narratives.

Regardless of the differences between narrative visualization and the visualization of narratives, the two concepts are not necessarily disjoint. A visualization which aims to expose the intricate narrative structure of a story can be a narrative visualization. Moreover, the application of techniques such as storyline visualizations in the news domain presents a fertile ground for research, and one that can help us understand if this more narrative-driven approach can indeed facilitate the comprehension of news stories.

2.7 Alternatives to traditional news formats

So far, the majority of the papers reviewed have tackled news stories that follow a traditional textual format. However, in recent years, there have been some research efforts into alternative means to produce journalistic content; some of these efforts are carried out by the BBC News Labs. The BBC News Labs have, for some years now, conducted several experiments into what they refer to as "structured journalism", a "structured approach to reporting the facts and constructing journalistic narratives" that allows for a scalable and versatile production of news stories [39].

One of the initiatives within the broad "structured journalism" umbrella is that of the News Storyline Ontology [3], which aims to describe and organize news stories. Another, Graphical Storytelling, explores the semi-automatic creation of news comics from text, originally intended to convey stories related to health in an "accessible and appealing way" [38]. These initiatives (among others within the same topic) are relevant sources of insight when conceiving a system that allows for the visualization of journalistic content and bring about important discussion topics. How can we take advantage of these alternative representations? How is the transition from text to visual currently handled (e.g. concerning the Graphical Storytelling project)? Are these scalable approaches?

Despite the experimental nature of the projects mentioned, this does not detract from their potential for real-life impact. The prototypes of tools such as Graphical Storytelling may be the standard in news rooms in a few years. As such, just as we must understand how news visualization took place in the past, we must also be aware of signals that indicate a change in the paradigm of journalistic content production.

2.8 Summary

The study of the 17 selected papers highlighted some common characteristics and goals. First, there is an overwhelming focus on large-scale analysis, where the news articles are mostly explored as a whole, thus losing some of details in the individual items. Second, the use of simple intermediate representations, such as Bag-of-Words vectors, are predominant, resulting in a lost of semantic complexity (e.g when a document is narrowed to a set of topics to which it may belong). Finally, a more narrative-driven perspective, in which narrative elements such as actors and interactions are placed at the forefront, is rarely adopted (with a few notable exceptions, such as the work of Vossen et al. [60]). From these conclusions, we have constructed some questions that will

serve as motivation for the remainder of the work: *is a small-scale analysis of the news (e.g. at the document level) useful? How can we visualize semantically richer intermediate representations? How can we adopt a perspective conscious of the narrative elements of news stories? What is the opinion of journalists in regard to this topic?*

Chapter 3

News from Different Perspectives

Having analyzed a representative set of papers tied to the visualization of the news, we are now better equipped to determine what the knowledge gaps are, what we can contribute to the field, and where our work might fit in. The hegemony of topic-based approaches in the study of large news collections leads us to question what other techniques would be adequate to perform this study. Furthermore, can we effectively analyze the news on a microscopic level, as opposed to the more common macroscopic solutions? In an attempt to answer these questions, we propose the implementation of a system that extracts elements usually associated with more traditional narratives from news articles and creates a visualization centered around these concepts. Figure 3.1 illustrates the dichotomy between the works we reviewed and what we intend to explore. We will first dive deeper into the problem definition, followed by a more thorough explanation of what we have done to tackle it.

3.1 Problem definition

Based on the literature review, we concluded that, while there is a considerable number of visualization tools built to study the news (or built generically enough that they allow for their use in this context), the scale of this analysis is frequently large. The interest lies mostly on artefacts that can only be extracted from large collections of news (e.g. topics or topic-driven events). The details of individual news articles are set aside in favour of a more holistic view of the news panorama. A single news article may be reduced to a set of keywords or topics to which it belongs, placing it alongside all other news articles which share those keywords or topics. These approaches allow you to quickly get a grasp of past or current trends; the user can ingest a quantity of news that would otherwise be, at the very least, considerably time-consuming to process manually.

But what if we want to look at less news, in more detail? What if we want to understand the story behind those news articles at a finer level of granularity? The visualization tools we have looked at so far rarely meet any of these two requirements. On the one hand, the features which

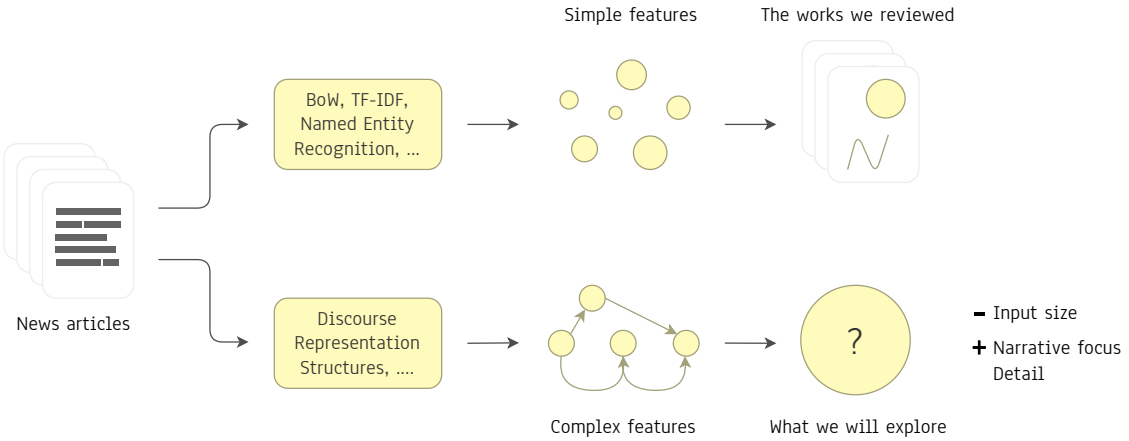


Figure 3.1: Comparison between the works reviewed and our goals.

are extracted from the news articles are relatively simple (e.g. keywords or entities); on the other hand, the collections used are large, usually in the thousands or even millions of individual news articles. While this type of approach is desired for some analytic tasks, we argue that there is value in exploring different ways of approaching the problem of news visualization, namely from a more narrative-driven, smaller-scale perspective.

To better visualize the differences between these two approaches, will we briefly explore two scenarios: a small collection of articles pertaining to a hypothetical ongoing trial case, and a single news story that reports on a hypothetical political scandal. For these two scenarios, we will explain why visualization has potential to help the reader understand the stories more efficiently, and how the two opposing approaches would deal with them.

Let us assume that our hypothetical trial case has been ongoing for a few years now, and a small collection of news articles pertaining to it has been steadily emerging. By now, most readers, while having a vague idea of what the trial is about, may find it difficult to know which stage of the trial is currently taking place, who has been judged for what crimes, what were the events that lead up to the trial, and so on. Scouring through that collection of articles and trying to piece together the narrative manually can be prohibitively time-consuming for most. Introducing a visualization tool that could condense those articles into an informative summary can help the readers situate themselves in the narrative without demanding too much from them.

A topic-based approach might, for example, tell the reader that the initial articles report on the investigation and that later articles pertain to the trial itself, which entities are mentioned in a certain time frame, and the frequency of articles related to a particular topic along time. However, characteristics such as the connection between the entities evolved and more detailed events that make up the timeline of the narrative (e.g. person X colluded with person Y to scam person Z) are left out and the reader is left with a somewhat shallow analysis of the underlying narrative. A narrative-driven approach might try to tackle these questions, exploring in greater detail who did what, with whom, where, and when.

As for the case of a single news articles such as the one we described above, there may be too little content for the system to extract any meaningful topics. Switching our focus to the most relevant keywords in the document, they may tell us who is involved in said scandal, but issues such as the narrative flow and the interaction between entities remain undiscovered. Once more, a visualization tool that has a greater focus on the details enclosed in the underlying narrative might be more useful in this context. For example, whereas the paragraphs explaining the events might not be ordered chronologically, the visualization can extract the most relevant events in those paragraphs and rearrange them in a more intuitive fashion in a timeline. Since some details can slip by, and others would be hard to encode visually, a system such as this one may serve as a complement to the written article, thus providing the reader with more ways to interpret the narrative.

From these two examples, we can hypothesise that the techniques we have looked at so far are not adequate for certain analytical tasks conducted at a smaller scale with a focus on the narratives. We have briefly ventured into the topic of narrative visualization in Chapter 2, and in the subsequent chapters we will explore the subject from a practical point-of-view. From the information we have gathered so far, we are left with this question: *how does a more narrative-driven visualization of the news look like?*

3.2 Proposed solution

To explore the problem space a bit further, we conducted a series of semi-structured interviews with professionals in the field of Journalism, described in Chapter 4. Our focus during these interviews was bipartite. On the one hand, we wanted to understand the written journalistic narrative a bit further. On the other hand, we wanted to gauge the journalists' perception of visualization in the news, and whether or not they believed it could be a helpful complement to a news story. From here, we implemented a storyline visualization focused on highlighting the sequence of events in a story, alongside its participants and their interactions. The input data for the visualization was manually produced based on existing news articles, as our focus was not on the automated textual analysis of said articles. This visual metaphor is present in works such as StoryFlow [43] and Storyteller [59]. With our prototype, we wanted to expand the interaction possibilities and enrich the visualization enough contextual information to allow for both an overview of the story, as well as an in-depth analysis of it. To assess the efficacy of the implementation in helping users explore a news story, we conducted a user test with five journalists, as well as a survey targeting the academic community of the University of Porto. Splitting the evaluation process in such a way allowed us to collect both in-depth testimonies of experts (i.e. journalists) in direct contact with the tool, as well as the broader, less detailed perception of a more varied pool of participants that closely resemble the average user.

In summary, we collected information on the journalistic narrative and the role of visualization in the news through semi-structured interviews with five journalists. Next, we implemented a prototype centered around a storyline visualization which aims to highlight narrative elements in

a news story, such as entities, events, dates, and locations. Finally, we conducted user tests and a survey to assess the efficacy of the prototype in facilitating the analysis of news stories.

Chapter 4

Exploring the Problem Space

Reviewing papers that have already tackled the visualization of news from multiple perspectives has allowed us to understand what some of the most common solutions are, their advantages and disadvantages, and their adequacy when considering a more narrative-driven perspective. Furthermore, it also lead us to conclude that the journalistic take was rarely considered. We believe that journalists, the creators of the content which we are trying to visualize, can provide crucial insight into the fundamentals of journalistic narratives.

Firstly, they can help us characterize the impact of information overload, both on the news room as well on the readers. Journalists and other content creators are at the forefront of this information crusade. Their active role in the dissemination of large amounts of information juxtaposes their constant struggle to capture the readers' attention, which is tugged at uninterruptedly by the amount of information made available. More, many journalists have themselves come across or been responsible for the creation of visualizations and, as such, can provide a first-hand account of their efficacy in the real-world. Finally, they are particularly familiar with the structure of journalistic narratives and are primed to understand, for example, whether or not they may lend themselves to visualization or what the most relevant characteristics to encode visually are.

These interviews fall roughly under the umbrella of the discovery phase found in the development of UX projects [49]. In it, the goal is to find and frame the problem and to explore what the opportunities might be. The discovery phase excludes the validation of hypotheses or solutions. Our approach does not strictly follow this criteria, as some of the questions are directly tied to the use of visualization as a potential solution. Nonetheless, the exploratory nature of these interviews is still present, particularly since visualization is a very broad field with countless research opportunities.

4.1 Methodology

Five semi-structured interviews were conducted through video calls. The interviewees are (or were) professionals in the field of Journalism with several years of experience at the time of the interviews. The group was made up of three men and two women. In addition to their experience

Code	Example
Helps situate the reader	*This tells people where they are in a story, I find it particularly useful in these complex narratives.*
More captivating	*If the visual narrative is adequately designed, it can captivate readers in a way different from text.*
Generates more interest	*Visual stories are often the ones that generate the most interest.*
Occasionally better than text	*I think it would help the reader in situations were it is harder to communicate solely through text.*
Allows for interaction	*When it is possible to add interactivity, it helps a lot.*

Table 4.1: Codes associated with the advantages of visualizations in the news.

in Journalism, two of the interviewees were also part of the academic sphere related to social sciences. Each interview had a duration of approximately 45 minutes and was centered around the set of questions showcased in Appendix A. The questions aimed to explore three main topics: the consumption of news and its main obstacles (e.g. information overload), the journalistic narrative and its main characteristics, and the use of visualization as a way to consume the news. The outline of the interviews was adapted iteratively as the they took place, to reflect the insight we had gathered thus far. The collected interviews were then assessed using thematic analysis [8]. The themes were determined in a data-driven fashion, in which the transcripts were analysed, extracting any relevant codes and grouping them into appropriate themes, sub-dividing or merging them whenever the grouping unit was too broad or too narrow. From this analysis we extracted five themes: advantages of visualizations, disadvantages of visualizations, retrospection and day-to-day analysis, structure of news articles, and how people consume the news. Throughout this chapter, we have included paraphrases from the journalists to better illustrate the arguments made. The original transcripts can be found in Appendix A.2.

While we tried not to tie the interviews to specific visualization methods and techniques, we believe it was important to situate the journalists and provide them with concrete examples of the types of visualizations we had in mind when designing the interviews. We presented them with two visualizations, one belonging to Cairo et al. [9], and another to Munroe [46]. With the first visualization, we wanted to present an example of an higher-level analysis, a visualization which focused more on topics rather than on specific stories, whereas with the latter we wanted to highlight a more focused, narrative-driven approach. The choice of examples could have influenced the journalists' opinion and conditioned their ability to generalize and consider other hypothetical visualizations. Nevertheless, we believe that the benefits outweigh the costs, as not providing any real examples and working purely on an abstract level could mean that two journalists answered the questions with two fairly distinct ideas of what a visualization might look like.

4.2 Advantages of visualization in the news

Table 4.1 presents a summary of the findings for this section. Journalists generally agreed that the use of visualization in the news has potential to help in the exploration of journalistic narratives. Visualizations can help the reader to locate themselves in a story, particularly when it is complex. They are also likely to engage the reader more efficiently in particular contexts for several reasons. Certain topics in which the chronological component of the events is relevant lend themselves to visualization. One hypothetical example given during the interviews was that of a news story that covers the first year of the Covid-19 pandemic. While text could be used to describe the related events, an interactive visual timeline would be more appropriate, as it would allow the reader to interact with the story in a number of different ways, choosing what is and is not relevant to them.

> *For example, a story focused on the first year of the pandemic would work much better in the shape of a visual timeline rather than in pure text, which would undoubtedly be very boring.*

A written report would be constricted by certain journalistic guidelines (e.g. the most important events should be written first) which would break the chronological order of events, and sometimes this order is desirable. Furthermore, exclusively using text could result in a "dull" article which fails to captivate the reader.

> *In a journalistic text, we avoid going from A to Z, we try to start with the most important information. Generally, we do not follow a chronological order, but often a timeline is the most effective.*

Journalists also pointed out that visual stories draw people in through more diverse ways than plain text, increasing the chances of the reader consuming that particular piece of content. In the same vein as the previous comment, another journalist claims that, in addition to conveying information, visualizations can add an emotional layer to a story. However, they point out that this can also be achieved through text.

> *These visual narratives can catch us off-guard and surprise us. Sometimes they are designed in a way that 'speaks' directly to the reader.*

4.3 Disadvantages of visualizations in the news

Table 4.2 presents a summary of the findings for this section. While the potential of visualizations in the news was recognized by all journalists, some warnings were also issued. One of the most frequent concerns is that of the learning curve necessary to interpret visualizations. Some of the visual codes used in visualizations such as the storyline of Munroe [46] (e.g. the correspondence between a continuous line and the passage of time) may not be immediately understood by readers.

> *Some of these things require a level of abstraction which some people do not have.*

Code	Example
Steep learning curve	*There needs to be a learning stage for this type of visual representation.*
Potentially noisy	*If I cannot immediately understand what a visualization is trying to communicate, then it produces noise.*
May not fit everyone's style	*Personally, I think that my rhythm of consuming information does not allow me to properly analyse this type of visualizations.*
Requires more resources	*How much time is needed to do something like this? We have a lack of resources and I question its viability.*

Table 4.2: Codes associated with the disadvantages of visualizations in the news.

Furthermore, an unclear visualization can lead to additional cognitive strain. Instead of helping to capture information, a poorly constructed visualization can hinder the interpretation of the underlying story, adding noise to the news reading process. The lack of quality in a visualization can stem from a number of factors. In addition to the previously mentioned learning curve, one of the issues most frequently brought up is the question of extraneous visual elements. A visualization, as with any journalistic product, should first and foremost tell a story, focusing primarily on the content and not on the presentation. As one journalist said, the form should submit to the function.

> *If they are too pretty for the sake of being pretty, then they create noise and distract us from the story.*

> *The form should work for the story, not the other way around.*

One journalist expressed their concern in regard to the visualization they were shown, the storyline of Munroe [46]. They claimed that the visualization not only did not help them decode information regarding the movie's narrative, it "complexified" the story as the journalist knew it. The information (i.e. how the action unfolded) is merely shown, and it does not help the reader interpret it.

> *This visualization does not help me feel less 'saturated', it does not help me decode anything. It does not even help me interpret [the story], it just shows it to me.*

People like the aforementioned journalist may not benefit from visualizations, as their consumption habits do not allow them to process them adequately. Finally, visualizations require more resources to be produced than written articles, resources which may not always be available.

4.4 Retrospection and day-to-day analysis

Table 4.3 presents a summary of the findings for this section. Journalists generally agreed that visualizations such as storylines can potentially work both as tools to understand current events,

Code	Example
Hard to updates stories online	*It is very hard to keep readers engaged with an online story that needs to be updated periodically.*
Need distance from the facts	*I think we need distance from the facts to understand what is important.*
Need a complex story	*To do something like this, the story needs to have a lot of ´threads'. A single-day story or event generally does not generate that.*
Possible to accommodate for both	*I think we can use this technique in both situations.*

Table 4.3: Codes associated with the discussion regarding the use of narrative-driven visualizations as an exercise of retrospection or as a day-to-day tool.

as well as past occurrences. One journalist argued that the difficulty in captivating readers to an ongoing story can work against the use of these visualizations in recent, developing stories. With these types of stories, the visualizations would potentially need to be updated often, to reflect new findings or correct past assumptions. This would, in turn, require the reader to periodically check up on the story, which is something that cannot be expected of the average consumer. One other argument against visualizations targeting more recent stories pertains to the relevance of the documented events. Allowing some time to pass between the occurrences and the reporting can help us define what is or is not relevant to the story, as well as determine if the observations remain accurate or whether new findings invalidate past assumptions. For a visualization focused on narrative elements to work, the story needs to be "meaty", which is usually not the case for recent or very recent events. Nevertheless, the adequacy of narrative-driven visualizations should be considered on a case-by-case basis.

4.5 How people consume the news

Table 4.4 presents a summary of the findings for this section. When asked which approach would be more interesting to the average news reader, a bird's eye view of the news or a detailed analysis of particular news stories, the responses were mixed. Some journalists argued that visualizations

Code	Example
High-level analysis	*If we look at the average consumer, they are much more likely to pick a visualization like Alberto Cairo's.*
Detailed, focused analysis	*From the perspective of the reader, it seems to me that visualizations that focus on a story and that help the reader understand it are more useful.*

Table 4.4: Codes associated with the readers' preference for high-level analysis or a detailed, focused one.

Code	Example
Inverted pyramid	*Speaking of news articles, they usually follow the inverted pyramid rule, which goes from the most to the least important information.*
Dynamic narrative focus	*The focus is different in different moments. The first narrative focus is on factual data. Then, some context starts to show up. Afterwards, I start to add 'density' with personal accounts.*
Can be similar to traditional narratives	*The order of the events is different, but the news story itself can have all elements of a good narrative.*
The "human factor" is important	*I create a connection with my reader or listener by transforming a generic story into a personalized one.*
Function over form	*I think that the idea that the form should submit to the function also applies to Journalism*

Table 4.5: Codes associated with the characteristics of news articles.

such as that of Cairo et al. [9] might be more useful to the journalists, whereas more focused approaches would better fit the needs of the average reader.

> *Generally, those high-level stories are more interesting to the journalist or to a very specific audience, they are almost meta-stories.*

On the other hand, one journalist proposed that the average news reader would skew more towards a broader visualization, that would allow them to get an idea of which topics are trending. If our focus shifted towards more "dedicated consumers", who may feel a stronger connection to a particular news outlet, then they would more likely appreciate a focused visualization on a particular subject.

> *An avid reader with a subscription [to a newspaper] is more likely to enjoy [a story-driven visualization] and appreciate the creativity behind that particular news story.*

4.6 Structure of the news articles

Table 4.5 presents a summary of the findings for this section. This theme highlights some of the characteristics of a journalistic narrative found in a news article, along with some guidelines that should generally be followed in when producing this type of content. Some journalists described the classic structure of a news article, the "inverted pyramid", in which the narration of the events should be ordered by relevance, with the most relevant topics at the start of the article.

> *In Journalism, the idea is to 'spoil' the story right away.*

As a consequence of this, the chronological sequence of the events may be lost, and, as we have mentioned in Section 4.2, this ordering is occasionally beneficial to the comprehension of the story. Although the order of narration differs somewhat from what is traditionally seen in other literary genres, the journalistic narrative can hold elements which are closely tied to traditional narratives. When news stories are spread out across articles, the focus will most likely change over time. One journalist described a three-step process, in which the focus will initially lay on the facts, followed by a contextualization of the event, and ending with the inclusion of more subjective, personal accounts of the events. This structure allows for the journalist to establish a connection with the reader by gradually narrowing the focus of the story and personalizing it. This "humanization" can attract the attention of the reader, and is one of the fundamental rules of journalism, according to the journalist in question.

> *Journalism anchors itself on the human element; that is a fundamental rule of journalistic work.*

As with visualizations, news articles should also focus primarily on telling a story, without superfluous elements clouding its interpretation. Here too, function trumps form.

4.7 Discussion

The majority of the journalists held a favourable stance towards visualization in the news, with one journalist expressing considerably negative feelings towards them. Nonetheless, all agreed in some way or another that visualizations, if properly constructed and used in the appropriate contexts, can be useful to the news reader. Despite its advantages, the use of visualizations should abide by certain guidelines, and failing to do so can turn an opportunity of enlightenment into a dreadful cognitive exercise. An unclear visualization that does not immediately convey information about the story will add noise to the news consumption process, the opposite of what we seek to do. This lack of clarity can stem from a number of different factors, two of which being the visualization's complexity, and an excessive focus on form over function. On the one hand, the learning curve of visualizations such as storylines should be taken into consideration, as certain visual metaphors may not be obvious to some (or even the majority) of the readers. On the other hand, the design of a visualization should primarily work towards conveying a story, and all visual elements must have a purpose.

Negative stances towards visualizations should also be taken into consideration. As one journalist said, other people who share their news reading habits may not benefit from visualizations. However, we must also acknowledge that this distaste might not stem solely from the way some people consume the news, but also from the quality of the visualizations with which they came across. As discussed above, visualizations that do not serve the story only add noise, and a series of poorly-executed visual stories can shine a negative light on visualizations as a whole.

While a narrative-driven visualization could work as a tool for both retrospection and day-to-day analysis, the arguments provided skew slightly in favour of the former. We need some distance

from the facts to properly evaluate their relevance to the over-arching story, and these types of visualizations need a complex story to truly showcase their potential over other representations. As time passes, the complexity of a story tends to increase, as new facts are uncovered and past events are revisited.

On the matter of abstract, high-level visualizations and concrete, narrative-driven ones, the data does not allow us to conclude whether any of the two is more likely to be preferred by the average news reader. Journalists provided arguments for both approaches. Some stated that a more focused, narrative-driven visualization of a particular story would interest the average reader more, while a higher-level approach with an hypothetical focus on news topics would be more attractive to a smaller subset of readers and to journalists. Others argue that a more abstract view could appeal to the average reader who may want to get a gist of the most popular news stories. In the matter of storyline visualizations, our data shows that they work best with stories which are factually dense, and where the chronological order of events is relevant. This chronological order is often lost when translating a news story to text, as journalists tend to prioritize relevance over timing, and thus present the facts from most to least important.

One topic that was not discussed previously in this chapter is that of information overload. Despite not having enough content to warrant a separate section, we want to highlight a concept brought up by one of the journalists, that of a 'spiral of sameness' that characterizes the abundance of repetitive articles in the news sphere. This concept, attributed to Boczkowski [7] by the journalist, describes one of the main contributors to the massive amount of information to which we are exposed to on a daily basis. In the case of Portuguese media, the news articles released by Lusa[1], the country's biggest news provider, are replicated by tens of news outlets, all vying for the readers' attention. As another journalist pointed out that, even though news rooms suffer from this excess of content, they are discouraged to do anything about it, as other outlets would take their place if they were to slow down the rhythm of production, for example.

The positive responses we gathered in relation to the use of visualizations in the news and, in particular, narrative-driven ones, validated our choice to explore them further during the implementation phase of the project. Despite the useful insight we collected, these interviews have some limitations. The sample pool is fairly small, and further research with more participants could be conducted to complement the aforementioned findings. Furthermore, journalists are at the forefront of the news production cycle, with direct access to the opinions and characteristics of news readers. However, reaching out directly to the news readers who do not have a professional connection to journalism would enrich our understanding of the problem space. Finally, one of the biggest challenges when designing the interviews was deciding how to clearly define the concept of visualization, especially considering the extent of the field. From simple bar charts, to complex visual stories, discussing visualizations purely in abstract terms would mean the participants could have very different interpretations in mind when expressing their opinions on this subject. The examples we provided during the interviews potentially primed the interviewees to skew towards more positive or negative opinions of visualizations as a whole. Notwithstanding, we believe this

[1] https://www.lusa.pt/

was a necessary compromise, as establishing a concrete baseline against which to compare the insights was more important to us than to allow complete freedom in the interpretation of what a visualization should be.

4.8 Summary

Before and during the early stages of implementation, we conducted five semi-structured interviews with journalists with the intent of exploring three main topics: the consumption of news, the journalistic narrative, and the relation between visualizations and news stories. From the artefacts gathered during these sessions, we were able conclude that journalists were generally optimistic about the potential that visualizations have to aid the reader in grasping the content of a news stories. These visualizations, if well-designed, can help overcome some limitations inherent to the traditional text format. For example, perhaps a timeline is more appropriate than text when the chronological order of events plays a big role in the comprehension of the narrative.

Chapter 5

Visualization Tool Implementation

As a starting point to the implementation of a visualization tool that highlights narrative elements in a news story, we began by exploring existing visualization tools, namely D3 Narrative Layout and Comic Book Narrative Charts. Both implementations make use of D3[1], a visualization engine built in JavaScript that allows for the creation of visualizations in browser-based environments. From there, we analysed the available features, what could be improved, and what could be added in order for the tool to better fit our needs. Having this analysis as a rough guide, we developed a browser-based interface based on elements from both existing tools that, given a JSON document with the specifications of a narrative, outputs a storyline visualization which allows for a series of interactions. The subsequent chapters will describe in greater detail the third-party implementations mentioned above, their characteristics, how they were adapted, the program flow, the input and output of the visualization tool, and the features of our implementation.

5.1 Existing tools

5.1.1 D3 Narrative Layout

Narrative Charts is an open-source (licensed under the MIT License) narrative charts layout engine for D3 [19] which aims to mimic the storyline visualizations of Munroe [46]. Given a JSON file of characters and events (an example of which can be seen in Appendix B), the engine outputs the position of each event in a the storyline. The styling and presentation of the output is left to the discretion of the user. If the user wishes to optimize the layout of the events in any way, they must modify the engine itself. The source-code is thoroughly documented, which greatly facilitates its adaptation.

Figure 5.1 depicts an example visualization based on the layout built by the engine and styled by the author and Figure 5.2 showcases a more complex use case, also designed by the author as

[1] https://d3js.org/

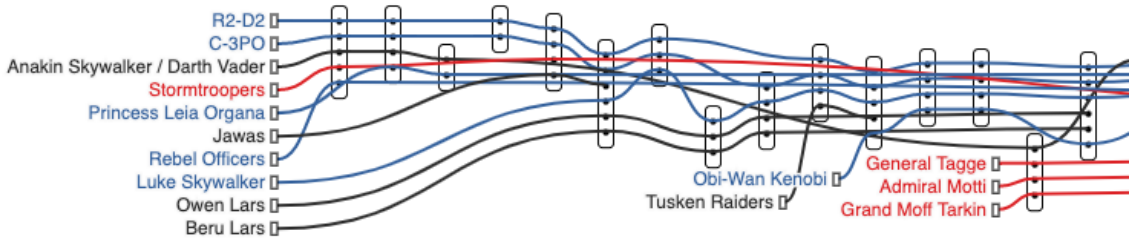


Figure 5.1: An example of the layouts produced by the Narrative Charts D3 engine [19].

part of an effort to chart every event in the first six films in the Star Wars saga [22]. While the example visualization includes only a static storyline, the more complex visualization features a number of complementary elements. For example, the reader can highlight certain events and read their description, filter the highlighted events by character, and sequentially focus on each event in the storyline (through the use of the two arrow buttons positioned under the visualization).

The contrast between these two examples is a testament to the versatility of the engine. However, the layout of the nodes (i.e. the output of the engine) can often lead to confusing storylines, with a considerable number of line crossings (as pictured in Figure 5.1). Nevertheless, the possibility of applying custom styles and interactions can help alleviate the information density of the visualization without directly altering the engine itself.

5.1.2 Comic Book Narrative Chart

Figure 5.3 depicts the Comic Book Narrative Charts, a visualization tool developed by Iskander et al. [30]. At the time of writing, no license information was present in the project's website; as such, we contacted the authors to ensure that we could use and adapt the source code (a request to which they agreed). The goal of this project is similar to that of the previously discussed implementation (i.e. to generate a narrative chart based on the work of Munroe [46]), and so is the input format (with slight differences in the information provided for each event). The placement of the nodes ensures a cleaner presentation of the storyline. Furthermore, the visualizations allow the user to drag the nodes and to hover over them to activate a tooltip. However, while the output of the tool may outperform the out-of-the-box results of the previous implementation, it is not as versatile, and modifying the source-code is not trivial (since documentation is sparse).

5.2 Choosing a starting point

As a starting point to our prototype, we chose to work with D3 Narrative Layout [19]. One of the major reasons behind this choice was the thorough documentation that accompanied the project. Nevertheless, it does have a number of drawbacks, the most salient of which: the example visualization provided by Elvery [19] was not as feature-rich nor as visually-appealing as those of Iskander et al. [30], and the implementation of the Louvain's method for community detection [5]

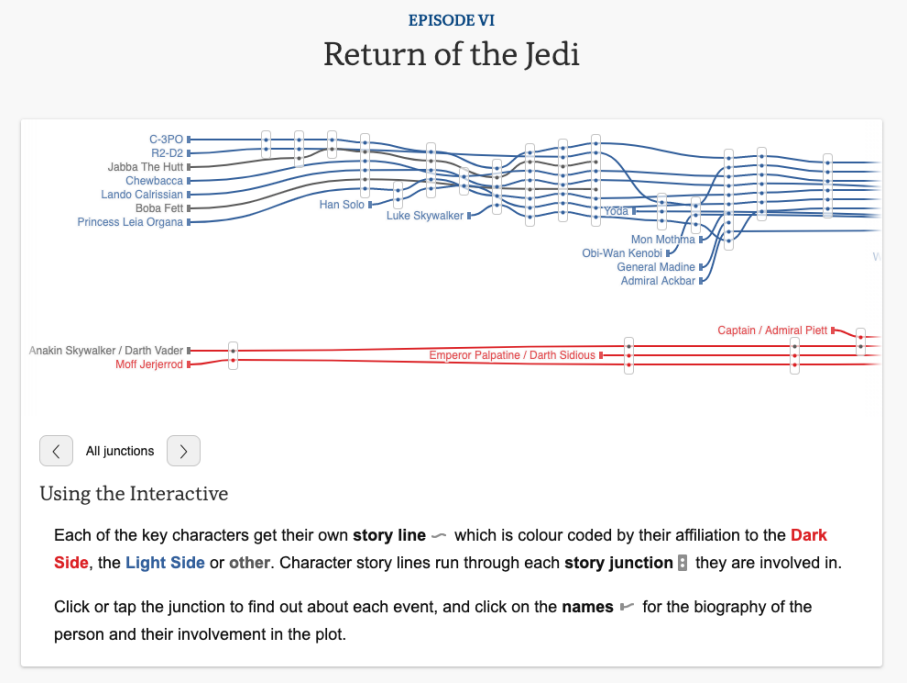


Figure 5.2: A complex interactive visualization built on the layouts produced by the Narrative Charts D3 engine [19].

used to group the characters and subsequently order them in the visualization did not appear to yield any relevant results in the examples tested (i.e. the characters were all placed in the same group, thus a default ordering based on the characters position in the corresponding data structure was used). Despite this, and given that this default ordering provided an acceptable event distribution upon which to work, we opted for the D3 Narrative Layout engine.

5.3 Overview

Based on the previously mentioned engine, we built a browser-based prototype which, given as input a JSON document detailing entities and events belonging to a story, presents said entities in a visualization that allows the user to explore said story through a narrative-driven lens, focusing on the relationships between entities, events, locations, and dates. The source-code of the prototype can be found on GitHub [12]; a live version is also available[2]. Appendix C.1 describes the installation instructions. The content is licensed under the MIT license. The interface, depicted in Figure 5.4, is composed of three main views: the visualization, the event information module, and the sidebar.

The visualization (**A** in Figure 5.4) is the core element of the interface. We worked with the layouts provided by D3 Narrative Charts [19] (i.e. the positions of the introduction nodes, event nodes, entity lines, and appearances within an event) to produce an interactive storyline

[2] https://marianafcosta.github.io/news-story-viz/

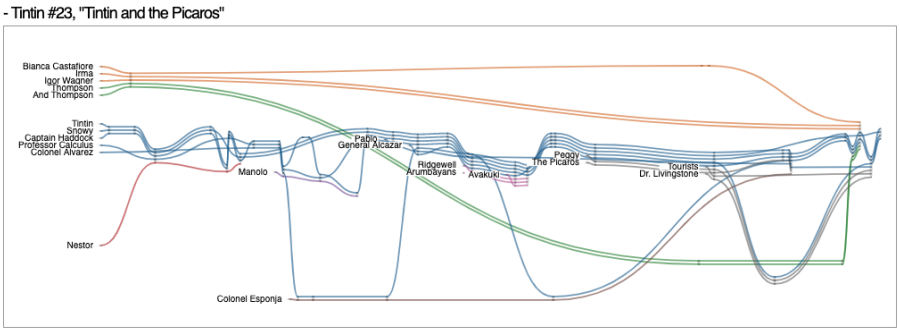


Figure 5.3: A Comic Book Narrative Charts visualization of the story in the 23rd issue of Tintin, "Tintin and the Picaros" [30].

visualization, complemented by additional features such as: a timeline, the ability to drag events and introduction nodes, the ability to focus on a event to obtain more information on it, and the ability to hover a event to activate a tooltip (containing the date and location of the event, if available).

The event information module (**B** in Figure 5.4), placed at the bottom of the interface, displays a detailed description of the event in focus (if any), along with some more contextual information such as the event's title, date, and location. The entities are highlighted using the color of the entity's line in the visualization. While the latter provides a bird's eye view of the events that make up that storyline, the event description module allows for a more detailed view of each individual event.

Finally, a collapsible sidebar (**C** in Figure 5.4) adds a number of additional interactions with the visualization, including: entity, event, and time frame selection, updates to the storyline based on those selections, and the option to filter elements according to a number of criteria (e.g. hiding entities which are not selected). In addition to these three main visual components, the user can also download the visualization as a PDF (reflecting all filters and updates applied to it) and visit an about page which contextualizes the project and provides information such as licenses and links to the source-code of the project and of the D3 Narrative Charts implementation. Figures 5.5 and C.1 depict these two features, respectively.

5.4 Interface

5.4.1 Visualization

Figure 5.5 provides a description of the visualization. Central to this view is the notion of events, characterized by a description, a title, participants (i.e. appearances), a date (optional), and a location (optional). These events are translated visually in the form of ellipses, indicated as element **D**. Inside those ellipses are circles which depict the appearance of entities in that respective event (element **C**). The color of said circles correspond to the entity to which that appearance belongs.

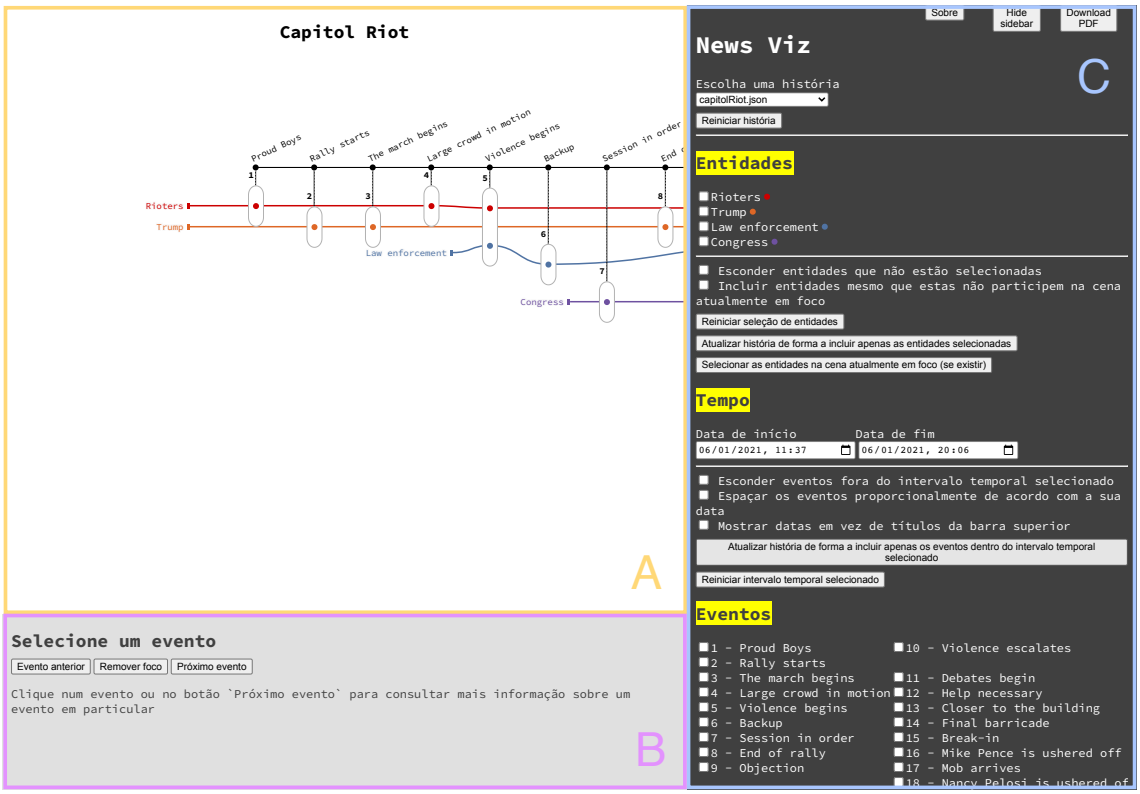


Figure 5.4: An overview of the interface of our application. Element **A** depicts the visualization (the core element of the interface), element **B** depicts the event information module (which provides more information on a particular event), and element **C** depicts the sidebar (the hub for the majority of the interactions available).

The lines flowing across the visualization represent the entities participating in the storyline (element **E**). Each line is attributed a random color if no user-specified color exists. The source and destination of the lines can either be an event or an introduction node (text legends that associate a line to a character's name, depicted as element **A**). Above the events and entity lines is a timeline that connects the events to a specific point in time or the title of the event in question (element **B**).

5.4.1.1 Interactions

While the sidebar unlocks the majority of interactions with the visualization (e.g. filtering events by date), some do not require access to that module. Table 5.1 list the interactions which are directly available through the visualization itself. The user can access a contextual tooltip (depicted in Figure 5.6) by hovering over a event, which allows them to quickly access the date and location of the event. The tooltip is only visible if the event in question is in focus, or if no events are in focus. It is also possible to drag the events in the vertical direction, the introduction nodes in both vertical and horizontal directions, and the titles or dates in the timeline in the horizontal direction. Finally, clicking on an event will focus on it, updating the event information module with the details of that particular event.

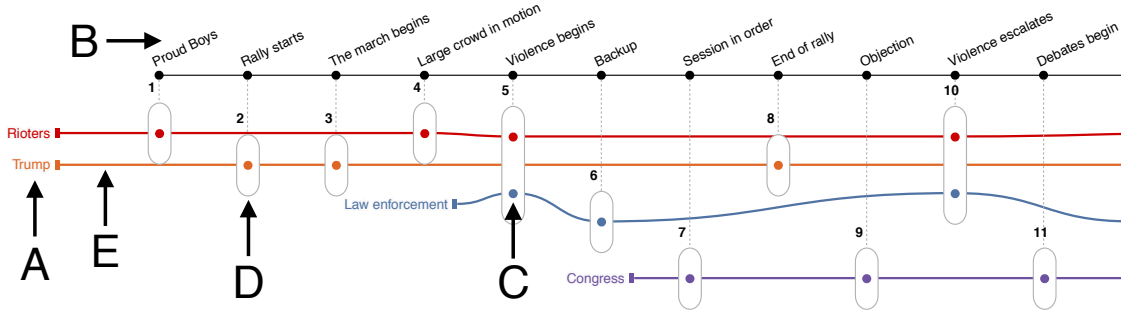


Figure 5.5: Crop of a PDF provided by the application, depicting the visualization.

5.4.2 Event information module

The event information module (present in Figure 5.7) is tightly coupled to visualization; it displays the information relative to the event in focus (or placeholder text in case no event is in focus). The user can also use the "Next event" and "Previous event" buttons to sequentially focus on the events. To reset focus, the user should press the corresponding button. Table 5.2 summarizes these functionalities.

As mentioned above, names in the text that correspond to entities who participate in the event are highlighted using that character's color. One of the obstacles faced during the implementation was the correct detection of names that represent a particular entity but do not share their exact name. For example, in our Capitol Riot example in Figure 5.4, we have a collective entity named "Congress". A number of entities mentioned in the description of the events should be highlighted as being part of the "Congress" entity (e.g. Nancy Pelosi, House, or Senate). To ensure this, each entity can be attributed a set of names associated to said entity. For each event in focus, the system will search the text for any mentions of the entities or of the terms associated with them, highlighting them accordingly (as is the case with the "rioters" and "police" in Figure 5.7).

5.4.3 Sidebar

Figure 5.8 depicts the sidebar, which facilitates the majority of the interactions available to the user. It is divided in four distinct sections: storyline selection, time, entities, and events. In addition to these sections, three fixed buttons grant the option to open the about page, to hide the

	Functionality	Taxonomy
FV1	Drag events	Reconfigure
FV2	Drag event titles	Reconfigure
FV3	Drag entity names	Reconfigure
FV4	Hover for tooltip	Elaborate
FV5	Focus on event	Elaborate

Table 5.1: Functionalities associated with the visualization itself.

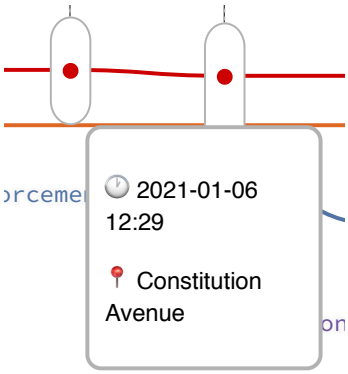


Figure 5.6: Whenever the user hovers over a node, a tooltip containing the event's date and title is displayed.

sidebar, and to download the visualization as a PDF. The storyline selection section allows the user to pick a story, as well as to reset the storyline. All other functionalities are summarized in Table 5.3. In the time section, the user can select a particular time frame, hide events which fall outside of the selected timeframe, space out the events so that their position in the visualization is proportional to the date in which that event occurred, update the storyline so that any event which falls outside of the selected time range is removed, and reset the timeframe selected. Moving on to the entity section, the user can select entities (a selected entity is encoded as thicker line with a bold introduction node), hide entities which are not selected, include the selected entities even if they do not participate in the event currently in focus, select the entities in the event currently in focus, update the storyline so that any event which does not contain any of the selected entities is removed, and reset the entity selection. Finally, in the event section, the user can select events (a selected event will have a pastel yellow background), hide events which do not include the selected entities, hide events which are not selected, update the storyline to so that any unselected event is removed, and reset the event selection.

5.5 Architecture

The system follows a Model-View-Controller model, described in Figure 5.9. The models contain all of the information pertaining to the narrative (e.g. events' descriptions, titles, and absolute positions in the visualization). The views are responsible for handling the visual interface, and

	Functionality	Taxonomy
FI1	Scroll through events	Select/Elaborate
FI2	Remove event focus	Abstract

Table 5.2: Functionalities associated with event information module.

	Functionality	**Taxonomy**
Time		
FT1	Hide events outside of the selected time frame	Filter
FT2	Space out the events according to their timestamp	Reconfigure
FT3	Show dates instead of titles	Explore
FT4	Update storyline to only include events inside the selected time frame	Filter/Reconfigure
FT5	Reset the time frame	Filter
Entities		
FEn1	Select entities	Select
FEn2	Hide entities which are not selected	Filter
FEn3	Show selected entities even if they are not on the event currently in focus	Filter
FEn4	Select the entities from the event currently in focus	Select
FEn5	Update the storyline to include only the selected entities	Filter/Reconfigure
FEn6	Reset entity selection	Filter
Events		
FEv1	Select events	Select
FEv2	Hide events which do not include the selected entities	Filter
FEv3	Hide events which are not selected	Filter
FEv4	Update the storyline to include only the selected events	Filter/Reconfigure
Other		
FO1	Select story	Explore

Table 5.3: Functionalities associated with the sidebar.

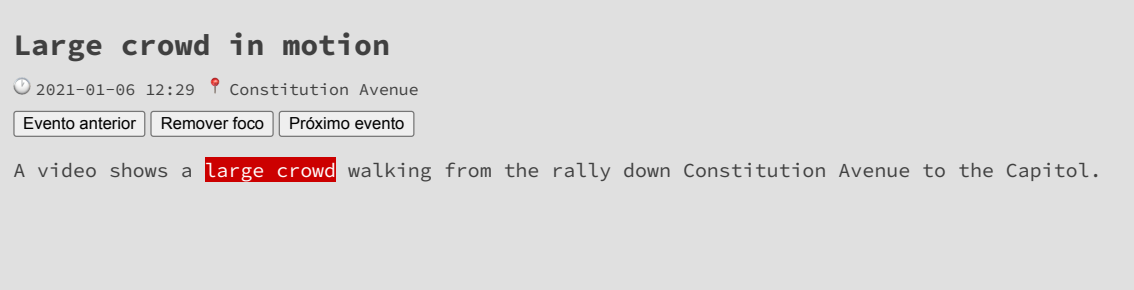


Figure 5.7: The event information module allows the user to scroll through the events, sequentially focusing on each one. Furthermore, the entities mentioned in the event description which are part of one of the entities participating in the event are highlighted with the color associated with said character.

are in charge of tasks such as connecting input elements to the event listeners and updating the interface according to the changes made to the model. The controllers establish a point of contact between the models and the views. They subscribe to the events triggered by both the models and the controllers, and are responsible for updating the models or the views according to those events. This structure ensures that there are no direct dependencies between the models and the views. Figure 5.10 demonstrates how the elements interact with one another, as well as where the event system comes into play.

5.6 Data

The storyline originates in a JSON file similar in format to the one that was originally used by the D3 Narrative Layout [19], with a few modifications. Both the original and the updated format are available in Appendixes B.2 and B.3, respectively. The JSON files used for testing were produced manually, as studying the automatic generation of these files from already existing documents falls beyond the scope of this document.

On initialization, the system loads a pre-defined narrative. The JSON file is parsed to conform to the format expected by the application. The narrative is then set up and the interface is updated to reflect these changes. From this point on, the user can interact with the narrative elements in the visualization (e.g. select events, drag events around, or choose to hide events that fall outside of the selected timeframe). Despite all changes that may occur to the narrative that the user is currently seeing (e.g. when they choose to update the narrative to only include the selected entities), a copy of the original file is kept as a way to ensure that resetting the storyline to its original format is possible. When the user picks another storyline, a new JSON file is loaded and the process repeats itself.

5.7 Use case

To demonstrate the capabilities of the system, we will explore a narrative originally published in the New York Times that covers the Capitol riots of 2021 [41]. In the online version of the news article, two visualizations are used: a storyline that is similar to the one we have implemented, and a more traditional timeline. In the former, entities are encoded as colored lines that flow from left to right, with significant events depicted by dark vertical lines topped by a short description of the events. We will briefly compare our storyline with the one in the aforementioned article in a subsequent section. As for the timeline, the events are ordered vertically, with small colored dots indicating the main entity behind the events. Figures 5.11 and 5.12 illustrate the storyline and timeline, respectively.

For this exercise, we will describe a set of interactions that a potential user might go through when exploring a narrative with our visualization tool, starting with obtaining an overview of the events. Appendix C.3 contains some of the tasks described below. A reader who is not familiar with the story might start by simply scrolling through the storyline, reading the titles in the top bar and getting a feel for how often entities interacted with each other, which entities appear the most often, and so on. Following this high-level exploration, the reader can sequentially scroll through the events, gathering more insight into the details of each one and expanding their knowledge of the story. After this first pass through the events, they might want to know, in greater detail, how certain entities interacted. For example, they might want to select the “Rioters” and “Law enforcement” entities and hide all others in order to understand if the encounters between the two were violent, or whether the law enforcement was quick to react to the actions of the rioters (see Figure C.2). From here, they might space out the events according to their timestamp and realize that a considerable number of events took place between two and three p.m. (see Figure C.3). Additionally, they may also conclude that the National Guard took two hours to arrive on scene, after the Capitol was declared secure. Finally, the reader may want to download a PDF with the visualization and the two aforementioned entities select to share it amongst their peers and discuss their findings.

While the New York Times article presents a cleaner visualization, some of the tasks described above are not easily accomplished. The storyline visualization in the article lets the reader understand the most important events, as well as periods during which entities were absent (something our visualization does not allow); however, you need to scroll through each event in the timeline below to get a gist of the whole story. More, interactions between entities are not easily conveyed, as each event is attributed to one entity only (even if others are mentioned). Even though the timeline preserves relative order, and the spacing between events somewhat reflects the time passed between them, you cannot immediately understand which periods were more or less eventful. The addition of multimedia elements such as photos of the events helps contextualize the story, and is something that would benefit our visualization (as some journalists pointed out during the user tests described in Chapter 6). Adopting a vertical alignment such as the one in the article’s timeline is an interesting path for future work (something that also also brought up during the user tests pre-

viously mentioned). Overall, the New York Times's visualization is cleaner and simpler, whereas ours offers more options through an interface with rough edges that is more overwhelming. Nevertheless, we should keep in mind that we are presenting a prototype with a number of accessibility and usability concerns that would be addressed in a later stage of the product development cycle.

5.8 Comparison with other tools

Of the 17 works analysed in Chapter 2, only Storyteller [59] utilizes a storyline visualization. In that project, the storyline is part of a tripartite system aimed at exploring rich event data from text. Storyteller uses the Simple Event Model [58] (SEM) as an intermediate representation for their input data. We cannot compare the two applications in depth, since there is common storyline between the two. Questions related to the disposition of visual elements and to the clarity of the visualization would require similar input data and, as such, we will not discuss them. Nevertheless, it is possible to conclude that our visualization offers more opportunities for interaction. Furthermore, it is harder to understand exactly which entities participated in an event in Storyteller, whereas in our implementation this information is readily conveyed through the concept of occurrences. However, their system accepts data following a standard format, whereas we have no way of automatically producing input documents from text sources. This facilitates reuse, integration with other tools, and founded comparisons between tools (for the reasons we have listed above). In future iterations, it would be interesting to explore ways to accept input data in standardized formats such as SEM.

5.9 Summary

We based our prototype on an open-source implementation of a storyline visualization framework in JavaScript. In addition to the adaptions made to the visualization itself, we built an information hub around the storyline to allow for multiple levels of exploration. The interface is thus composed of three main elements: the visualization, the event information module, and the sidebar. The visualization depicts the sequence of events that make up a news story, as well as the interactions between the involved parties in said events. The event information module presents additional information on a particular event. Finally, the sidebar is the "control center" of the visualization, unlocking a number of interactions. The architecture of the application follows a Model-View-Controller model, with an auxiliary event system to facilitate communication between the different elements. The input documents follow the format specified in Appendix B.3, and the files used for test purposes were produced manually.

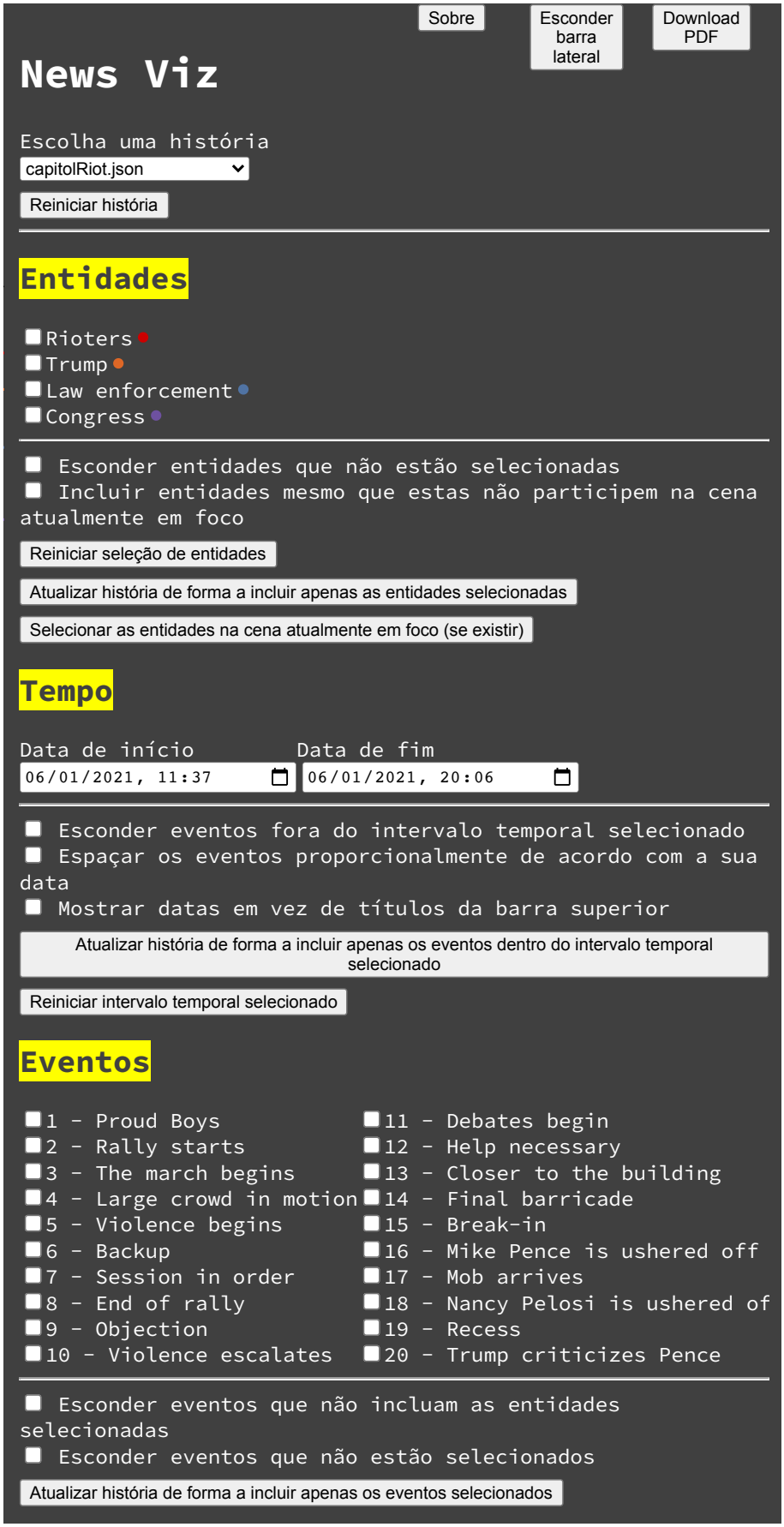


Figure 5.8: The sidebar is akin to a control center for the visualization. It allows for most of the interactions available to the user.

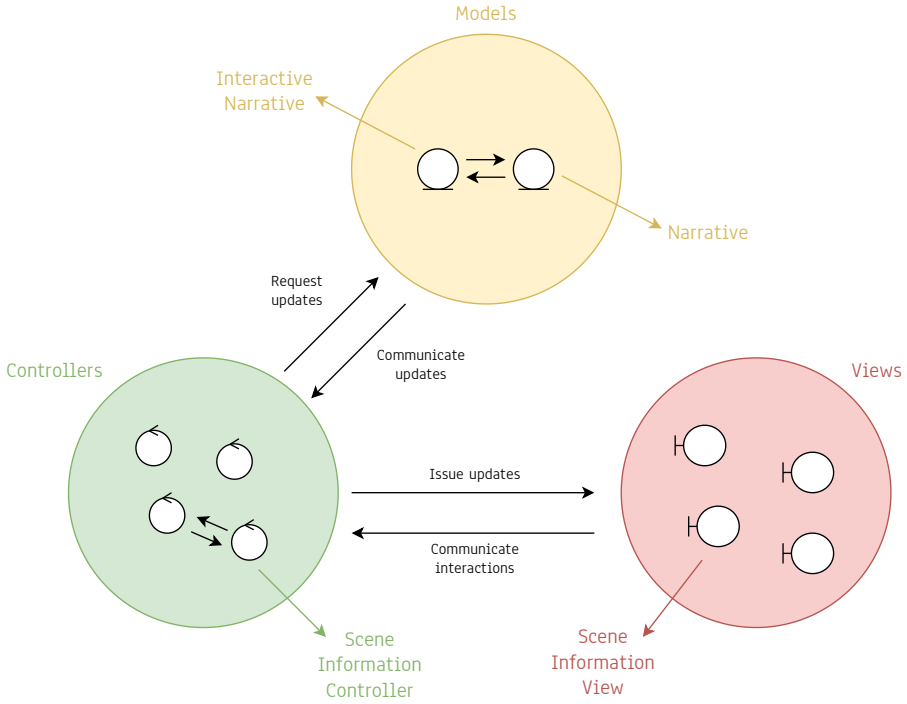


Figure 5.9: A diagram describing the adopted Model-View-Controller approach, along with examples of each element in the application.

Figure 5.10: An example of an interaction with the system, along with the communication between the applications components.

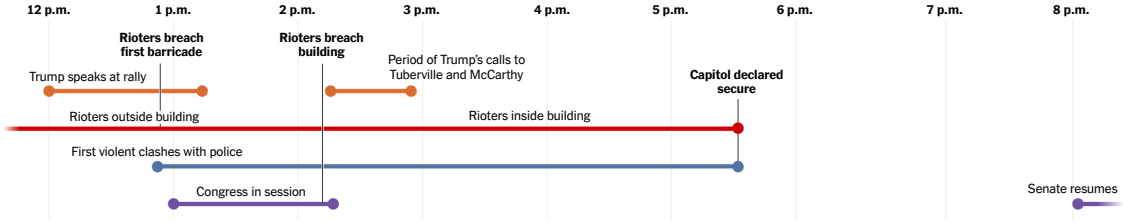


Figure 5.11: Storyline from the New York Times's article on the Capitol Riots [41].

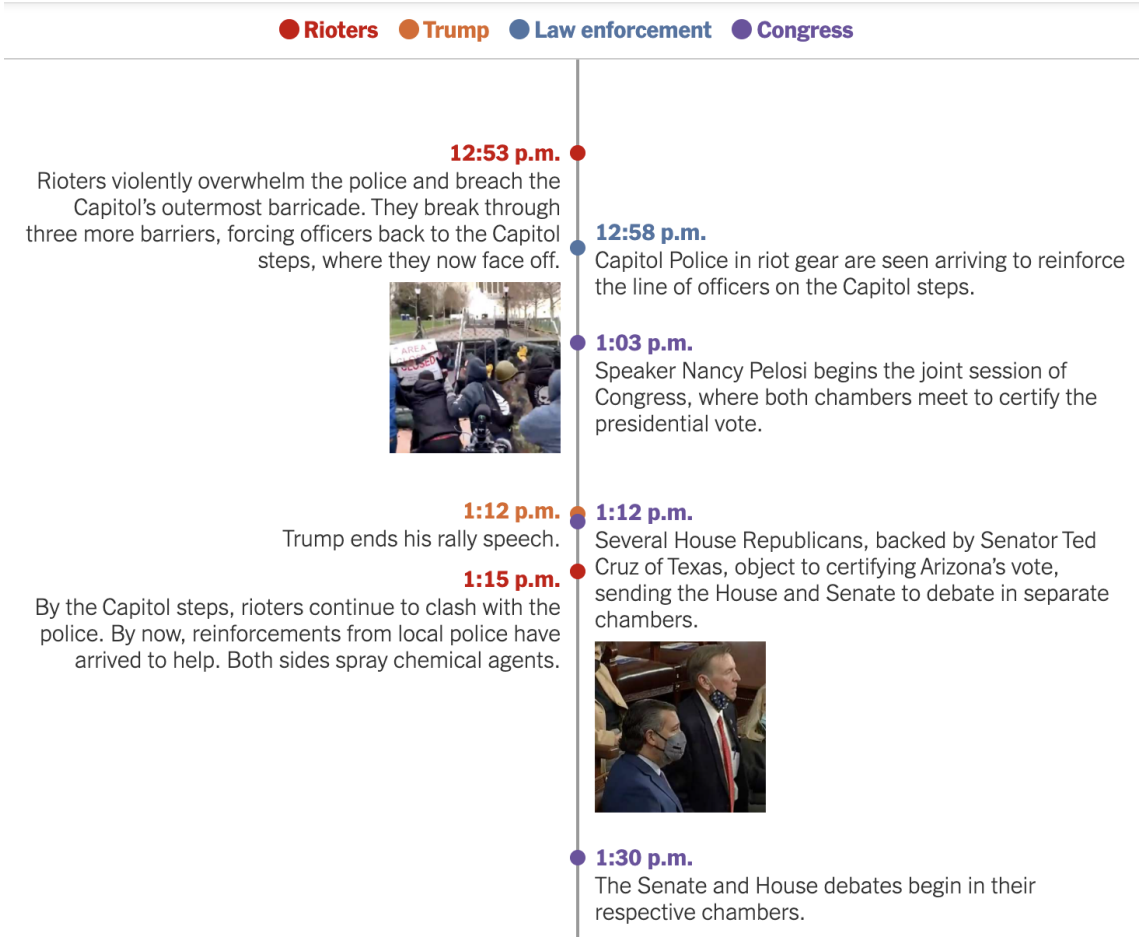


Figure 5.12: Timeline from the New York Times's article on the Capitol Riots [41].

Chapter 6

Solution Evaluation

With both the user tests and survey, we wish to assess the efficacy of the developed prototype in exploring news stories from a narrative perspective. Our focus was on collecting qualitative data which would in turn allow us to identify problems and opportunities [45]; however, some quantitative data was also gathered, in particular during the survey. We will go over the results of the user test and survey, prefacing each section with a description of the methodology adopted, along with brief discussions of some of the design choices behind it. The original quotes from the participants can be found in Appendix E.1.

6.1 Tests

6.1.1 Methodology

While the format of the evaluation methods varied between the works reviewed in Chapter 2, user tests were a common approach for evaluation. By user tests, we are referring to methods involving one or more participants in a controlled setting, in which the participants are asked to complete a set of tasks or to explore the system freely. Five people participated in our individual test sessions. Three of the participants also took part in the interviews conducted prior to the implementation of our prototype. All had strong ties to Journalism, and all are currently working in the field. The participant pool was comprised of five men. First, the participants were introduced to the prototype; the interviewer walked through the main components of the interface and some of the main concepts pertaining to the visual metaphor used. Then, the participants were asked to complete a number of tasks and, between each task, were asked to rate the usefulness of the underlying feature, on a scale from one (i.e. not useful at all) to five (i.e. extremely useful); these tasks are listed in Table 6.1. The number of questions asked between tasks was kept to a minimum in order to reduce the interference with the flow of test [40]. We also collected difficulty ratings for the tasks, but ultimately decided against using them, since they were relatively simple and not conducive of an interesting analysis for our purposes. This exercise was followed by a period of open exploration, in which the participants were asked to interact with the system in whichever way they saw fit, voicing any comments or suggestions along the way. Finally, the test session

	Functionality	Interaction taxonomy
T1	Focus on event	Select, explore, abstract/elaborate
T2	Move the events	Reconfigure
T3	Filter by date	Filter
T4	Select entities	Select
T5	Filter events	Filter
T6	Filter entities	Filter
T7	Update storyline	Reconfigure
T8	Reset storyline	Reconfigure
T9	Download PDF	-

Table 6.1: Tasks in the first part of the test.

concluded with a set of post-test questions related to the overall perception of the prototype (listed in Table 6.2).

Since the implementation is still a prototype, we were mostly focused on understanding which features are desirable and which are dispensable; the usability of the interface was a secondary concern when designing the tests. Nevertheless, we found it difficult to complete detach the concepts of usefulness, desirability, and usability. If we were to ask the participants conceptual questions regarding the types of features they would expect to see in a visualization of this sort, then this distinction could be easily made (since we cannot measure the usability of an idea). However, when participants are confronted with a real-life implementation, their perception of usefulness is bound to be affected by how usable the features actually are. Initially, we planned to have a demonstration of features, as opposed to asking the participants to complete a number of tasks. We eventually decided against this, redesigning the tests in a way that allows us to quantitatively and qualitatively reason about the aforementioned metrics, while still focusing on the usefulness of what is provided to the user, and not on the details of the interface.

The tasks were designed to encapsulate one single feature of the interface. Since the time of each session was limited, we chose a set of features which we believe to be some of the most common when analysing a storyline. All categories of the interaction taxonomy of Yi et al. [63] that can be found in the implementation are represented in the tasks. The answers were not timed nor was the success rate calculated, as it is still an early prototype and we believe that little could be gained from this at this stage. Counteracting the rigidity of the previous section, the open exploration period allowed us to collect any feedback or suggestion that the participants might have had in relation to specific features or the interface as a whole, thus gathering some qualitative insight.

The second question that comes about is related to the context in which usefulness should be perceived by the participants. Since this concept is almost completely dependent on the circumstances under which the tool is expected to be used, we preface each session by stating that what the participant will evaluate is a prototype targeting the average news consumer, to be used as a way to complement a news article and to facilitate its consumption. One of the later questions

	Question
Q1	How hard was it to understand the base elements of the visualization?
Q2	How much did you enjoy using the tool?
Q3	How much would you be interested in having access to this tool on a daily basis?
Q4	How useful is this tool to the average news reader?
Q5	How useful is this tool to journalists?

Table 6.2: Questions asked in the second part of the test.

urged the participants to explore other possible contexts.

6.1.2 Quantitative data

Figures 6.1 and 6.2 ratings provided by the participants for the first and last parts of the test, respectively (the question listings can be seen in Tables 6.1 and Tables 6.2). For the first set of questions, resetting the storyline was the highest rated functionality in terms of usefulness, followed by filtering and selecting entities. This sentiment is replicated in the comments made during the open exploration phase, which will be discussed in the next sections. On the other hand, moving an event gathered the least amount of "five" ratings and the most of "two" ratings, in part because the journalists, as they explained after the tasks, could not find any relevant use for it in the storyline used. Five out of the nine functionalities exclusively received ratings of four or above: focusing on an event, selecting entities, filtering entities, updating the storyline, and resetting the storyline. Downloading a PDF, filtering events, and filtering by date share the same distribution of votes: three "five" ratings, one "four" rating, and one "three" rating.

As for the second set of questions, the participants found the prototype to be more useful to journalists than to the average readers, with Q4 registering the same number of "five" ratings, less "four" ratings, and more "three" ratings than Q5. This may be linked to the complexity of the interface, as average news reader might be overwhelmed by it (as was suggested in the comments made by the participants). Questions Q3 and Q2, related to the interest in and enjoyment while using the tool respectively, had the same distribution of votes, with three "four" ratings and two "five" ratings. Finally, four participants gave a rating of "two" or less to the difficulty of understanding the visualization. Once more, the overall sentiment of the comments made by the journalists support these ratings. As we will discuss in the subsequent sections, the participants were generally interested in the tool, and found the visual metaphor easy to grasp; however, all voiced concerns about the number of options available to the user and to the attrition felt while navigating the sidebar.

6.1.3 Positive feedback

The journalists had a generally positive stance on the application as a whole, as evidenced by the results of the quantitative questions presented earlier. Table 6.3 provides an overview of the topics related to this theme. Some of the comments pertain to the ease with which the visual encoding

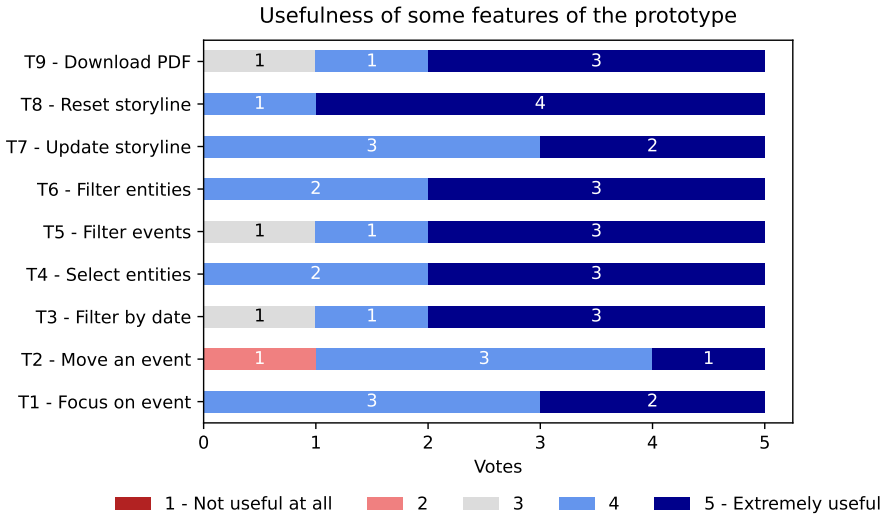


Figure 6.1: Results of the first part of the test.

was understood. The prototype provides a different perspective on a news story, one which allows the user to quickly grasp key information such as the participants, the dates, and the location of events.

6.1.4 Negative feedback

The most frequent feedback we received is related to the amount of options offered in the sidebar. While a power user may be interested in having considerable control over the tool, to the average reader the number of options available might be overwhelming.

> *I think this might have too many options, especially when you are in a newspaper with a giant audience which will probably not dedicate much time to this.*

One journalist likened the prototype to a more 'technical' tool, as opposed to an application that might be used in a journalistic context. Another journalist warned of the possibility of having readers focus too much on the options, as opposed to the story.

Codes	Example
Appealing visualization	*I think the visualization is very simple, in a good way; it is easy to absorb.*
Situates readers	*It allows the reader to quickly visualize who participated, when, and where.*
Different way of seeing things	*I think it is very interesting, it is a different way to look at things.*

Table 6.3: Codes related to the positive feedback gathered from the journalists.

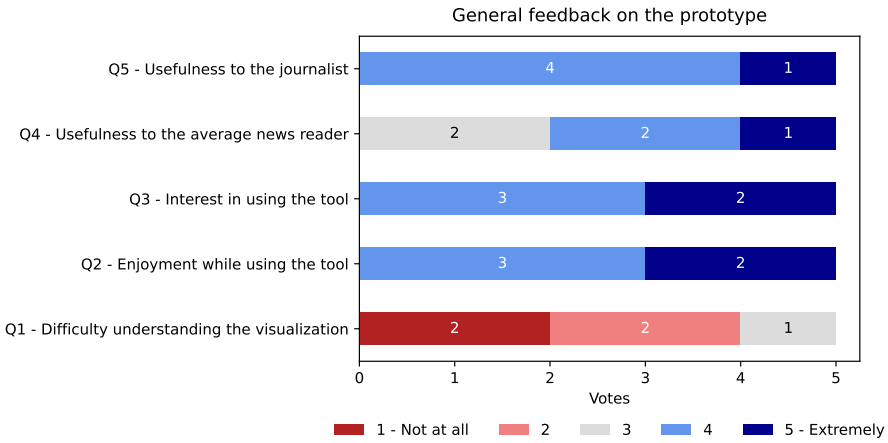


Figure 6.2: Results of the second part of the test.

> *It can turn into a 'technical' application instead a journalistic one, which generally has very few options.*
>
> *If I have too many options, then I can get lost in them instead of paying attention to the story.*

Not only is there too much choice as to what the user can do, some of the visibility options are redundant, as one journalist pointed out when asked about the possibility of hiding events or removing them from the storyline entirely. Moreover, journalists pointed out that having to select an option to hide the characters which are not selected is cumbersome and something that could be automated.

> *When I select something, the change should happen immediately without having to press another button [to hide the entities which are not selected].*

One other common theme brought up by the interviewees is the lack of clarity of certain functionalities. This is due to a number of factors, namely the excessive complexity of certain

Codes	Example
Too many options	*I am not sure if it is necessary to have so much control over the application.*
Unclear interactive elements	*I would not understand what that button [to space out the events according to the timestamp] does if you had not told me.*
Redundant visibility levels	*I would say they are somewhat redundant, the first option [to hide the entities] does the same thing [as removing the entities entirely]. I do not see a point in deleting the entities.*

Table 6.4: Codes related to the negative feedback gathered from the journalists.

Codes	Example
Reset the visualization	*You need to have the option to reset, to go back to the original view.*
Select/highlight entities	*I though the option to highlight entities was the most useful, it helps uncover certain narrative threads.*
Space out the events	*[Spacing out the events according to their timestamp] is important, it lets us know exactly which periods were the most intense.*
Event scroll	*It is probably the most natural way to explore the narrative.*

Table 6.5: Codes related to the features journalists explicitly claimed were the most important.

buttons (which could be replaced easily recognizable icons such as the circular reset arrow), the wording of certain options, and the lack of a visible scroll bar in both the visualization and the sidebar. Table 6.4 synthesizes the findings for this section.

> *The buttons do not stand out, they were a bit tough to read; besides, they have a lot of text.*

6.1.5 Important features

Table 6.5 synthesizes the findings for this subsection. The option to space events out according to their timestamps (as opposed to being evenly spaced out) was often mentioned as providing an alternative perspective (and, subsequently, other interpretations) of the events. Nevertheless, this option comes at the cost of readability, as events that have a small time difference between them overlap and their titles become hard to read.

> *Despite the mess that is created around [the titles in the timeline], it helps understand certain dimensions [of the story].*

Resetting the visualization and highlighting characters were seen as one of the most important functionalities. However, the latter has room for improvement, as we will discuss in the following subsection. Finally, journalists agreed that the possibility to sequentially focus on each event is relevant, particularly when a reader is not familiarized with the story.

> *This is the most reasonable way to navigate, is it not? Step by step.*

6.1.6 Suggestions

For the suggestions, we have divided them into four distinct categories (as seen in Table 6.6): sidebar, visualization, event information module, and other. The participants overwhelmingly agreed that the options should be cut down, some proposing that entities be selected by default, removing the need for the user to act on the choice of characters and thus simplifying the interaction process.

While one journalist said the reset options in the various sections (i.e. entities, time, events, and the full reset) should be "standardized", another said that having a different wording for each

Codes	Example
Sidebar	
Concise wording	*It would help to have more concise sentences.*
Hover to explain options	*You should have an icon that to explains what that option does on hover.*
Replace buttons with icons	*A fast-forward icon could also be simpler.*
Multiple option screens	*There should be a distinction between the basic options and those geared towards the power users*
Hide characters automatically	*For example, when I select 'rioters' and 'Trump', it should highlight these entities by hiding the ones which I did not select.*
Standardize the resets	*Maybe you should repeat [the reset icon] in the various options [to reset].*
Different wording for the resets	*Using a different wording would probably help me keep track of what each option does.*
Event info on highlight	*If I deliberately move the cursor to select an event, why do I still need to go to the visualization to read up on that event?*
Visualization	
Show info on title click	*My first instinct was to click on the dot [to focus on the event].*
Toggle entity on click	*If I clicked [on the entity's name on the visualization], could it not hide the 'rioters'?*
Entities selected by default	*By default, you would have all entities selected with a check mark; then, the option to remove the check mark [to hide the entity] would be intuitive.*
Scroll with drag	*I was hoping that dragging the visualization would allow me to scroll through it.*
Vertical storyline	*I wonder if it would make sense to place this storyline vertically instead of horizontally.*
Add date to title	*I think it would be more useful to have both [the date and the title].*
Format the dates	*The date format could be more user-friendly.*
Event information module	
Highlight event info	*I wonder if it would help to highlight this sequential navigation of events.*
Side-by-side with vis.	*Maybe it would be nice to have the information on the event in the same reading plane as the event itself.*
Other	
Add a top bar	*I would add a top menu, without the 'weight' of a sidebar.*
Other download formats	*If we had a PNG instead of a PDF, it would be useful to share on social media.*
More multimedia content	*You could enrich [the event information] with a photo or video.*
Tutorial animation	*You could have a short animation at the beginning.*

Table 6.6: Codes related to the journalists' suggestions.

reset might help them remember what each one does. To accommodate for both average and power users, two journalists proposed the inclusion of two different option views, one with the more basic options, and another one which allows for a more in-depth customization of the visualization.

> *In a first screen, you would have the basic options; then, if it was really necessary, you could add another with advanced options.*

The journalists also commented on the interactions with the visualization itself, covering mostly two topics: the dissonance between the expected behaviour and the actual behaviour, and the desire for more multimedia content. With respect to the former, two journalists suggested that dragging the visualization should scroll through it. Furthermore, clicking on the outside area of the event in focus should remove that focus. On the topic of entities, clicking on an entity in the visualization should select it, something that is currently only possible by using the entity list in the sidebar. Moving on to events, the small black circles on the timeline above the storyline should also allow the user to focus on the event in question. Finally, one journalist suggested a small, explanatory animation on load, as there are a considerable number of interactions, and some slip by unnoticed.

As for multimedia content, two main suggestions arise: adding video and images to the descriptions of events, and allowing the user to download the visualization in more formats. One journalist says that, by adding videos and photos, we are left with two levels of information: one more "schematic" (the visualization), and another more in-depth (the description of the event). On the other hand, facilitating the conversion to a PNG format would make the visualization more appropriate for social media, whereas allowing the download of the input file for the visualization, for example, would allow the user to extract more information on their own.

> *If you could add a photo or video, then you would have a 'whole' story, a more 'schematic' layer complemented by a more in-depth one.*

> *I wonder if it would be useful to have this [download] in other formats. I would like to extract data from this.*

Between having the title of the event or the date in the timeline, journalists generally favoured the title. However, most also said that having the date alongside it would be beneficial. Moving on to the storyline, three of the journalists suggested that the visualization be rotated so as to flow vertically. This would facilitate navigation in smaller screens and encourage scrolling.

Finally, some general remarks pertaining to the general arrangement of the interface were made. One journalist said that it would be interesting to place the event information in the same reading plane as the visualization, side-by-side as opposed to stacked vertically. Another journalist also suggested that the sidebar be replaced by a slimmer top bar, which would favour the visualization and event information module.

Contexts
Organizing the activities of a business
Management of public information
Management of information related to financial markets
Exploration of historic narratives
Analysis of the performance of public entities in the safety and health departments

Table 6.7: Suggested use cases for the prototype.

6.1.7 Other use cases

In addition to exploring the capabilities of the prototype we developed in the journalistic field, we also wanted to make note of any other contexts in which the tool could be useful. Table 6.7 summarizes these findings. One journalist suggested a visualization such as this one might be useful when planning a business's activities for a particular time frame. The management of communications from government entities in the health and safety departments were also common suggestions, with two journalists providing the same example: analysing the actions of safety authorities during high-risk fires. Another journalist provided an example pertaining to the study of historic narratives. The possibility of exploring well-established events through a narrative visualization was also discussed in the interviews prior to the implementation, as described in Chapter 4.

6.2 Surveys

Whereas the user tests provided us with qualitative and quantitative measures through which to evaluate the prototype from multiple angles, surveys allowed us to collect a wider range of opinions on a shorter range of questions. Our survey was designed to be completed under five minutes, and some compromises were made as a consequence. We followed a number of criteria in an attempt to maximize the potential turn-out: minimize open questions, minimize the amount of text, and pose simple and straightforward questions. The survey was sent out to all members of the University of Porto, including students and teaching and non-teaching staff members. A test run of the survey was conducted prior to the official launch, in which five people were asked to provide comments and suggestions. The feedback gathered from this experiment was reflected in the final version of the survey sent out to the general public.

The survey questions can be seen in Table 6.8 and Appendix E. They were divided into two main topics: news reading habits and demographic information (questions S1 through S8), and analysis of the prototype (questions S9 through S12). The first section allowed us to establish the profile of the participant. We collected information such as the participant's age, their affiliation within University of Porto, the frequency with which they consumed online news, how familiar they were with visualizations in the news and their perception of them. The possible answers for question S8 are present in Table 6.9. The second and final section was designed to explore

	Question
S1	Age group
S2	Affiliation
S3	Occupation
S4	How frequently do you read online news?
S5	How overwhelmed do you feel with the amount of news to which you are exposed everyday?
S6	How frequently do you come across visual elements (e.g. charts) in online news?
S7	When a news article has visual elements (e.g. charts), how useful are they to the comprehension of that news article?
S8	Which of the following expressions better describes your news reading style?
S9	Rate the importance you attribute to the following visualization elements
S10	Would you be interested in using this tool as a complement to written news articles?
S11	Would you be interested in editing the visualization (e.g. ordering the entities in a different way)?
S12	Suggestions/comments regarding the visualization

Table 6.8: Questions asked in the survey.

how the general public might perceive a tool such as this one. First, the respondents were shown two (appropriately annotated) screenshots of the visualization and scene information module, a brief video clip demonstrating some of the core features (e.g. scrolling through the events and dragging the event nodes around), and a brief textual description of the interface and its elements. Keeping in mind the time constraints, we chose to hide the sidebar, as its interest in a static context is quite limited. This introduction was followed by a series of questions related to the perceived importance of each element (question S9), which the respondent was asked to rate from one to five, one being "not important at all" and five being "extremely important". Two quantitative, broader questions related to visualization followed, and an open-ended question regarding any comments or suggestions to the visualization marked the end of the survey (questions S10 through S12). We gathered 178 survey responses, a summary of which can be found in Table 6.10.

6.2.1 General information

The majority of participants (64.7%) are between 18 and 24 years old; Figure 6.3 contains the distribution of results. This is in line with our expectations, as the survey was distributed using

Options
I would rather have a panoramic view of the news, even if I cannot read every article in detail
I would rather read the news articles that I come across carefully, analysing all of the details
I have no preference for any of the methods listed above

Table 6.9: Possible answers for question S8.

Question	Answers	Mode
S1 - Age	178	18-24
S2 - Faculty	170	FEUP
S3 - Occupation	172	Student
S4 - Reading frequency	178	Several times per day
S5 - News overload	178	4
S6 - Visualization in news	178	3
S7 - Usefulness of visualizations	178	4
S8 - News reading preference	178	Panoramic view
S10 - Interest in the visualization	175	4
S11 - Interest in editing the visualization	177	3
S12 - Suggestions	14	-

Table 6.10: Summary of the survey responses.

a university-wide mailing list targeting the University of Porto's members. The top three faculties are FEUP (62), FCUP (25), and FLUP (21), and the overwhelming majority of participants (93.2%) are students. The participant pool varies little in terms of occupation, with a greater disparity observed in age group and faculty.

6.2.2 News reading habits

Figures 6.4 and 6.5 showcase the frequency with which the participants consume online news and the preferred method of doing so, respectively. The distribution of answers for questions S4, S5, and S6 can be found in Figure 6.6. For question S6, there does not appear to be a significant skew towards any of the ratings, which may indicate that, while visualizations are present in the news, they are still not supplanting text as the main source of information. When visualizations are present, around 73% of participants claim that they are useful to the comprehension of the news article. This may indicate that visualizations are a valuable addition to text articles. Finally, the answers to question S5 show a slight skew towards a positive rating; however, no strong conclusions can be drawn from this, as around 57% of participants provided a rating of "three" or below, which can indicate that information overload is not as prevalent in the average consumer base as we thought.

6.2.3 Visualization prototype

For question S9, the participants were asked to rank the importance of several visualization elements from one to five. The results can be see in Figure 6.7. Two elements stand out as the most important in the eyes of the participants: the events and the entities. The descriptions were ranked third in terms of relevance, which is in line with our expectations, as they provide crucial information on the events that make up the story. Surprisingly, the element which gathered the least amount of positive ratings is the title. This somewhat contradicts the findings extracted from the user tests, as the titles were one of the elements which received a considerable amount of positive

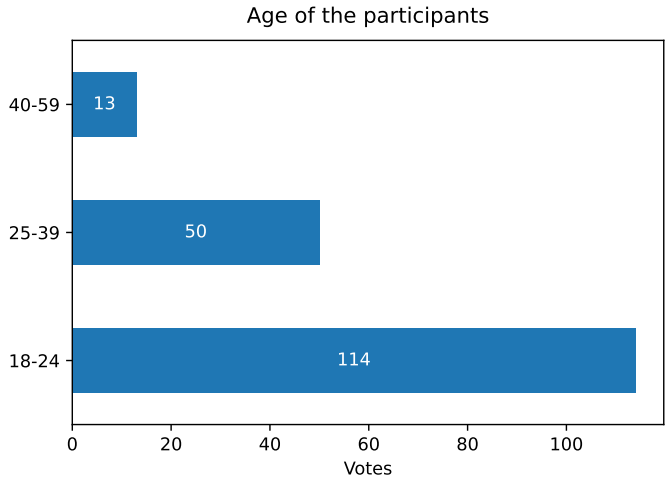


Figure 6.3: Age group of the participants of the survey.

feedback. The occurrences (i.e. the participation of entities in a particular event, encoded as a small colored circle inside of events) gathered the most neutral responses, which may indicate that their purpose is not obvious. All features had a majority of ratings above "four", which means we cannot single out any that may be unnecessary.

Figure 6.8 summarizes the results for questions S10 and S11. Participants showed more interest in using the visualization tool as opposed to editing the visualization, with 63% of participants rating their interest in using the prototype as a "four" or "five", in contrast with the 39% who provided those same ratings for the question related to editing. This is in line with what we expected, as editing the visualization is a feature that is perhaps more desired by power users. Moreover, the example we included in the survey did not lend itself to any major edits; there were no significant line overlaps and there were no obvious problems with the visualization's readability. The need for manual adjustments may have been clearer if the visualization was more cluttered.

6.2.4 Cross-references

We calculated the Spearman's rank-order correlation coefficients for questions S5, S6, S7, S10, and S11, and no significant values could be extracted. The highest value registered (0.356) correlated the interest in using the visualization with the interest in editing it; however, the relation between these two variables is not relevant for our purposes. As showcased in Figure 6.9, participants between the ages of 25 and 39 are the most enthusiastic about the application, with 24% rating their interest with a "five", compared to 17% of participants between 18 and 24, and to 8% of participants between 40 and 59. Despite some variation in the number of ratings equal to or under "two" (with the highest incidence being registered in the 40-59 age group), the majority of all age groups skews towards ratings equal to or above "four".

From the three faculties with the highest number of responses (i.e. Faculty of Engineering with 63 responses, Faculty of Sciences with 25 responses, and Faculty of Arts with 21 responses),

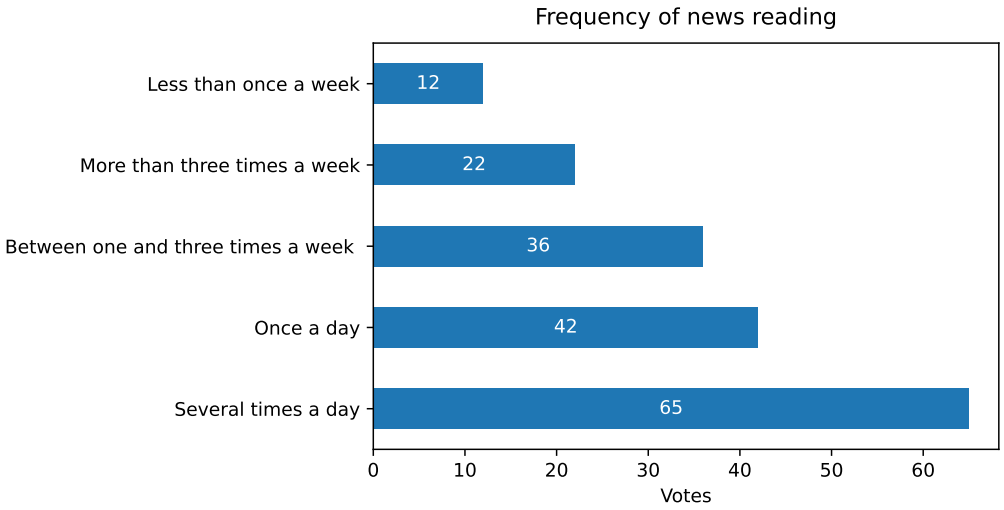


Figure 6.4: Frequency with which the participants consume news online.

members from the Faculty of Engineering (FEUP) showed the most interest, followed by members of the Faculty of Sciences (FCUP) and members of the Faculty of Arts (FLUP), respectively. These numbers do not include the participants who listed two faculties simultaneously. Figure 6.10 summarizes these results. Since all of the faculties mentioned above offer a variety of academic programs in a wide range of fields, we cannot narrow this down to a binary comparison between Science and Engineering and Arts. Nevertheless, it is interesting to note the lack of interest from FLUP students, since it may imply that the visualization, as it is, does not appeal to a non-STEM audience.

As for the relation between reading frequency and interest in the prototype, those who read the news more than three times a week show the highest percentage of "five" ratings, at 29%. Between those who read the news once a day and those who do so multiple times daily, the percentage of "three" and "four" ratings vary significantly, with those who read once a day providing 14% more "four" ratings and 16% less "three" ratings, with all other values remaining roughly the same between the two participant groups. This may indicate that participants who read the news multiple times per day do not benefit as much from visualizations focused on particular stories, since their consumption rhythm allows them to keep up with current news more easily. Additionally, people who read more often are already accustomed to the news sphere and may be more reluctant to changes. For people who read more than three times per week, the number of "four" ratings decreases, but the number of "five" ratings increases. For those who read between one and three times per week, the inverse happens. Finally, the participants who read the news less than once a week showcase the highest number of positive ratings; however, only 12 of the 178 responses belong to this group and the ratings may not be representative. Figure 6.11 demonstrates this distribution of ratings.

As for the interest in the application according to the participants preferred method of reading the news, there does not appear to be any significant trend, since the distribution of ratings is very

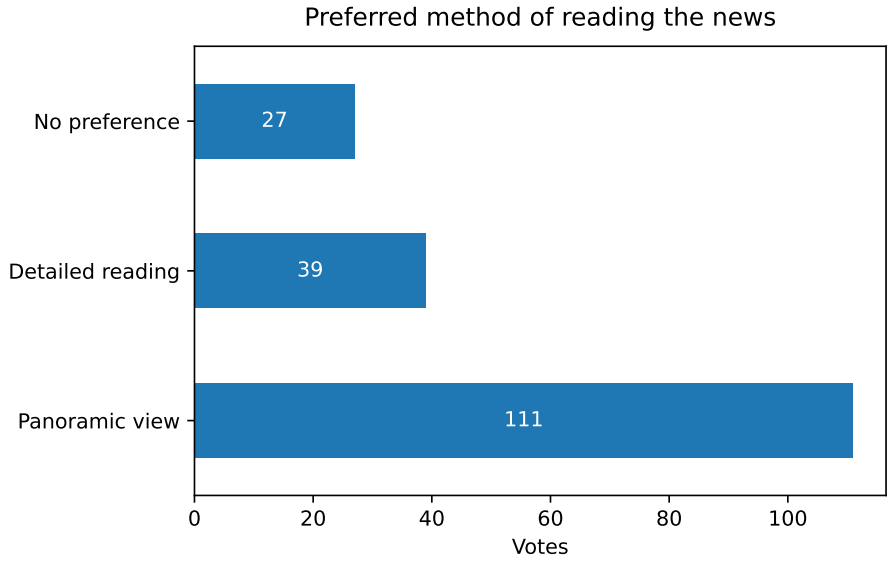


Figure 6.5: News reading methods of the participants.

similar between the two methods presented. This may indicate that the visualization adequately answers the needs of these two groups. These results can be seen in Figure 6.12

6.2.5 Open-ended question

For the only open ended question of the survey, we conducted a brief thematic analysis of the results. We gathered 14 answers of varied length, from which we then extracted relevant codes and subsequently organized them in themes. The result of this analysis can be found in Table 6.11. Four participants expressed interest in the visualization. One participant claimed that the tool allows for a detailed reading without losing a high level perspective. Another said that it would be interesting to study historic events with it. Nine participants expressed concern regarding the interface's appeal and usability. One participant, despite finding the visualization interesting, said that the design needs some work to be more accessible. Another one said that, at first glance, it is hard to interpret the visualization. Finally, five participants made some suggestions. These include adding standardized icons to the next and previous event buttons, adding photos, and adding various zoom levels to the visualization. Two participants said they wanted to see a timeline. While this is present in the tool, the video and images used in the survey only showed the timeline with event titles, not dates. Overall, while some found the visualization interesting, there was also a considerable number of negative comments related the interface usability. The general feeling seems to be that the visualization is interesting, albeit not quite accessible to the general public at this stage.

Codes	Example
Positive feedback	
Interesting	*I think it is a very interesting tool, it seems like an excellent way to visually complement a news story.*
Useful	*It allows for a detailed reading without losing the high-level perspective.*
Negative feedback	
Not accessible	*The visualization is hard to interpret at first glance, limiting its accessibility.*
No point in moving the events	*Regarding the video, I question the point of moving the events up and down.*
Not appealing	*The visualization is not simple nor appealing to the average reader in this format, only to those who are very interested and wish to study the topic in depth.*
Suggestions	
Icons instead of buttons	*The previous and next event buttons could use a standard icon instead of text, to stand out more.*
Hyperlinks to more content	*If the descriptions are directly extracted from a news article, it would make sense to have a link to that page or paragraph.*
Variable zoom	*It would be interesting to have a zoom option, in which the time scale shifts according to the level of zoom.*
Automatic scroll	*It would be interesting to have the option to automate scrolling through the timeline.*
Timeline	*In the future, I would like to see a timeline that would place the events according to their timestamp.*
Add photos	*If, for example, [the visualization] was centered around photos, then it could be more appealing.*

Table 6.11: Written feedback gathered in the survey.

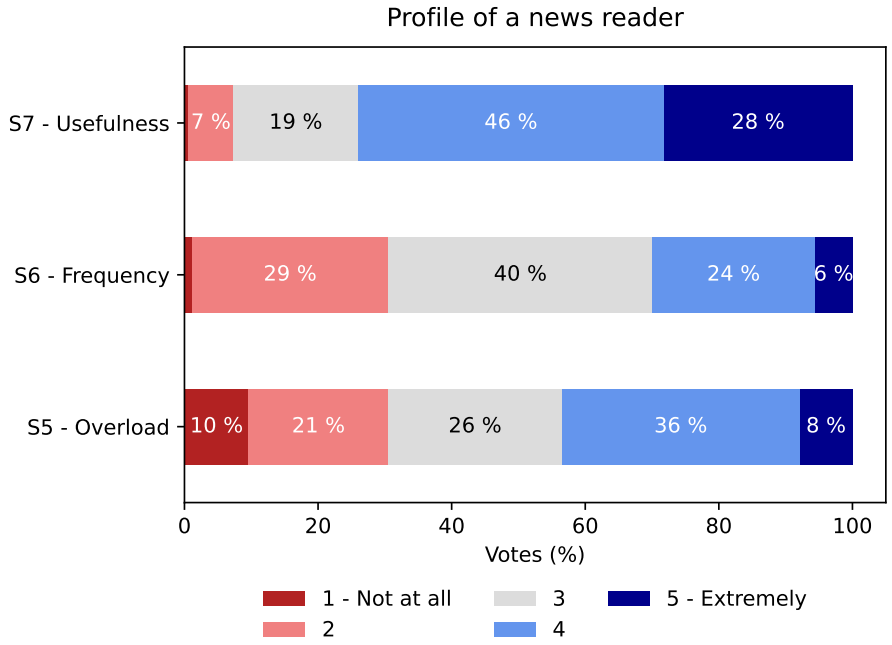


Figure 6.6: Questions pertaining to the reading habits of the participants.

6.3 Limitations

In the matter of user testing, we can identify three main causes for concern associated with the validity of the results: the difficulty of the tasks, the number of participants, and the sample characteristics. The tasks chosen for the test have a one-to-one correspondence with the functionalities provided by the prototype. The simplicity may have been a limiting factor in understanding how users might realistically interact with the application; however, we are still at an early stage of development, where the discovery of opportunities and weaknesses is more significant to us than observing specific workflows. Our primary goal was to determine which features are perceived as the most relevant, which are perceived as dispensable, and which are missing. Nevertheless, concerns regarding usability were also taken into consideration. The participant pool was made up of five journalists, some of whom participated in our previous interviews. We do not believe this poses a significant concern to the validity of the feedback gathered, given the aforementioned goals for the study and, particularly, since we do not generalize any of the findings. Regardless, in future tests it would be beneficial to both expand the number and diversity of participants, as well as to plan tasks based on high-level goals (e.g. discovering how often two entities interacted within a particular time frame).

As for the survey, the difficulty in accurately conveying the prototype's functionalities and the length of the questionnaire stand out as the most relevant limiting factors. In order to keep the survey short, we provided only two annotated images of the visualization, along with a brief 27-second video demonstrating some basic functionalities. A lot of information is lost when condensing the prototype into a still image and brief video. Besides, there are no opportunities for

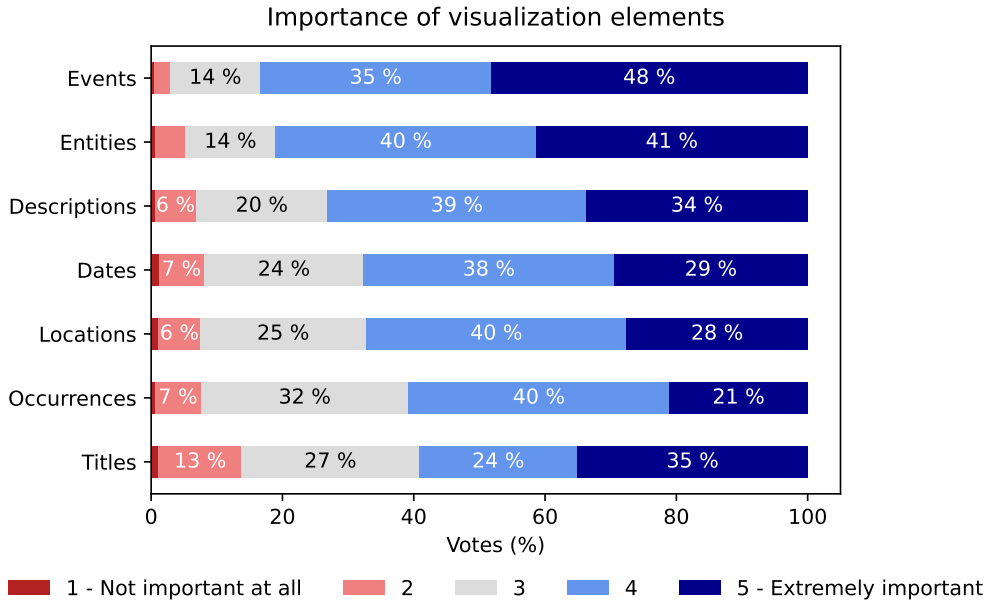


Figure 6.7: Importance of certain visualization elements.

interaction, a significant part of applications of this sort. Additionally, being conditioned on the amount of questions we could ask without potentially compromising the response rate meant we could not explore the application to the extent we desired. Despite this, we were still able to measure the overall interest in the prototype from a significantly-sized sample pool of 178 participants.

6.4 Summary

The feedback collected both during the user tests and in the survey corroborated the potential that storyline visualizations have in the study of journalistic narratives. The five journalists who

Validity concerns	Mitigating factors
User test	
The tasks were simple	We were more focused on discovering opportunities and weaknesses rather than on specific workflows
The sample pool was small and homogeneous	The results were not generalized
Survey	
Hard to demonstrate the prototype	-
There were few questions related to the prototype	We were still able to evaluate the overall interest in the application

Table 6.12: Concerns regarding the validity of the evaluation methods used, followed by arguments against these potential threats.

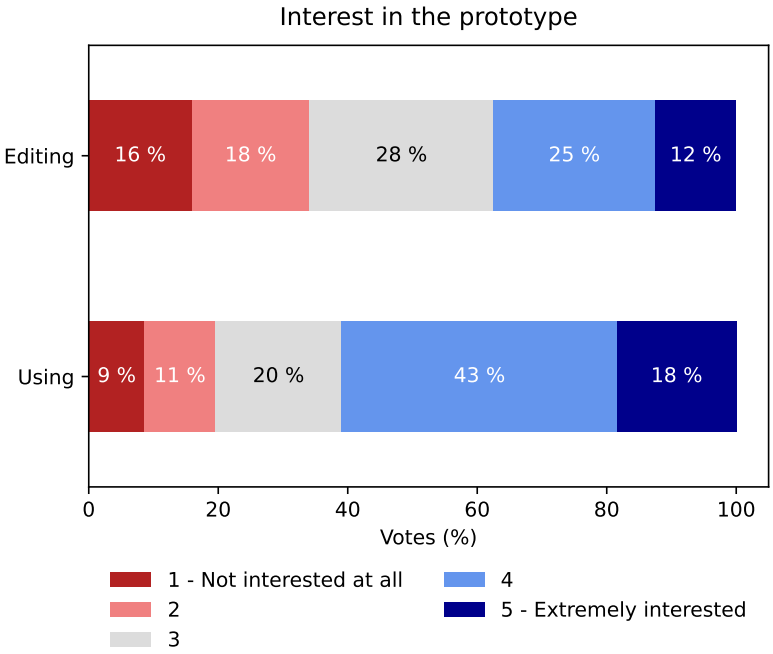


Figure 6.8: Interest in using and editing the visualization.

participated in the user tests claimed that application provides an interesting perspective on news stories, one that is likely to be of interest both to the average reader, as well as to the journalist. From the 178 survey responses collected, 106 (60.6%) provided a rating of "four" or above when asked quantify their interest in using the application. Overall, the participants generally found the application to be useful, but in need of some work in order for it to be more accessible to a broader public. Moreover, the suggestions made during these evaluation sessions can potentially serve as a guide to future development cycles for the prototype. It should be noted that, as with the interviews conducted prior to the implementation of the prototype, these evaluation methods have some limitations. Since the survey was designed to be completed in a short period of time, we could not delve too deep into the functionalities of the application, as that could have compromised the response rate to the survey. With respect to the user tests, the participant pool could be expanded in the subsequent sessions, both to include more participants, as well as to expand our target audience. Furthermore, the tasks could be redesigned in a way that better translates a potential workflow.

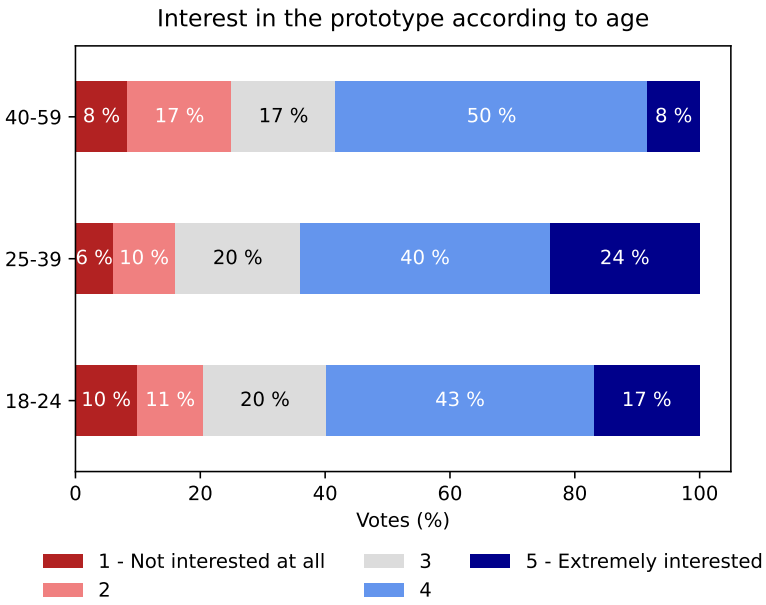


Figure 6.9: Association between the participants' age and their interest in the application.

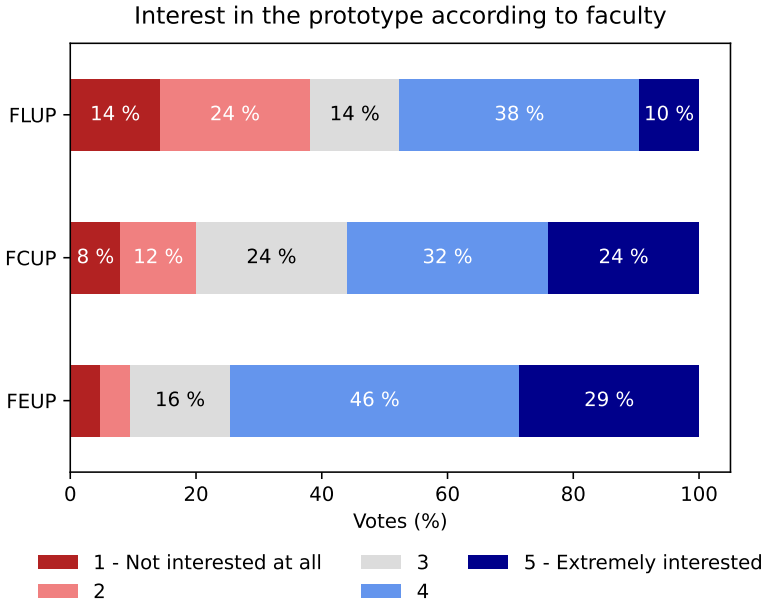


Figure 6.10: Association between the participants' association and their interest in the application.

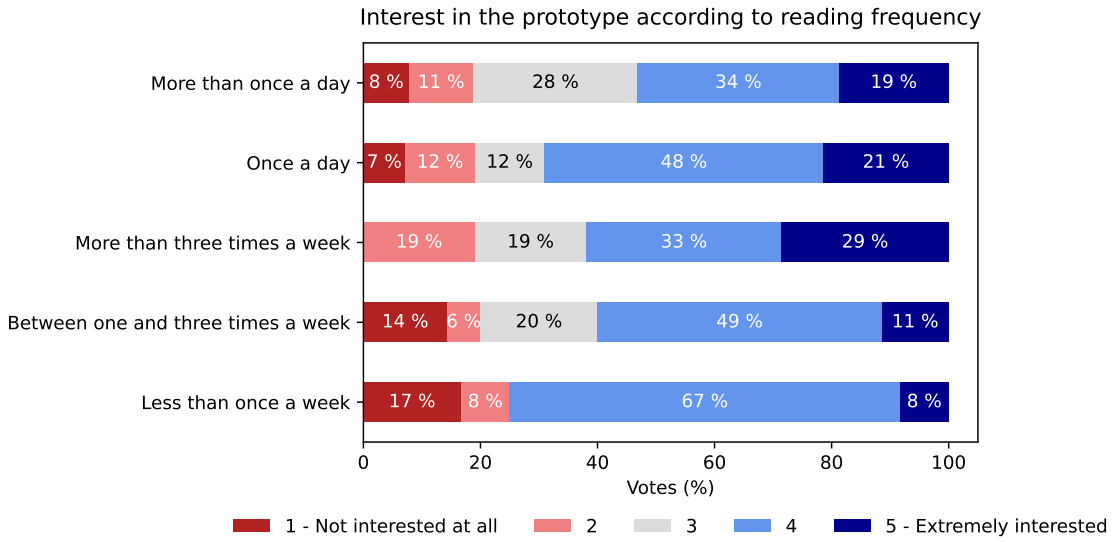


Figure 6.11: Association between the participants' news reading frequency and their interest in the application.

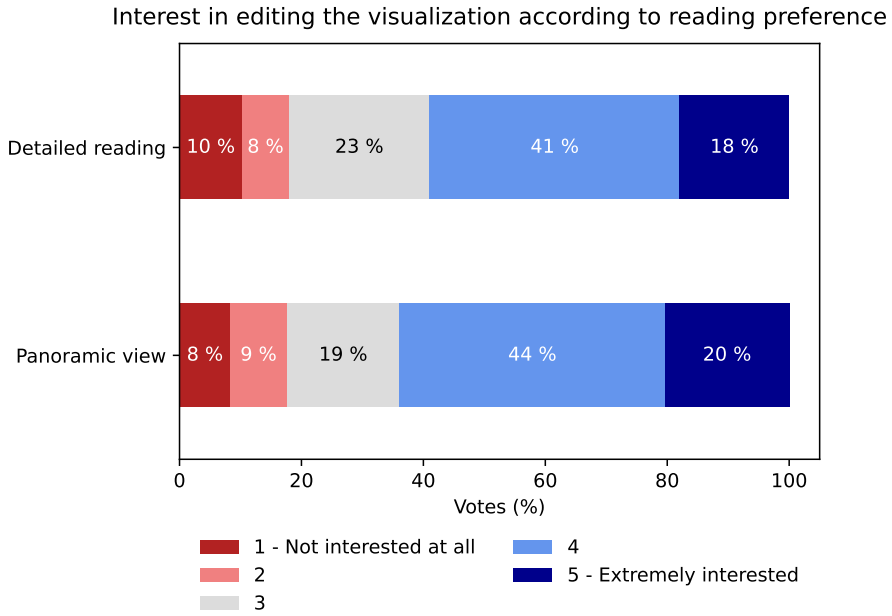


Figure 6.12: Association between the participants' news reading methods and their interest in the application.

Chapter 7

Conclusions

In this work, we have explored the topic of news visualization in a two-pronged approach. First, we surveyed existing work in order to understand how the news were typically visualized. While the majority of the collected research papers focus primarily on broader fields such as text analysis, they nonetheless explore news visualization, often as a way to validate the proposed techniques. For example, LeadLine [16] is presented as an interactive visual analysis system for text data that, among other things, can aid in the exploration of news events. It would be preferable to work solely with documents that focused exclusively on news content; however, we deemed that the few papers we found that abide by that criteria were not enough to conduct a thorough study. Furthermore, we would be remiss to exclude other works which tackle this subject matter with a lesser focus.

From the 17 information visualization papers studied, we gathered three main conclusions. The works are overwhelmingly focused on a large-scale, usually topical analysis. Large text corpora are condensed into topics that group similar text units together as a way to make sense of the input data. Some focus heavily on the time factor, exploring the evolution of documents across a time period. This can be seen in works such as TextFlow [13] and TIARA [42]. Others, such as Story Tracker [35] and the work of Krstajić et al. [33] share this focus on temporal development, but target news content specifically. Overall, the individuality of the input data is somewhat lost. For example, with this type of analysis, we might conclude that a particular time period was rife with news pertaining to crime, but we can't easily follow a particular court case we might be interested in, as this information was absorbed into a broader, abstract topic.

The intermediate representations of the textual data used in the works also contribute to this loss of individuality. The input data commonly goes through some sort of text processing which reduces a document to a set of words that aim to summarize it, thus significantly reducing its semantic complexity. An example of text summarization techniques can be seen in TIARA [42], with its implementation of the Latent Dirichlet Allocation model. When analysing some news narratives, their semantic intricacies may be desirable, as details important to the comprehension

of the underlying story (e.g. interactions between entities) can be lost during the process of summarization.

Finally, even in works which specifically tackle the news, the narratives told within those news stories are often overlooked. Characteristics such as narrative structure and the interaction between the parties involved do not show up frequently in the analysis conducted by the works reviewed, with some exceptions such as the work of Vossen et al. [60]. Other systems, such as Story Tracker, do take into consideration the complexities of news stories, how different news articles may be related to a single narrative or how some narratives may split into two distinct ones. Nevertheless, they still rely mostly on topics and text summaries, making it harder to clearly understand the details of the events, who participated in them and how those entities interacted with each other.

For the second part of our exploration of news visualizations, we decided to focus on the question of narrative-driven perspectives brought up during the survey. We wanted to understand how visualizations with stronger ties to narrative elements, such as StoryFlow [43], could be applied to the news domain. We started by conducting a series of semi-structured interviews with five journalists in order to understand the problem space more thoroughly and, ultimately, to assess whether the approach we had in mind was viable. From the feedback we gathered, we conclude that the journalists generally agreed that visualizations can help in understanding news stories, if adequately designed. While not all journalistic narratives are appropriate for a visualization such as a storyline, similar to the one in StoryFlow [43] or in the comic book narrative chart of Munroe [46], news stories can be similar in structure to the more traditional narratives visualized in the aforementioned systems.

Following the positive feedback we collected in these interviews, we implemented a browser-based prototype centered around a storyline visualization which highlights the events and entities in a story. This visualization tool takes as input a adequately formatted JSON document detailing the events and entities of a story, and presets said elements in such a way that facilitates the study of the underlying narrative: how events relate to each other, how entities interact with one another, how events are spread out in time, and so on. We then evaluated the implemented solution through two methods, a survey to the members of the University of Porto, and a user test with five journalists, two of which also participated in the interviews mentioned above.

Overall, the participants of the user test enjoyed the application, nevertheless pointing out some rough edges. While they considered the storyline visualization itself to be simple and quick to interpret, the rest of the interface was deemed confusing and too complex. One of the most common complaints pertains to the excess of options available, which may be overwhelming to the average consumer. They also provided a number of modifications that would be interesting to explore in a future developments of the prototype. With the surveys, we wanted to reach a broader audience and determine whether a visualization tool such as the one we developed would interest the general population. Given the brevity of the questionnaire, we were limited in terms of which parts of the visualization we could evaluate. Regardless, from the brief demonstration we included in the survey, the participants showed considerable interest in using the application as a complement to news articles. Similarly to the user tests, the comments left by some participants

expressed concern about the usability of the application, which they found lacking in clarity and appeal.

7.1 Main contributions

Our main contributions stem from the survey conducted on the topic of news visualization and from the publicly available and open-source implementation of a visualization prototype which focuses on exploring journalistic narratives. To our knowledge, the field of news visualization has not been analysed as in-depth as we have done so in this document, with surveys frequently focusing on broader topics such as text visualization. Through this survey, we were able to identify gaps in the literature, namely in how news stories are visualized and how techniques from other visualization fields, such as narrative visualization, could be re-purposed. With our prototype, we demonstrated that storyline visualizations can be successfully applied to the news domain and hope to have sparked interest in the application of other narrative-driven visualization techniques to study news stories. The evaluations conducted have shown that the general public, as well as people with stronger ties to Journalism, have interest in an application of this sort. Furthermore, the feedback and suggestions gathered have set the foundations for any potential future iterations of the project.

7.2 Future work

As we have said previously, the feedback gathered during both the user tests and surveys has opened up paths for future development. One of the most pressing questions which we could tackle pertains to the complexity of the interface, since it was one of the most frequently mentioned issues during the user tests. The poor accessibility of the application is also a concern that was brought up in the surveys that we could focus on alleviating. After another round of product development, it would be interesting to conduct more user tests, both with Journalists and with people from other fields. The scope of this document did not allow us to expand on the question of automatic textual processing, which would also be a possible avenue for future work.

Appendix A

Interviews

A.1 Questions

	Question
Consumption of news	Who are the main culprits behind the information overload we come across when reading the news?
Journalistic narrative	What characterizes the journalistic narrative?
	What are the differences and similarities between traditional narratives and journalistic narratives?
Visualizations in the news	Which approach is more enticing to the average reader: high-level visualizations or focused, narrative-driven visualizations?
	Would a narrative-driven visualization make more sense in a day-to-day context or as way to analyse distant narratives?
	Do news stories have the potential to provide the content needed for a narrative-driven visualization?

Table A.1: Questions explored during the semi-structured interviews.

A.2 Original quotes

Code	Example
Helps situate the reader	"Isto diz às pessoas onde é que estão, porque às vezes, nestas histórias complexas, utilizar este tipo de alternativa para situar visualmente as pessoas na história, eu acho muito útil."
More captivating	"Se a narrativa visual for bem feita, tocamos às vezes em pontos dos leitores e conseguimos que eles se embrenhem na história de uma forma diferente do que se for um texto (...)"
	"A força que (...) estas narrativas visuais têm é que surpreendem os leitores, apanham-nos desprevenidos (...) às vezes são construídas de maneira a falar para eles diretamente (...)"
Generates more interest	"Histórias visuais são muitas vezes as que geram mais interesse."
Occasionally better than text	"E acho que podia fazer sentido para o leitor no sentido de ajudar a entender melhor aquele tipo de informação que no texto corrido provavelmente se perderia mais."
	"(...) visualizações de cronologias, por exemplo o ano da pandemia. (...) é um caso em que é muito mais eficaz mostrar assim do que escrever um texto sobre isto que seria necessariamente um texto aborrecidíssimo, quase uma ata."
	"(...) num texto jornalístico evitamos ir de A ao Z, tentamos começar pelo mais importante que pode ser o M. Geralmente não seguimos uma ordem cronológica, mas muitas vezes a cronologia é mesmo o mais eficaz."
Allows for interaction	"Com a interatividade, quando a conseguimos acrescentar, facilita imenso (...)"

Table A.2: Codes associated with the advantages of visualizations in the news in their original language.

Code	Example
Steep learning curve	"Precisa de haver uma fase intermédia de habituação da audiência a este tipo de representação."
	"Algumas destas coisas exigem níveis de abstração de leitura que algumas pessoas não dominam."
Potentially noisy	"Se eu não extrair dessas visualizações uma ideia imediata do que ela me está a tentar transmitir, então ela produz ruído (...)"
	"Se forem demasiado bonitas 'for the sake of' serem bonitas cria ruído e distraí-nos da história."
	"O formato tem de estar ao serviço da história e não ao contrário."
	"Isto [a visualização] não me ajuda a não estar saturado, isto não me descodifica nada. (...) Ou seja, não me ajuda sequer a interpretar, mostra-me. Coloca-me num nível indesejado de informação que eu queria ler de outra forma."
May not fit everyone's style	"Eu acho que o ritmo que eu pessoalmente tenho de consumo de informação não me permite olhar para este tipo de visualizações como deve ser."
Requires more resources	"Qual é o tempo necessário para isso ser feito? Temos falta de recursos e eu questiono a viabilidade (...)"

Table A.3: Codes associated with the disadvantages of visualizations in the news in their original language.

Code	Example
Hard to updates stories online	"Na 'net', ter coisas que têm de ser atualizadas (...) é muito difícil tu conseguires prender o leitor àquela história."
Need distance from the facts	"Acho que precisamos de ter distância dos factos para perceber o que é importante"
Need a complex story	"Para fazer uma coisa deste género, é preciso que a história tenha muitos fios. Uma história ou um evento de um dia geralmente não gera isto."
Possible to accommodate for both	"Acho que podemos usar esta técnica nas duas situações."

Table A.4: Codes associated with the discussion regarding the use of narrative-driven visualizations as an exercise of retrospection or as a day-to-day tool in their original language.

Code	Example
High-level analysis	"Se olharmos para aquele consumidor comum (...) eles serão muito mais propensos a ver uma visualização como a do Alberto [Cairo], não é?"
	"Geralmente, aquelas histórias alto-nível às vezes até interessam mais ao jornalista ou a um público muito específico, são quase meta-histórias."
Detailed, focused analysis	"Pensando do ponto de vista do leitor, parece-me que tem mais utilidade as visualizações que estão ao serviço da história."
	"Um leitor (...) assíduo que tenha uma assinatura, acho que aí já gostaria e valoriza esse lado (...) de criatividade na apresentação da notícia."

Table A.5: Codes associated with the readers' preference for high-level analysis or a detailed, focused one in their original language.

Code	Example
Inverted pyramid	"Falando da notícia, ela tem por regra uma forma que é a pirâmide invertida, que segue do mais importante para o menos importante."
	"No jornalismo, a ideia é 'estragar' logo a história"
Dynamic narrative focus	"Em momentos diferentes, o foco é diferente (...) o primeiro foco narrativo são dados absolutamente factuais (...) depois começa a surgir o trabalho de contexto (...) começo a acrescentar densidade com esses relatos pessoais"
Can be similar to traditional narratives	"A narrativa jornalística pode potencialmente ter todos os elementos que uma história como um filme tem. A ordem é outra (...) Mas que a história em si pode ter tudo o que uma boa narrativa tem, as personagens, os arcos das próprias personagens, os momentos pivot, isso sim, pode ter."
The "human factor" is important	"Eu crio relação com o meu leitor ou com o meu ouvinte transformando uma história global numa história personalizada, é a melhor maneira"
	"O jornalismo procura a ancoragem no elemento humano, é uma regra fundamental do trabalho."
Function over form	"Acho que a ideia de design de que a forma submete-se à função também se aplica ao jornalismo."

Table A.6: Codes associated with the characteristics of news articles in their original language.

Appendix B

Input Data

B.1 JSON Schema

Listing B.1 depicts the JSON Schema used by our implementation.

```
{
  "$schema": "https://json-schema.org/draft/2020-12/schema",
  "title": "News story",
  "description": "Main story elements necessary to the creation of a
      storyline visualization",
  "type": "object",
  "properties": {
    "title": {
      "description": "The title of the story",
      "type": "string"
    },
    "characters": {
      "description": "Entities who participate in the story",
      "type": "array",
      "items": {"$ref" : "#/$defs/character"}
    },
    "scenes": {
      "description": "Events in a story",
      "type": "array",
      "items": {"$ref" : "#/$defs/scene"}
    }
  },
  "$defs": {
    "character": {
```

```
      "type": "object",
      "required": [ "id", "name" ],
      "properties": {
        "id": {
          "type": "string",
          "description": "The ID of the entity"
        },
        "name": {
          "type": "string",
          "description": "The name of the entity"
        },
        "affiliation": {
          "type": "string",
          "description": "RGB value of the entity line in the
              visualization in the format 'rbg(R, G, B)'"
        },
        "synonyms": {
          "type": "array",
          "descriptions": "Other names by which this entity can be
              referred to in the description texts",
          "items": {"type" : "string"}
        }
      }
    },
    "scene": {
      "type": "object",
      "required": [ "characters", "description", "title" ],
      "properties": {
        "characters": {
          "type": "array",
          "description": "ID's of the entities that participate in
              this event",
          "items": { "type": "string"}
        },
        "description": {
          "type": "string",
          "description": "Description of this event"
        },
        "title": {
          "type": "string",
```

```
          "description": "Title of this event"
        },
        "date": {
          "type": "string",
          "descriptions": "Date of this event. Must be in a valid
              date format"
        },
        "location": {
          "type": "string",
          "descriptions": "Location of this event"
        }
      }
    }
  },
  "required": [ "title", "characters", "scenes" ]
}
```

Listing B.1: JSON Schema.

B.2 Original JSON format

Listing B.2 depicts the original JSON format used by the D3 Narrative Charts implementation [19].

```
{
  "characters": [
      {
        "id": "R2D",
        "name": "R2-D2",
        "affiliation": "light"
      },
      {
        "id": "C3P",
        "name": "C-3PO",
        "affiliation": "light"
      }
    ],
  "scenes": [
    [
      "R2D",
```

```
      "C3P"
    ]
  ]
}
```

Listing B.2: Original JSON format.

B.3 Modified JSON format

Listing B.3 depicts the modified JSON format used by our implementation and adapted from the original format discussed in the previous section.

```
{
  "characters": [
    {
      "id": "rioters",
      "name": "Rioters",
      "affiliation": "rgb(204, 0, 0)",
      "synonyms": ["proud boys", "large crowd", "mob", "ashli babbit
          "]
    }
  ],
  "scenes": [
    {
      "characters": ["rioters"],
      "description": "Members of the Proud Boys, a far-right group,
          join protesters, who had been assembled on the Capitol lawn
           since 10 a.m.",
      "title": "Proud Boys",
      "date": "2021-01-06 11:37:00",
      "location": "Capitol"
    }
  ]
}
```

Listing B.3: Modified JSON format.

Appendix C

Prototype

C.1 Instruction manual

A live version of the application is available; alternatively, it is possible run the application locally. First, clone the source-code from the public repository. Then, using Python's *http.server* or an equivalent service, start a local server in the application's directory. Listing C.1 summarizes the process.

```
git clone https://github.com/marianafcosta/news-viz-protos.git

# Using Python 3.8.6
python3 -m http.server
```

Listing C.1: Installation instructions.

C.2 Screenshots

Figure C.1: The about page, which contains some information about the project, source-code, and licenses.

C.3 Use case

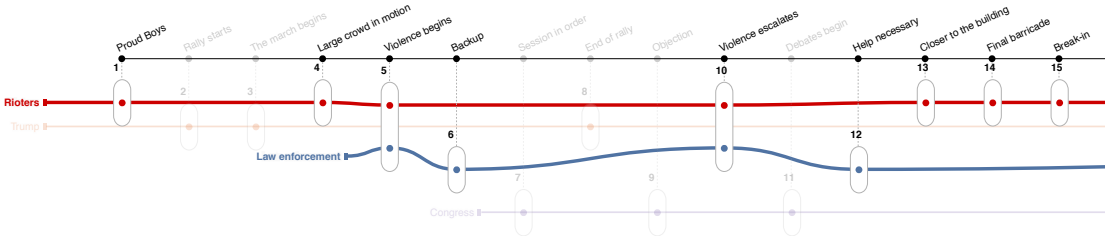


Figure C.2: Focus on the “Rioters” and “Law enforcement” in the Capitol riot storyline.

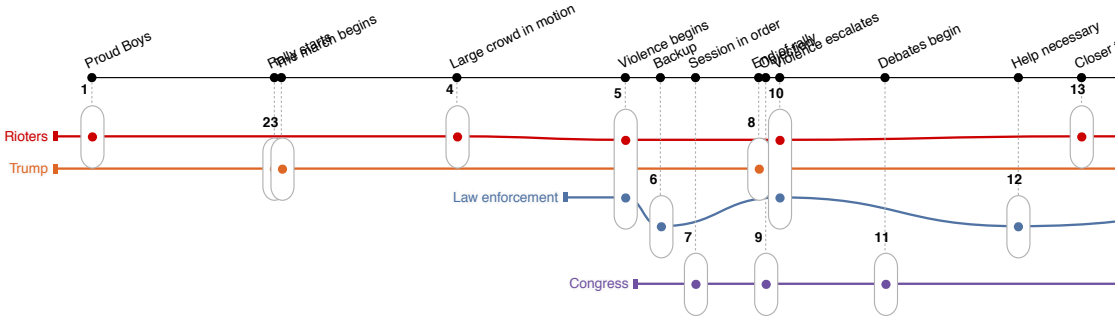


Figure C.3: Spacing out the events in the Capitol riot storyline according to their timestamp.

Appendix D

News visualization tools

D.1 Screenshots

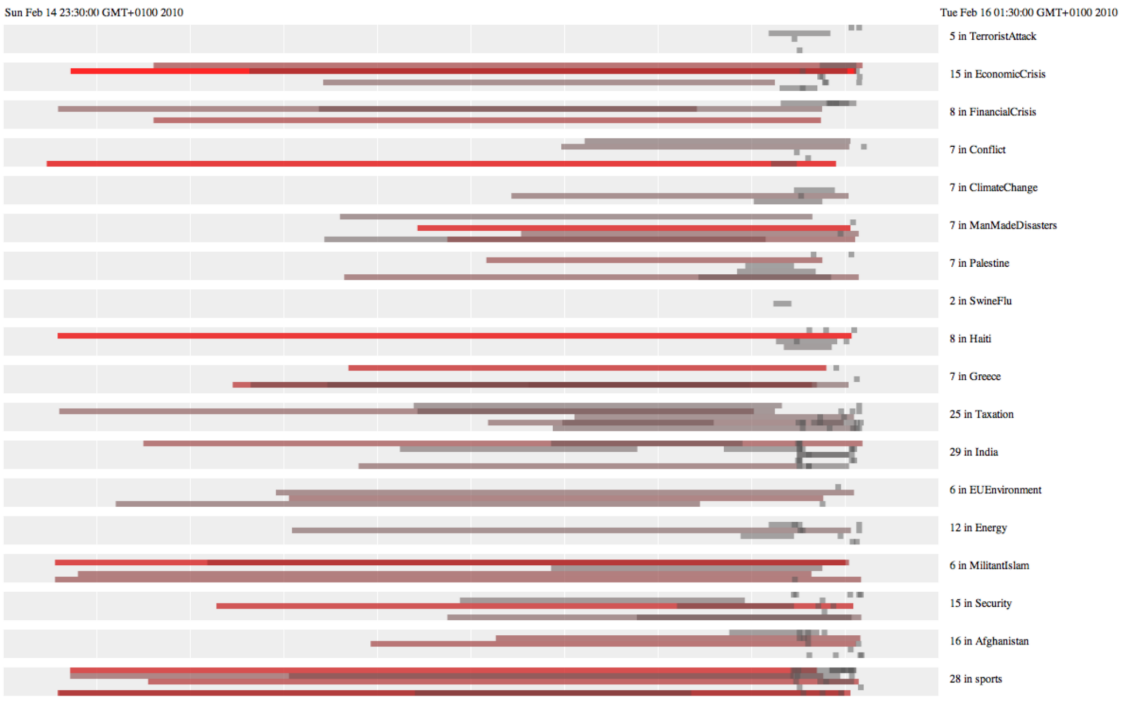


Figure D.1: Visual analysis of news streams by Krstajić et al. [33].

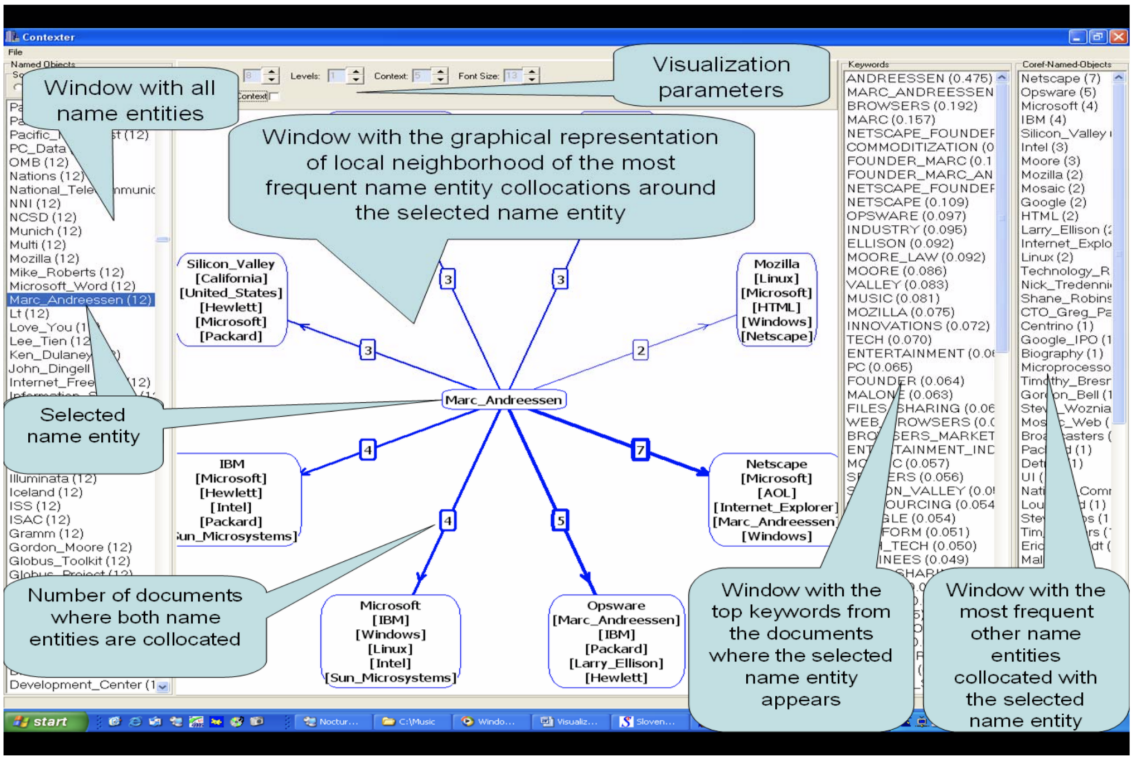


Figure D.2: Interface of Contexter [26].

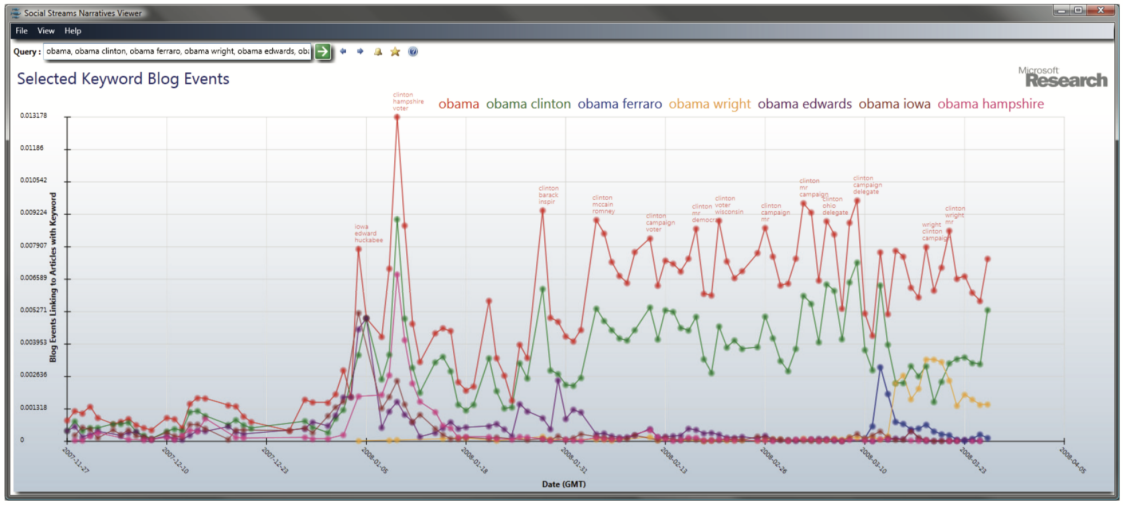


Figure D.3: Interface of Narratives [21].

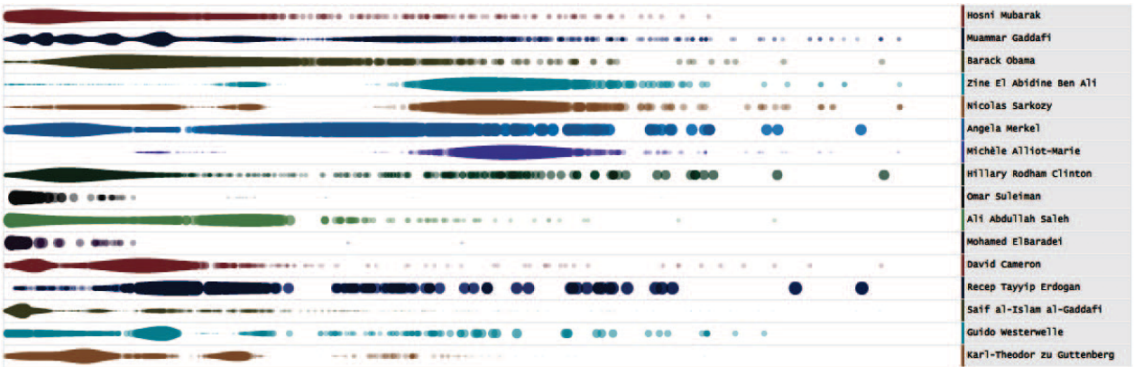


Figure D.4: Interface of CloudLines [34].

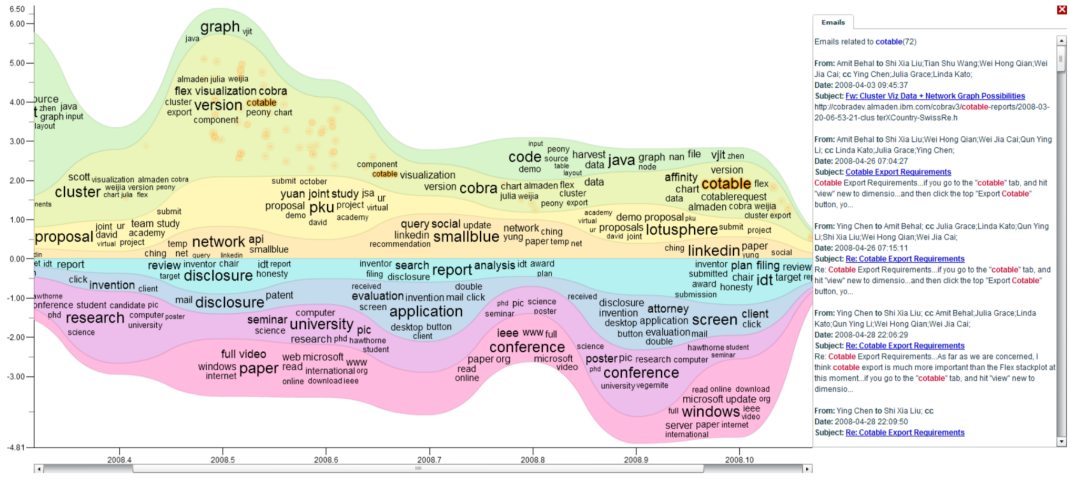


Figure D.5: Interface of TIARA [42].

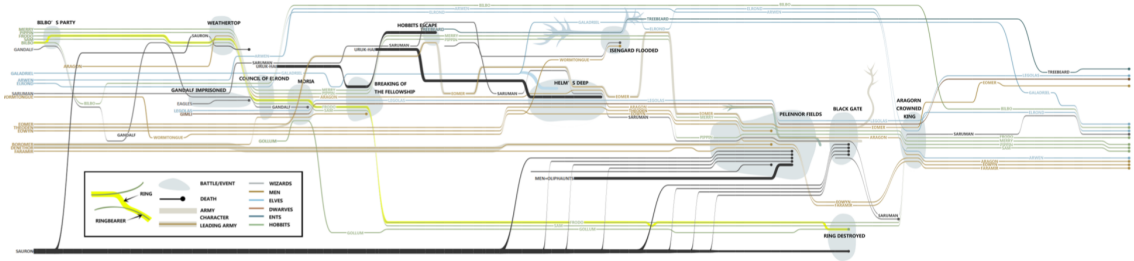


Figure D.6: Interface of StoryFlow [43].

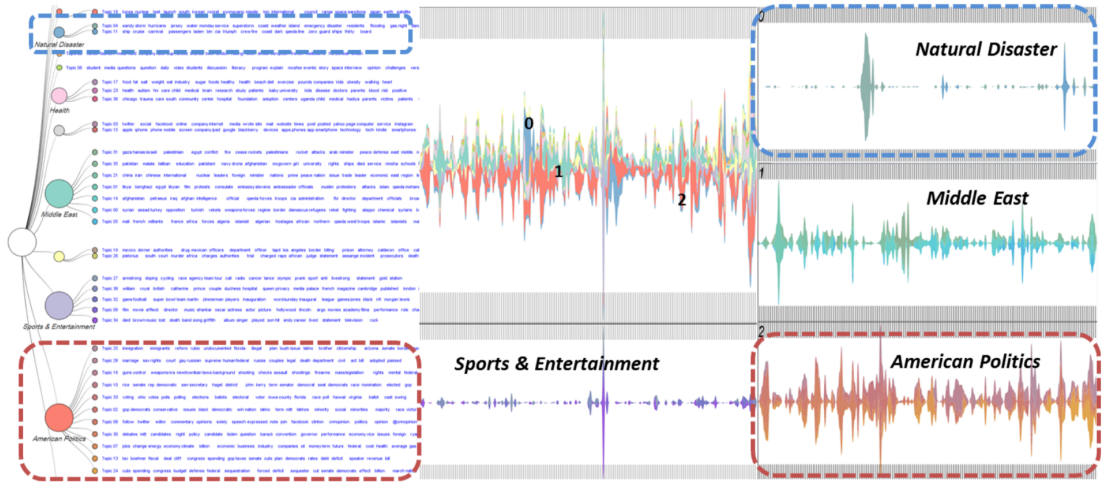


Figure D.7: Interface of HierarchicalTopics [17].

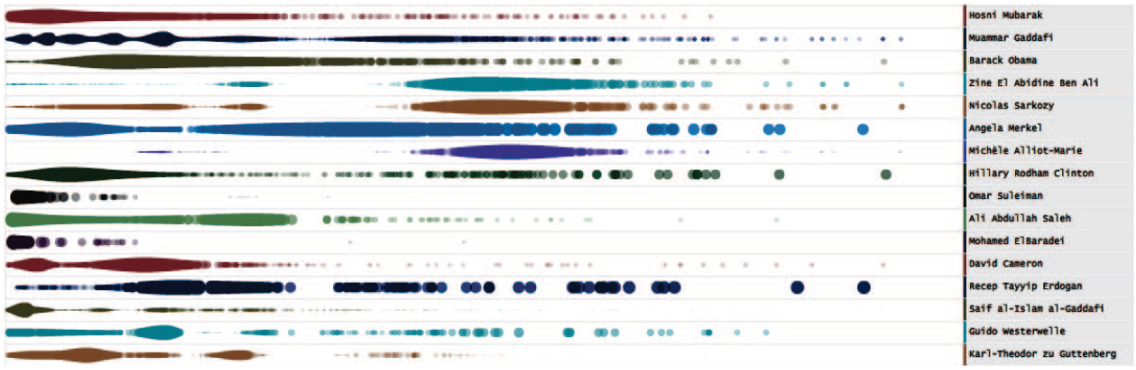


Figure D.8: Zoomable metro maps by Shahaf et al. [51].

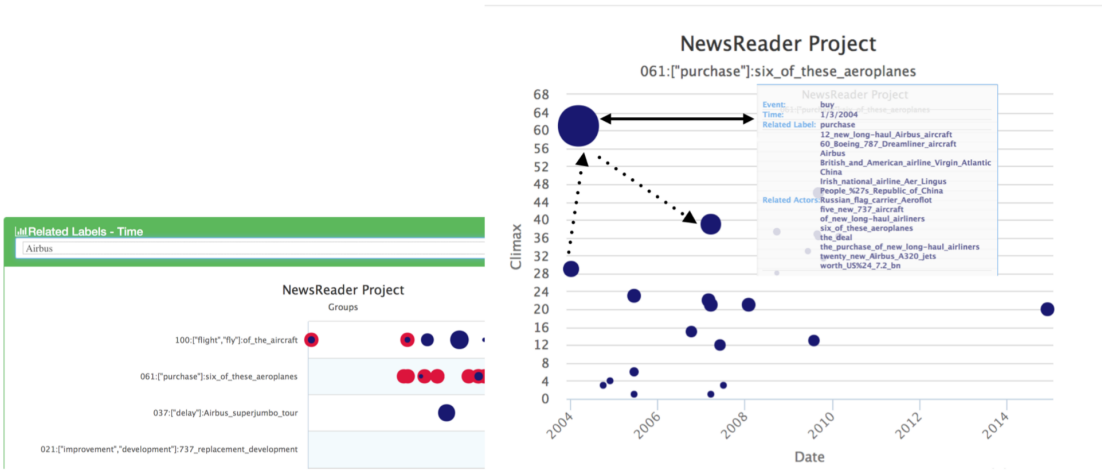


Figure D.9: Interface of the visualization tool provided by Vossen et al. [60].

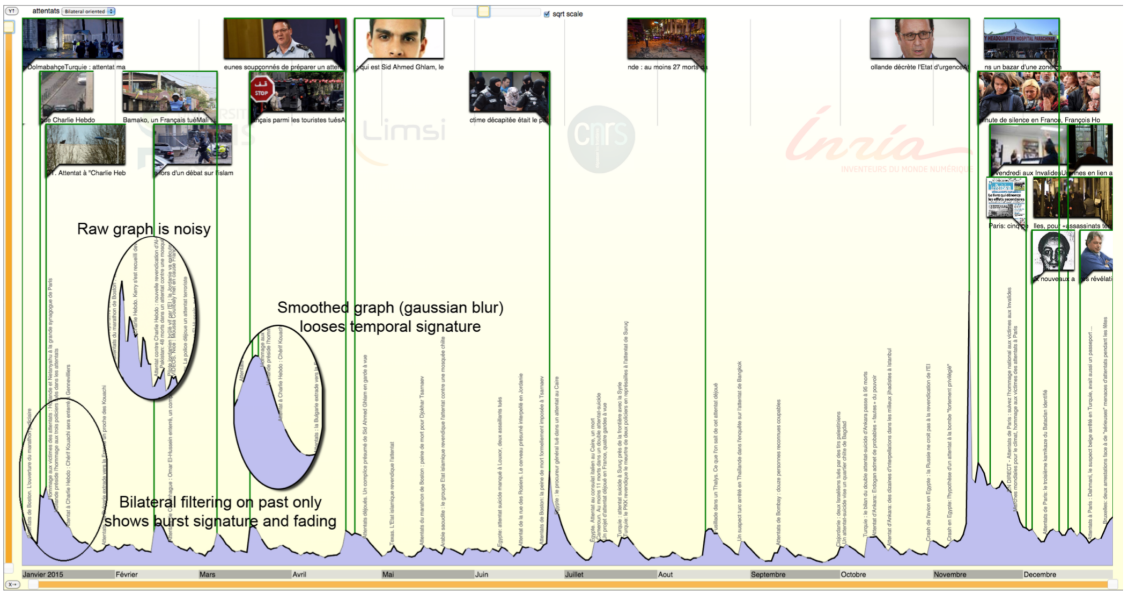


Figure D.10: Visualization of thematic timelines Tannier and Vernier [57].

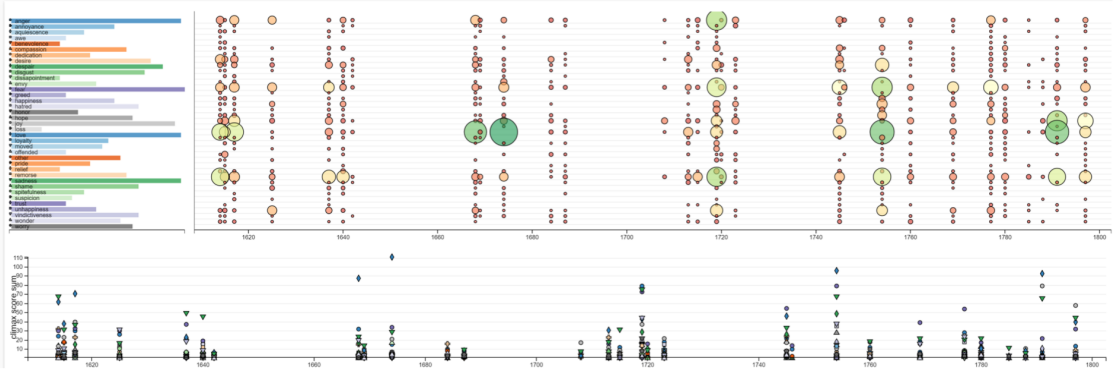


Figure D.11: Interface of Storyteller [59].

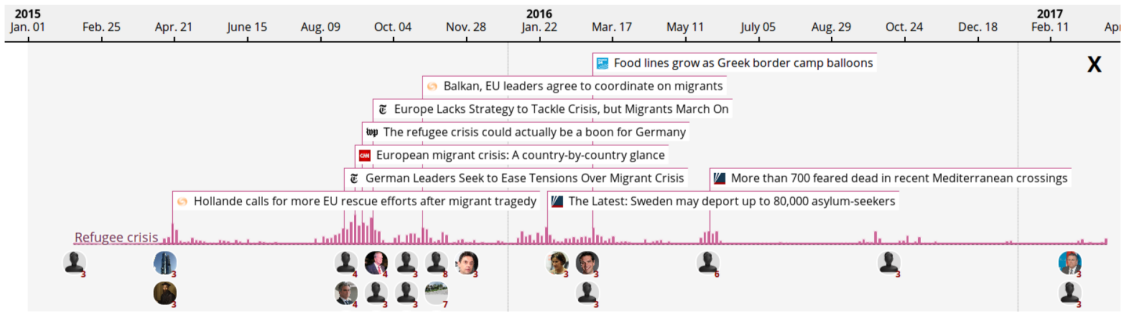


Figure D.12: Interface of newsLens [37].

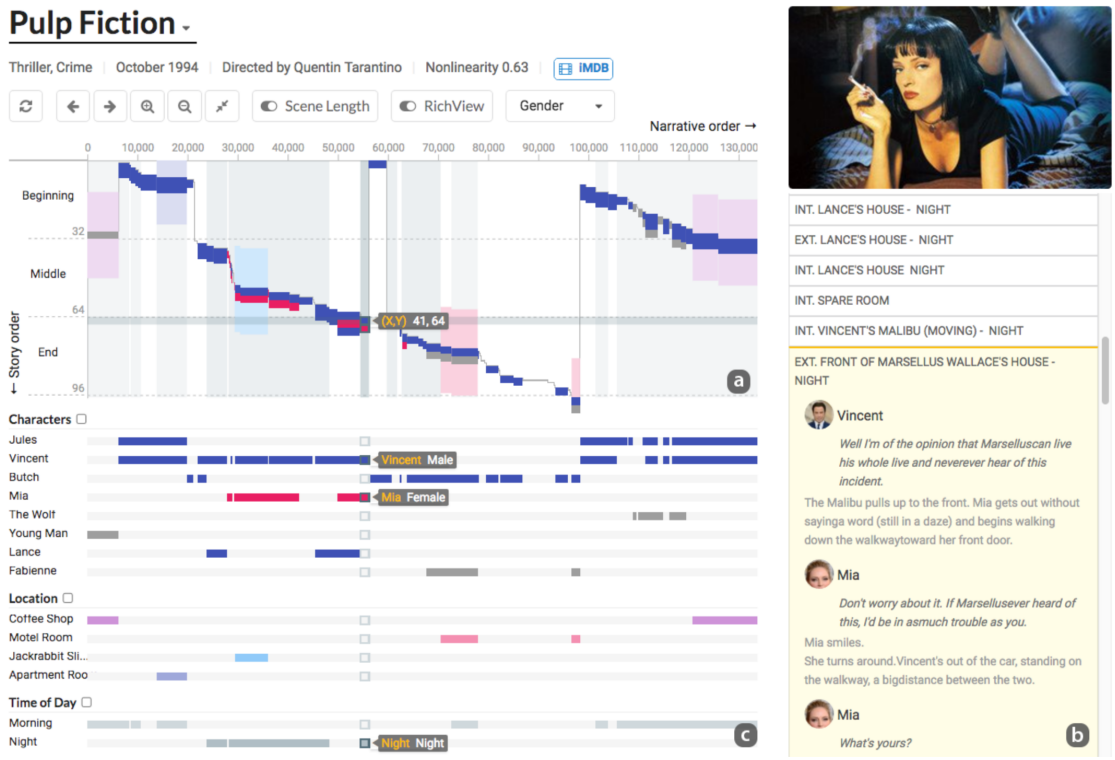


Figure D.13: Interface of Story Curves [32].

Appendix E

Evaluation

E.1 Original quotes

Codes	Example
Appealing visualization	"Eu acho a visualização muito simples e fácil, simples no bom sentido, de fácil absorção."
Situates readers	"Permite uma visualização rápida de quem participou, quando, onde (...)"
Different way of seeing things	"Acho que está bastante interessante, uma forma muito diferente de ver as coisas."

Table E.1: Codes related to the positive feedback gathered from the journalists in their original language.

Codes	Example
Too many options	"Não sei se será necessário tanto detalhe de controle."
	"Eu acho que isto é capaz de ser opção a mais (...) quando se está num jornal com um público gigante e que provavelmente não vai ter muito tempo para estar aqui (...) as pessoas não dedicam tanto tempo a uma coisa."
	"(...) depois fica programa informático e não aplicação jornalística, que geralmente tem mesmo muito poucas opções."
	"(...) se eu tiver muitas opções de controle eu às vezes posso perder-me nelas e não prestar tanta atenção à história"
Unclear interactive elements	"A linguagem [do botão de espaçamento temporal] (...) como entrei há um bocado, sem saber nada disto, não iria chegar lá."
	"Ao selecionar uma coisa, ele faz a mudança e depois não preciso de clicar [para efetivar a seleção]"
	"Os botões não têm muito destaque, era um bocadinho difícil para mim ler, e têm muita coisa escrita."
Redundant visibility levels	"Eu diria que são em parte redundantes, a primeira opção faz a mesma coisa (...) não vejo utilidade em apagar."

Table E.2: Codes related to the negative feedback gathered from the journalists in their original language.

Codes	Example
Reset the visualization	"Tem que se ter a opção de reiniciar, voltar à visão original."
Select/highlight entities	"O que eu achei mais útil é as entidades (...) poder sublinhar, destacar uns para facilitar certos fios narrativos."
Space out the events	"(...) é importante, para se ter a noção exata de que houve aqui uma hora em que isto foi mesmo dramático"
	"Tirando esta grande confusão que se gera junto destes pontos, ajudará a perceber algumas dimensões (...)"
Event scroll	"É provavelmente a forma mais natural de navegar esta narrativa."
	"Esta é a maneira mais razoável de navegar, não é? Passo a passo."

Table E.3: Codes related to the features journalists explicitly claimed were the most important in their original language.

Codes	Example
Sidebar	
Concise wording	"Ajudava serem frases mais concisas (...)"
Hover to explain options	"(...) ter um ícon que me explicasse a sério o que faz aquela opção."
Replace buttons with icons	"Setinha de 'fast-forward' também poderia ser mais simples."
Multiple option screens	"Deveria haver uma diferenciação entre as opções básicas e as opções mais de power user."
	"Num primeiro ecrã, ter as opções básicas e, eventualmente, se for mesmo necessário ter outro, quase um nível de opções avançadas (...)"
Hide characters automatically	"Por exemplo, eu ao selecionar 'rioters' e 'Trump', por exemplo, automaticamente destacaria essas duas escondendo as outras."
Standardize the resets	"Se calhar repetia-o [o ícon de 'reset'] nas várias opções (...)"
Different wording for the resets	"Usar verbos diferentes provavelmente ajudava a manter na minha cabeça o que é que era que fazia cada uma das coisas."
Event info on highlight	"Se eu tenho um gesto deliberado de ir selecionar um evento porque é que ainda necessito de ir ali à história para ter mais informação sobre esse evento?"
Visualization	
Show info on title click	"O meu primeiro instinto foi carregar no círculo [para focar no evento]"
Toggle entity on click	"Se eu ao clicar aqui também não podia fazer desaparecer os 'rioters' (...)"
Entities selected by default	"Ele entrava na história e por defeito tinha as quatro [entidades] selecionadas com os vistos, e depois é quase intuitivo, se tenho visto eu posso tirá-lo"
Scroll with drag	"Estava na esperança de que ao fazer drag ele me permitisse fazer scroll."
Vertical storyline	"Questiono-me se não faria sentido esta linha temporal ser na vertical e não na horizontal."
Add date to title	"Achava mais útil ter as duas coisas [data e título]."
Format the dates	"A formatação das datas (...) se fosse de alguma forma mais inteligível (...)"
Event information module	
Highlight event info	"Pergunto-me se não ajudava que esta navegação por eventos estivesse de alguma forma mais evidenciada (...)"
Side-by-side with vis.	"Mas a informação sobre o evento estar mais ou menos no mesmo plano de leitura do evento talvez não seja desagradável"
Other	
Add a top bar	"Tentava criar um menu em cima, sem o peso da barra [lateral]."
Other download formats	"Se em vez de ser um PDF fosse um PNG isto seria útil para as redes sociais."
	"Pergunto-me se não seria útil ter isto noutros formatos (...) eu gostava de alguma forma extrair dados a partir daqui (...)"
More multimedia content	"Podendo enriquecer com uma foto ou vídeo (...) aí já tens toda uma história (...) tens o nível mais esquemático e depois em baixo tens um nível quase de reportagem."
Tutorial animation	"No início, não sei se dá para ter uma 'animaçãozita'"

Table E.4: Codes related to the journalists' suggestions in their original language.

Codes	**Example**
Positive feedback	
Interesting	"Acho a ferramenta muito interessante, parece uma excelente forma de graficamente complementar uma notícia."
Useful	"(...) permite uma leitura detalhada sem perder a perspetiva de mais alto-nível."
Negative feedback	
Not accessible	"A visualização é difícil de interpretar à primeira, limitando a acessibilidade."
No point in moving the events	"Sobre o vídeo, fiquei com dúvidas sobre o objetivo de mover os pontos para cima/baixo."
Not appealing	"A visualização não é visualmente simples e apelativa ao leitor normal neste formato, apenas a quem tem bastante interesse a ir ao âmago do assunto."
Suggestions	
Icons instead of buttons	"Previous event / next event poderiam utilizar um ícone standard, em vez de texto, para se destacarem melhor."
Hyperlinks to more content	"Caso as descrições dos eventos sejam extraídas diretamente de um artigo/notícia faria sentido haver um link para essa páginaparágrafo."
Variable zoom	"(...) talvez fosse interessante ter uma opção de zoom, em que a escala de tempo variava com o nível de zoom (..)"
Automatic scroll	"Era interessante ter uma opção para animar a cronologia."
Timeline	"No futuro gostaria de ver uma linha cronologica que fixasse os eventos no eixo horizontal."
Add photos	"Se, por exemplo, fosse mais à base de fotos, talvez o tornasse mais apelativo."

Table E.5: Written feedback gathered in the survey in its original language.

E.2 Survey

Visualizações no consumo de notícias

Este questionário tem como objetivo perceber de que forma o uso de visualizações poderá ajudar no consumo de notícias. A primeira parte do questionário aborda os hábitos de consumo de notícias. A segunda e última parte relaciona-se com o protótipo desenvolvido no âmbito da minha tese de mestrado, uma ferramenta de visualização que procura realçar elementos narrativos como entidades, eventos, e localizações em textos noticiosos (e de outros domínios). O questionário tem uma duração aproximada de 5 minutos.

Hábitos de leitura de notícias online

1. Faixa etária

   Marcar apenas uma oval.

   - ◯ Menos de 18
   - ◯ 18-24
   - ◯ 25-39
   - ◯ 40-59
   - ◯ 60 ou mais

2. Afiliação (sigla da faculdade, p.ex. FEUP, FBAUP)

   _____

3. Ocupação

   Marcar apenas uma oval.

   - ◯ Estudante
   - ◯ Docente
   - ◯ Técnico
   - ◯ Investigador
   - ◯ Outra: _____

4. Com que frequência lê notícias online?

   Marcar apenas uma oval.

   - ◯ Menos de uma vez por semana
   - ◯ Entre 1 a 3 vezes por semana
   - ◯ Mais de 3 vezes por semana
   - ◯ Uma vez por dia
   - ◯ Várias vezes ao dia

5. Quão sobrecarregada/o se sente com a quantidade de notícias a que é exposto/a diariamente?

   Marcar apenas uma oval.

	1	2	3	4	5	
Nada sobrecarregado/a	◯	◯	◯	◯	◯	Muito sobrecarregado/a

6. Com que frequência se depara com elementos visuais (p. ex. gráficos) em notícias online?

   Marcar apenas uma oval.

	1	2	3	4	5	
Nunca	◯	◯	◯	◯	◯	Sempre

7. Quando uma notícia tem elementos visuais (p. ex. gráficos), quão úteis são para compreensão da notícia?

   Marcar apenas uma oval.

	1	2	3	4	5	
Nada útil	◯	◯	◯	◯	◯	Muito útil

8. Qual das seguintes expressões descreve melhor o seu método de leitura de notícias?

Marcar apenas uma oval.

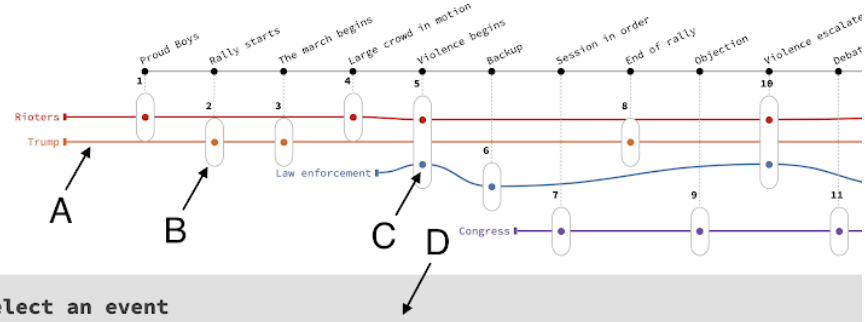
- ◯ Prefiro ter um visão panorâmica das notícias, mesmo que não consiga ler todos os artigos em detalhe
- ◯ Prefiro ler com atenção os artigos com que me deparo, analisando todos os detalhes
- ◯ Não tenho preferência por nenhum dos métodos listados acima

Ferramenta de visualização de notícias

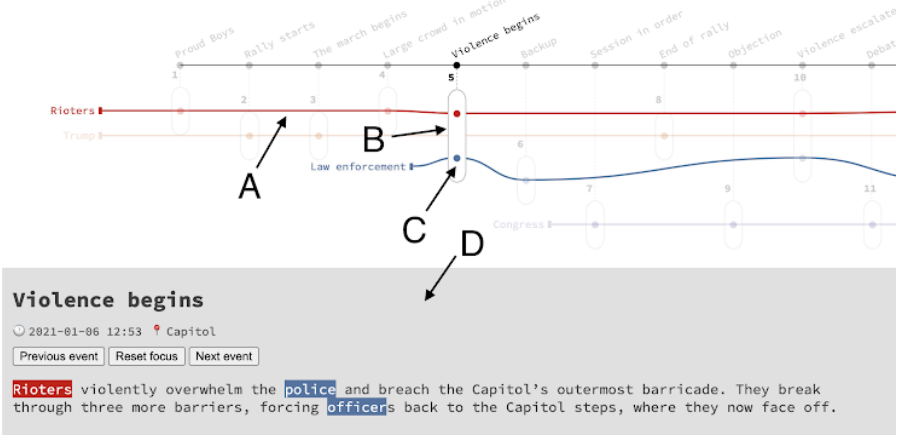
Abaixo encontram-se algumas vistas do protótipo de uma ferramenta de visualização de notícias que procura destacar e contextualizar os eventos de uma história, bem como os seus participantes. Assuma que este tipo de visualizações pode ser gerado automaticamente a partir de um artigo textual.

Legenda:
A - Pessoas/Entidades
B - Evento
C - Participação de uma personagem num evento
D - Descrição do evento em foco (se este existir)

Visualização dos eventos da invasão ao Capitólio



Exemplo de foco num evento



Exemplo de interação com a visualização

http://youtube.com/watch?v=NYyc7CgEr4I

9. Avalie a importância que atribui aos seguintes elementos da visualização

Marcar apenas uma oval por linha.

	1 (Nada importante)	2	3	4	5 (Extremamente importante)
Pessoas/Entidades	◯	◯	◯	◯	◯
Eventos	◯	◯	◯	◯	◯
Participações em eventos	◯	◯	◯	◯	◯
Datas	◯	◯	◯	◯	◯
Localizações	◯	◯	◯	◯	◯
Título do evento	◯	◯	◯	◯	◯
Descrição do evento	◯	◯	◯	◯	◯

10. Estaria interessado/a em usar esta ferramenta como complemento a artigos textuais?

Marcar apenas uma oval.

	1	2	3	4	5	
Nada interessado/a	◯	◯	◯	◯	◯	Muito interessado/a

11. Estaria interessado/a em editar a visualização produzida (p.ex. ordenar as personagens de uma forma diferente)?

Marcar apenas uma oval.

	1	2	3	4	5	
Nada interessado/a	◯	◯	◯	◯	◯	Muito interessado/a

12. Sugestões/comentários à visualização

Obrigada pela sua participação!

Caso tenha alguma dúvida, não hesite em contactar-me através do seguinte e-mail: up201604414@edu.fe.up.pt

References

[1] James Allan. *Introduction to Topic Detection and Tracking*, pages 1–16. Springer US, Boston, MA, 2002. ISBN 978-1-4615-0933-2. doi: 10.1007/978-1-4615-0933-2_1. URL https://doi.org/10.1007/978-1-4615-0933-2_1.

[2] Mieke Bal and Christine Van Boheemen. *Narratology: Introduction to the theory of narrative*. University of Toronto Press, 2009.

[3] BBC. Storyline Ontology. Available at https://www.bbc.co.uk/ontologies/storyline, 2021. [Online; accessed Feb 2021].

[4] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022, 2003. URL http://jmlr.org/papers/v3/blei03a.html.

[5] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, oct 2008. doi: 10.1088/1742-5468/2008/10/p10008. URL https://doi.org/10.1088/1742-5468/2008/10/p10008.

[6] Charles Blundell, Yee Whye Teh, and Katherine A Heller. Discovering non-binary hierarchical structures with bayesian rose trees. *Mixtures: Estimation and Applications. New York, NY, USA: Wiley*, pages 161–187, 2011.

[7] Pablo J Boczkowski. *News at work: Imitation in an age of information abundance*. University of Chicago Press, 2010.

[8] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, 2006. doi: 10.1191/1478088706qp063oa. URL https://www.tandfonline.com/doi/abs/10.1191/1478088706qp063oa.

[9] Alberto Cairo, Schema Design, Google Trends, and Axios. The Lifespan of News Stories. Available at https://www.newslifespan.com/, 2019. [Online; accessed June 2021].

[10] Nan Cao and Weiwei Cui. *Introduction to Text Visualization*, volume 1 of *Atlantis Briefs in Artificial Intelligence*. Atlantis Press, 2016. ISBN 978-94-6239-185-7. doi: 10.2991/978-94-6239-186-4. URL https://doi.org/10.2991/978-94-6239-186-4.

[11] Stuart K. Card, Jock D. Mackinlay, and Ben Shneiderman. *Readings in information visualization - using vision to think*. Academic Press, 1999. ISBN 978-1-55860-533-6.

[12] Mariana Costa. News Story Viz (version 1.0.0) [source-code]. Available at https://github.com/marianafcosta/news-story-viz, 2021. [Online; accessed June 2021].

[13] Weiwei Cui, Shixia Liu, Li Tan, Conglei Shi, Yangqiu Song, Zekai Gao, Huamin Qu, and Xin Tong. Textflow: Towards better understanding of evolving topics in text. *IEEE Trans. Vis. Comput. Graph.*, 17(12):2412–2421, 2011. doi: 10.1109/TVCG.2011.239. URL https://doi.org/10.1109/TVCG.2011.239.

[14] Alan Dix, Alan John Dix, Janet Finlay, Gregory D Abowd, and Russell Beale. *Human-computer interaction*. Pearson Education, 2003.

[15] Wenwen Dou, Xiaoyu Wang, Remco Chang, and William Ribarsky. Paralleltopics: A probabilistic approach to exploring document collections. In *6th IEEE Conference on Visual Analytics Science and Technology, IEEE VAST 2011, Providence, RI, USA, October 23-28, 2011*, pages 231–240. IEEE Computer Society, 2011. doi: 10.1109/VAST.2011.6102461. URL https://doi.org/10.1109/VAST.2011.6102461.

[16] Wenwen Dou, Xiaoyu Wang, Drew Skau, William Ribarsky, and Michelle X. Zhou. Leadline: Interactive visual analysis of text data through event identification and exploration. In *7th IEEE Conference on Visual Analytics Science and Technology, IEEE VAST 2012, Seattle, WA, USA, October 14-19, 2012*, pages 93–102. IEEE Computer Society, 2012. doi: 10.1109/VAST.2012.6400485. URL https://doi.org/10.1109/VAST.2012.6400485.

[17] Wenwen Dou, Li Yu, Xiaoyu Wang, Zhiqiang Ma, and William Ribarsky. Hierarchicaltopics: Visually exploring large text collections using topic hierarchies. *IEEE Trans. Vis. Comput. Graph.*, 19(12):2002–2011, 2013. doi: 10.1109/TVCG.2013.162. URL https://doi.org/10.1109/TVCG.2013.162.

[18] Graham Dove and Sara Jones. Narrative visualization: Sharing insights into complex data. In *Interfaces and Human Computer Interaction (IHCI 2012)*, 2012. URL https://openaccess.city.ac.uk/id/eprint/1134/.

[19] Simon Elvery. D3 Layout Narrative (version 1.0.0) [source-code]. Available at https://github.com/abcnews/d3-layout-narrative, 2016. [Online; accessed Feb 2021].

[20] Ana Figueiras. How to tell stories using visualization. In Ebad Banissi, Mark W. McK. Bannatyne, Francis T. Marchese, Muhammad Sarfraz, Anna Ursyn, Gilles Venturini, Theodor G. Wyeld, Urska Cvek, Marjan Trutschl, Georges G. Grinstein, Vladimir Geroimenko, Sarah Kenderdine, and Fatma Bouali, editors, *18th International Conference on Information Visualisation, IV 2014, Paris, France, July 16-18, 2014*, page 18. IEEE Computer Society, 2014. doi: 10.1109/IV.2014.78. URL https://doi.org/10.1109/IV.2014.78.

[21] Danyel Fisher, Aaron Hoff, George G. Robertson, and Matthew Hurst. Narratives: A visualization to track narrative events as they develop. In *3rd IEEE Symposium on Visual Analytics Science and Technology, IEEE VAST 2008, Columbus, OH, USA, October 19-24, 2008*, pages 115–122. IEEE Computer Society, 2008. doi: 10.1109/VAST.2008.4677364. URL https://doi.org/10.1109/VAST.2008.4677364.

[22] Katie Franklin, Simon Elvery, and Ben Spraggon. Star Wars: every scene from I-VI charted. Available at https://www.abc.net.au/news/2015-12-16/star-wars-every-scene/7013826?nw=0, 2016. [Online; accessed Feb 2021].

[23] Nahum D. Gershon and Ward Page. What storytelling can do for information visualization. *Commun. ACM*, 44(8):31–37, 2001. doi: 10.1145/381641.381653. URL https://doi.org/10.1145/381641.381653.

[24] Bart Geurts, David I. Beaver, and Emar Maier. Discourse Representation Theory. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2020 edition, 2020.

[25] Mohammad Ghoniem, Dongning Luo, Jing Yang, and William Ribarsky. Newslab: Exploratory broadcast news video analysis. In *2nd IEEE Symposium on Visual Analytics Science and Technology, IEEE VAST 2007, Sacramento, CA,USA, October 30 - November 1, 2007*, pages 123–130. IEEE Computer Society, 2007. doi: 10.1109/VAST.2007.4389005. URL https://doi.org/10.1109/VAST.2007.4389005.

[26] Marko Grobelnik and Dunja Mladenic. Visualization of news articles. *Informatica (Slovenia)*, 28(4):375–380, 2004.

[27] Susan Havre, Elizabeth G. Hetzler, Paul Whitney, and Lucy T. Nowell. Themeriver: Visualizing thematic changes in large document collections. *IEEE Trans. Vis. Comput. Graph.*, 8 (1):9–20, 2002. doi: 10.1109/2945.981848. URL https://doi.org/10.1109/2945.981848.

[28] Elizabeth G. Hetzler and Alan Turner. Analysis experiences using information visualization. *IEEE Computer Graphics and Applications*, 24(5):22–26, 2004. doi: 10.1109/MCG.2004.22. URL https://doi.org/10.1109/MCG.2004.22.

[29] Jessica Hullman and Nicholas Diakopoulos. Visualization rhetoric: Framing effects in narrative visualization. *IEEE Trans. Vis. Comput. Graph.*, 17(12):2231–2240, 2011. doi: 10.1109/TVCG.2011.255. URL https://doi.org/10.1109/TVCG.2011.255.

[30] Nancy Iskander, Matthew Thorne, and Craig Kaplan. Comic Book Narrative Charts. Available at https://csclub.uwaterloo.ca/~n2iskand/?page_id=13, 2021. [Online; accessed Feb 2021].

[31] Markus John, Steffen Lohmann, Steffen Koch, Michael Wörner, and Thomas Ertl. Visual analytics for narrative text - visualizing characters and their relationships as extracted from novels. In *Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 2: IVAPP, (VISIGRAPP 2016)*, pages 27–38. INSTICC, SciTePress, 2016. ISBN 978-989-758-175-5. doi: 10.5220/0005669800270038.

[32] Nam Wook Kim, Benjamin Bach, Hyejin Im, Sasha Schriber, Markus H. Gross, and Hanspeter Pfister. Visualizing nonlinear narratives with story curves. *IEEE Trans. Vis. Comput. Graph.*, 24(1):595–604, 2018. doi: 10.1109/TVCG.2017.2744118. URL https://doi.org/10.1109/TVCG.2017.2744118.

[33] Miloš Krstajić, Enrico Bertini, Florian Mansmann, and Daniel A. Keim. Visual analysis of news streams with article threads. In *Proceedings of the First International Workshop on Novel Data Stream Pattern Mining Techniques*, StreamKDD '10, page 39–46, New York, NY, USA, 2010. Association for Computing Machinery. ISBN 9781450302265. doi: 10.1145/1833280.1833286. URL https://doi.org/10.1145/1833280.1833286.

[34] Miloš Krstajić, Enrico Bertini, and Daniel A. Keim. Cloudlines: Compact display of event episodes in multiple time-series. *IEEE Trans. Vis. Comput. Graph.*, 17(12):2432–2439, 2011. doi: 10.1109/TVCG.2011.179. URL https://doi.org/10.1109/TVCG.2011.179.

[35] Miloš Krstajić, Mohammad Najm-Araghi, Florian Mansmann, and Daniel A. Keim. Story tracker: Incremental visual text analytics of news story development. *Inf. Vis.*, 12(3-4): 308–323, 2013. doi: 10.1177/1473871613493996. URL https://doi.org/10.1177/1473871613493996.

[36] Kostiantyn Kucher and Andreas Kerren. Text visualization techniques: Taxonomy, visual survey, and community insights. In Shixia Liu, Gerik Scheuermann, and Shigeo Takahashi, editors, *2015 IEEE Pacific Visualization Symposium, PacificVis 2015, Hangzhou, China, April 14-17, 2015*, pages 117–121. IEEE Computer Society, 2015. doi: 10.1109/PACIFICVIS.2015.7156366. URL https://doi.org/10.1109/PACIFICVIS.2015.7156366.

[37] Philippe Laban and Marti A. Hearst. newslens: building and visualizing long-ranging news stories. In Tommaso Caselli, Ben Miller, Marieke van Erp, Piek Vossen, Martha Palmer, Eduard H. Hovy, Teruko Mitamura, and David Caswell, editors, *Proceedings of the Events and Stories in the News Workshop@ACL 2017, Vancouver, Canada, August 4, 2017*, pages 1–9. Association for Computational Linguistics, 2017. doi: 10.18653/v1/w17-2701. URL https://doi.org/10.18653/v1/w17-2701.

[38] BBC News Labs. Graphical Storytelling. Available at https://bbcnewslabs.co.uk/projects/graphical-storytelling/, 2021. [Online; accessed Feb 2021].

[39] BBC News Labs. Structured Journalism. Available at https://bbcnewslabs.co.uk/projects/structured-journalism, 2021. [Online; accessed Feb 2021].

[40] Page Laubheimer. Beyond the NPS: Measuring Perceived Usability with the SUS, NASA-TLX, and the Single Ease Question After Tasks and Usability Tests. Available at https://www.nngroup.com/articles/measuring-perceived-usability/, 2018. [Online; accessed June 2021].

[41] Lauren Leatherby and Anjall Singhvi. How Trump's Calls to G.O.P. Lawmakers Fit in the Timeline of the Capitol Riot. Available at https://www.nytimes.com/interactive/2021/01/15/us/trump-capitol-riot-timeline.html, 2021. [Online; accessed June 2021].

[42] Shixia Liu, Michelle X. Zhou, Shimei Pan, Weihong Qian, Weijia Cai, and Xiaoxiao Lian. Interactive, topic-based visual text summarization and analysis. In David Wai-Lok Cheung, Il-Yeol Song, Wesley W. Chu, Xiaohua Hu, and Jimmy J. Lin, editors, *Proceedings of the 18th ACM Conference on Information and Knowledge Management, CIKM 2009, Hong Kong, China, November 2-6, 2009*, pages 543–552. ACM, 2009. doi: 10.1145/1645953.1646023. URL https://doi.org/10.1145/1645953.1646023.

[43] Shixia Liu, Yingcai Wu, Enxun Wei, Mengchen Liu, and Yang Liu. Storyflow: Tracking the evolution of stories. *IEEE Trans. Vis. Comput. Graph.*, 19(12):2436–2445, 2013. doi: 10.1109/TVCG.2013.196. URL https://doi.org/10.1109/TVCG.2013.196.

[44] Dongning Luo, Jing Yang, Miloš Krstajić, William Ribarsky, and Daniel A. Keim. Eventriver: Visually exploring text collections with temporal references. *IEEE Trans. Vis. Comput. Graph.*, 18(1):93–105, 2012. doi: 10.1109/TVCG.2010.225. URL https://doi.org/10.1109/TVCG.2010.225.

[45] Kate Moran. Collecting Metrics During Qualitative Studies. Available at https://www.nngroup.com/articles/metrics-qualitative/, 2021. [Online; accessed June 2021].

[46] Randall Munroe. xkcd: Movie Narrative Charts. Available at https://xkcd.com/657/, 2021. [Online; accessed Feb 2021].

[47] Michael Ogawa and Kwan-Liu Ma. Software evolution storylines. In Alexandru Telea, Carsten Görg, and Steven P. Reiss, editors, *Proceedings of the ACM 2010 Symposium on Software Visualization, Salt Lake City, UT, USA, October 25-26, 2010*, pages 35–42. ACM, 2010. doi: 10.1145/1879211.1879219. URL https://doi.org/10.1145/1879211.1879219.

[48] Kalpesh Padia. *Storyline Visualization Techniques for Linear, Non-linear, and Diegetic Narratives*. PhD thesis, North Carolina State University, Releigh, North Carolina, USA, 2019.

[49] Maria Rosala. The Discovery Phase in UX Projects. Available at https://www.nngroup.com/articles/discovery-phase/, 2020. [Online; accessed Feb 2021].

[50] Edward Segel and Jeffrey Heer. Narrative visualization: Telling stories with data. *IEEE Trans. Vis. Comput. Graph.*, 16(6):1139–1148, 2010. doi: 10.1109/TVCG.2010.179. URL https://doi.org/10.1109/TVCG.2010.179.

[51] Dafna Shahaf, Jaewon Yang, Caroline Suen, Jeff Jacobs, Heidi Wang, and Jure Leskovec. Information cartography: creating zoomable, large-scale maps of information. In Inderjit S. Dhillon, Yehuda Koren, Rayid Ghani, Ted E. Senator, Paul Bradley, Rajesh Parekh, Jingrui He, Robert L. Grossman, and Ramasamy Uthurusamy, editors, *The 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2013, Chicago, IL, USA, August 11-14, 2013*, pages 1097–1105. ACM, 2013. doi: 10.1145/2487575.2487690. URL https://doi.org/10.1145/2487575.2487690.

[52] Yang Shi, Chris Bryan, Sridatt Bhamidipati, Ying Zhao, Yaoxue Zhang, and Kwan-Liu Ma. Meetingvis: Visual narratives to assist in recalling meeting context and content. *IEEE Trans. Vis. Comput. Graph.*, 24(6):1918–1929, 2018. doi: 10.1109/TVCG.2018.2816203. URL https://doi.org/10.1109/TVCG.2018.2816203.

[53] Ben Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages, Boulder, Colorado, USA, September 3-6, 1996*, pages 336–343. IEEE Computer Society, 1996. doi: 10.1109/VL.1996.545307. URL https://doi.org/10.1109/VL.1996.545307.

[54] Herbert A. Simon. Designing Organizations for an Information-rich World. In *M. Greenberger (Ed.), Computers, communications, and the public interest.*, pages 37–72. Baltimore, MD: The Johns Hopkins Press, 1971.

[55] Robert Spence. *Information Visualization - An Introduction*. Springer, 2014. ISBN 978-3-319-07340-8. doi: 10.1007/978-3-319-07341-5. URL https://doi.org/10.1007/978-3-319-07341-5.

[56] Yuzuru Tanahashi and Kwan-Liu Ma. Design considerations for optimizing storyline visualizations. *IEEE Trans. Vis. Comput. Graph.*, 18(12):2679–2688, 2012. doi: 10.1109/TVCG.2012.212. URL https://doi.org/10.1109/TVCG.2012.212.

[57] Xavier Tannier and Frédéric Vernier. Creation, Visualization and Edition of Timelines for Journalistic Use. In *"Natural Language meets Journalism", workshop of the International Joint Conference on Artificial Intelligence (IJCAI 2016)*, New York, United States, July 2016. URL https://hal.archives-ouvertes.fr/hal-02407152.

[58] Willem Robert van Hage, Véronique Malaisé, Roxane Segers, Laura Hollink, and Guus Schreiber. Design and use of the simple event model (SEM). *J. Web Semant.*, 9(2):128–136, 2011. doi: 10.1016/j.websem.2011.03.003. URL https://doi.org/10.1016/j.websem.2011.03.003.

[59] Maarten A. J. van Meersbergen, Piek Vossen, Janneke M. van der Zwaan, Antske Fokkens, Willem Robert van Hage, Inger Leemans, and Isa Maks. Storyteller: Visual analytics of perspectives on rich text interpretations. In Octavian Popescu and Carlo Strapparava, editors, *Proceedings of the 2017 Workshop: Natural Language Processing meets Journalism, NLPmJ@EMNLP, Copenhagen, Denmark, September 7, 2017*, pages 37–45. Association for Computational Linguistics, 2017. doi: 10.18653/v1/w17-4207. URL https://doi.org/10.18653/v1/w17-4207.

[60] Piek Vossen, Tommaso Caselli, and Yiota Kontzopoulou. Storylines for structuring massive streams of news. In *Proceedings of the First Workshop on Computing News Storylines*, pages 40–49, Beijing, China, July 2015. Association for Computational Linguistics. doi: 10.18653/v1/W15-4507. URL https://www.aclweb.org/anthology/W15-4507.

[61] Franz Wanner, Christian Rohrdantz, Florian Mansmann, Daniela Oelke, and Daniel A. Keim. Visual sentiment analysis of rss news feeds featuring the us presidential election in 2008. In *Visual Interfaces to the Social and the Semantic Web (VISSW 2009), Sanibel Island, Florida, 8th February 2009*, 2009.

[62] Colin Ware. Chapter Nine - Images, Narrative, and Gestures for Explanation. In Colin Ware, editor, *Information Visualization (Third Edition)*, Interactive Technologies, pages 325–343. Morgan Kaufmann, Boston, third edition edition, 2013. ISBN 978-0-12-381464-7. doi: https://doi.org/10.1016/B978-0-12-381464-7.00009-0. URL https://www.sciencedirect.com/science/article/pii/B9780123814647000090.

[63] Ji Soo Yi, Youn ah Kang, John T. Stasko, and Julie A. Jacko. Toward a deeper understanding of the role of interaction in information visualization. *IEEE Trans. Vis. Comput. Graph.*, 13 (6):1224–1231, 2007. doi: 10.1109/TVCG.2007.70515. URL https://doi.org/10.1109/TVCG.2007.70515.