Comprehensive contextual visualization of a news archive for aiding story planning

Ishrat Rahman **Sami**[1], Dr Tony **Russell-Rose**[1] and Prof. Larisa **Soldatova**[1]

[1]*Goldsmiths, University of London, New Cross, London, SE14 6NW*

Abstract

Writing is a complex mental process of generating ideas and organizing the flow of information to convey the intended message to the appropriate audience for educating, enriching or entertaining. Strategic story planning in the pre-writing phase can enrich the quality of writing. The facts of news are encapsulated in five basic questions "Who", "Where", "What", "When" and "Why" which are fundamental for any readers' understanding. Focusing on these 5Ws, this paper demonstrates visualizations designed to provide cognitive guidance for planning editorial news stories that require comprehensive analysis using a news archive. The visualizations are contextual: global (considering the whole archive), relative (considering topic-based news collection) and local (considering single news). Global context visualizations are designed to aid the identification of a historically important or a decaying topic of interest that can be beneficial to review/compare against raising new topics in the current time. On selecting a topic, a relative context Terms Board is produced to aid brainstorming in the pre-writing phase. During reviewing related documents presented via Terms Board, Local Context visualization is presented to aid in recalling strategic terms' emphasis in the selected news.

Keywords

Natural Language Processing, Visualization, Story Planning, News Writing

1. Introduction

The journey of news is as old as human civilization [1]. Newsgathering and writing are the core fundamentals of journalism [2], and news writing falls under the genre of storytelling [3]. The story planning of the news material before writing aids speed, accuracy and influence via controlled information flow [4]. To win readers' attention, a journalist must expose the "Who", "Where", "What", "When" and "Why" of a news story consciously [5]. Missing any of these 5Ws is referred to as "holes" in journalism [5]. Considering the importance of these questions during story planning before writing, the main contributions of this article are interactive contextual visualizations for

- **Global context:** Archive-based time-bound visualization to identify important story topics for editorial recall.
- **Relative context:** Topic collection based Terms Board visualization for brainstorming.

- **Local context:** Single news based story plan visualization to demonstrate important terms' emphasis.

2. Related Work

Writing can be perceived as a complex mental process that involves four stages: planning, drafting, revising and editing [6]. Strategic story planning is the thinking process that leads to better writing [6]. Knowledge visualization can be used to augment cognition and facilitate thinking by helping to build a rich mental model [7]. Various comprehensive topic-based and cluster-based visualization works have been performed for text analysis. For example, Open Knowledge Maps (OKM) is a bubble chart based visual interface for discovering scientific knowledge from PubMed [8]. In another study, Latent Dirichlet Allocation (LDA) based topic modelling was used to identify hotspot topics using bubble charts and chord diagram [9]. Uniform Manifold Approximation and Projection (UMAP) [10] and directed graphs were used to visualize topics filtered by patterns and analyzed by LDA in a study by Ordun et al. [11]. Topic evolution can be represented with WordStream using word cloud and a stream graph [12]. Tools like VISTopic use Sunburst diagrams to represent topics and ThemeRiver timeline to demonstrate topic strength [13]. A study by Bras et al. introduced a theme-based visualisation based on a hierarchical topic model backed by LDA [14]. TextWheel visualization that consists of one or multiple keyword wheels, a document transportation belt, and a dynamic system that connects the wheels and belt is introduced in a study [15]. For news, the journalist ensures the presence of "Who, Where, What, When and Why" answers of the news in the document to address the facts [16, 17]. In this paper, we demonstrate writing-focused comprehensive interactive visualizations for story planning.

3. Contextual visualization

In this paper, we are presenting an interactive contextual visualization tool (as a demo), "Story Analysis" which is built on a collection of 730 news published by "The Pharmaletter" between March 2021 and May 2021. Global context visualizations are designed to aid the identification of a historically important or a decaying topic of interest that can be beneficial to throwback against raising new topics in the current time. It is a monthly visualization of the whole collection. On selecting a topic, a relative context Terms Board is produced to aid brainstorming in the pre-writing phase for intuiting creative thoughts. During reviewing related documents presented via Terms Board, Local Context visualization is presented to aid a better strategic understanding of the most important terms' emphasis in the selected news.

3.1. Global Context Visualization

The interactive Global Context visualization represents "When" (time-bound) based frequency analytics for a sample news archive to guide "Why" (motivation to write) and to identify "What" (most frequent topics), "Who" (most frequent characters/organizations) and "Where" (most frequent locations) for planning a story. Figure 1 shows a monthly Global Context visualization.

The answer to "Why" is contributed via an editor's point of view for achieving a task. This visualization can be useful for the following tasks:

- Identify historically important topic in the given time that can be beneficial to recall in the current time

- Identifying decaying topics that can be beneficial to review/compare against raising new topics

When?

For this demo, we have collected 730 articles published between 2021-03-01 and 2021-05-30 from the Pharmaletter News Publisher.

Currently displaying data from 2021-03-01 to 2021-03-31.

March 2021   April 2021   May 2021

What?

The following section provides the top 25 weighted topics identified in the specified date range. Hovering over a topic will display the number of documents the topic was highlighted per day. This descriptive information is helpful to understand whether the topic is having more / less / consistent importance in the selected date range. Clicking on the topic will lead you to the Terms Board view of the selected topic. Terms Board view provides a quick diagnostics analytics view for discovering ideas related to the selected topic.

Topic: treatment
Blocks count in range: 51
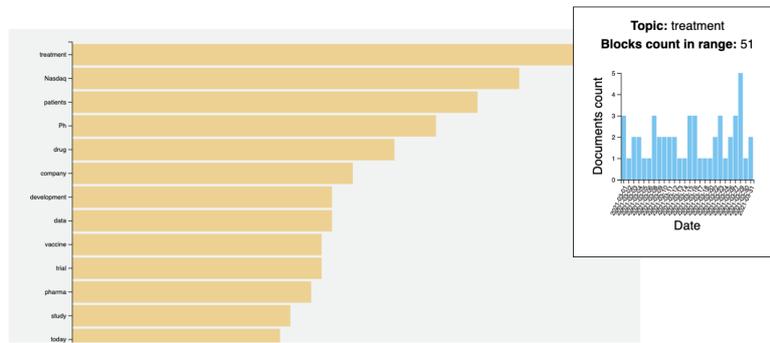
Who?

The following section provides the major characters and organizations that have been referred to in the news stories in the specified date range. Hovering over a topic will display the number of documents the topic was highlighted. This descriptive information is useful for identifying the importance of the character/organization in the corpus. Clicking on the topic will lead you to the Terms Board view of the selected topic. Terms Board view provides a quick diagnostics analytics view for discovering ideas related to the selected topic.
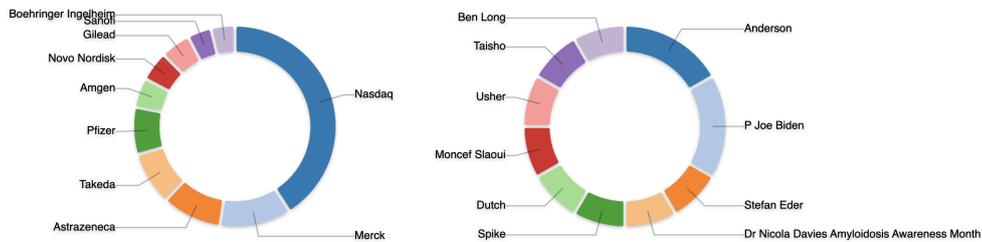
Where?

The following section provides the major countries in the world that have been referred to in the news stories in the specified date range. Hovering over a topic will display the number of documents the topic was highlighted per day. This descriptive information is helpful to understand whether the topic is having more / less / consistent importance in the selected date range. Clicking on the topic will lead you to the Terms Board view of the selected topic. Terms Board view provides a quick diagnostics analytics view for discovering ideas related to the selected topic.

Country: China
Blocks count: 8
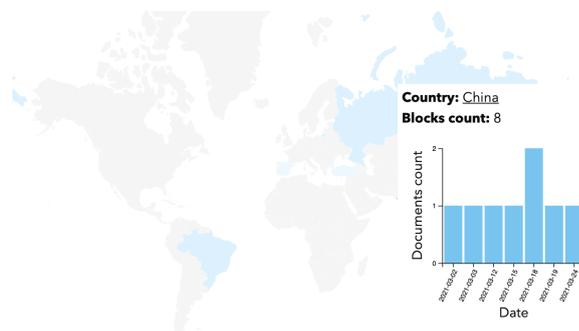
Figure 1: Global Context Visualization (https://storyanalysis.co.uk/demo/index.html).

Figure 2: Relative Context Visualization of a company "Nasdaq" (https://storyanaly-sis.co.uk/demo/termsboard.html?key=nasdaq).

This visualization can work to guide identifying historical important events that require editorial recall through comprehensive writing. Hovering on topics reveals frequency bars of the topic in a timeline. Clicking on the topic produces respective topic-based Relative Context Terms Board visualization to aid story planning on that topic (see the next section). During data analysis, for each news most important terms are selected as topics based on the algorithm proposed by Sami and Farrahi (2017) [18]. Frequency analysis was performed on the extracted topics of all documents. The most important top topics are categorically presented in these visualizations.

3.2. Relative Context Visualization

Relative Context is formed by using the news collection of a specific topic in a considered time period. In our interactive demo, when a user selects a topic of interest for writing, a comprehensive Terms Board visualization of the topic collection is presented as an aid for story planning. Figure 2 demonstrates a sample Terms Board. Brainstorming, clustering, outlining and drafting can improve writing quality in the pre-writing phase[6]. The use of a storyboard improves engagement and expressing ability [19]. It enhances storytelling power, enforces writing discipline and sharpens the narrative [20], reveals gaps in continuity, improves the

78
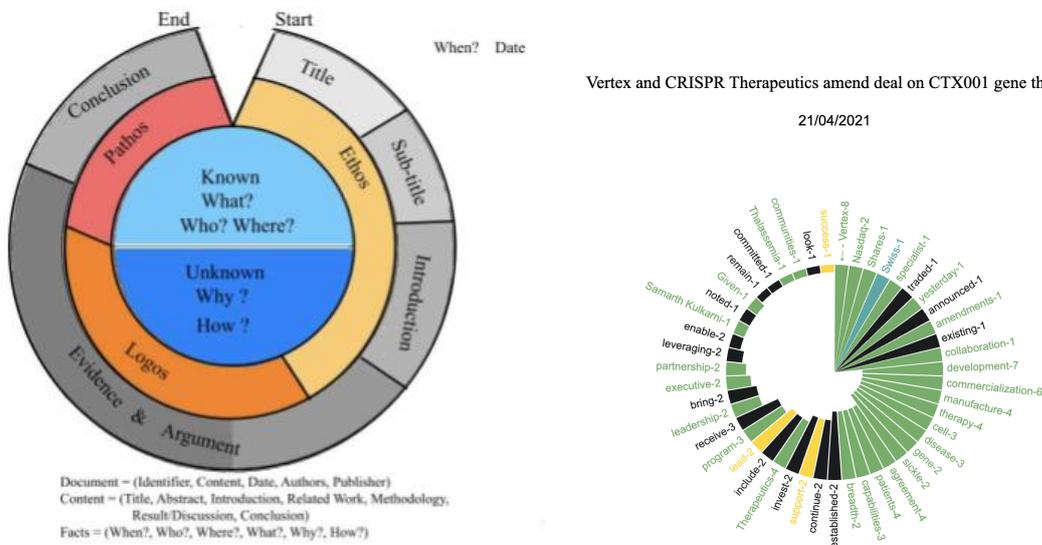
(a) Who/Where/What/Why mapping in content and When from date. (b) Clockwise terms visualization
.

Figure 3: Local Context Visualization of a news (https://storyanalysis.co.uk/demo/lc.html?condition=vizkey=mood-music-improves-for-type-1-diabetes-candidate)

quality of text organization and increases readability [20].

Inspired by the concept of a storyboard, Terms Board displays terms as a palette of thoughts to guide an individual to create a fact-driven story plan. For building a Terms Board, frequency analysis is performed on the terms/topics of the news collection and the most frequent top terms of each question category are selected for display. The terms are categorized into six story planning aspects: Who (person/organization), Where (location), What (related topics represented by nouns), What (actions represented by verbs), Why (positive sentiment words) and Why (negative sentiment words). Each story planning aspect forms a card in Terms Board. Each card groups topics based on historical importance based on a timeline. A card is further separated into three timeline-based aspects represented by circles based on the cumulative historical weight in a timeline. If a term appears early in the given date range, it gets a higher weight. The dark grey circle represents historically weighted terms, the light grey circle represents recently important terms and the turquoise colour circle represents consistent terms that are important both historically and recently. Terms Board is an interactive visualization. Clicking on the circles reveals related news.

3.3. Local Context Visualization

To gain readers' attention from the beginning to end, the news agencies generally adopt an inverted pyramid structure where a story starts with stating the most important material [5]. Likewise, "Aristotle's Rhetoric" has guided writers to create effective communication using "Ethos", "Logos" and "Pathos" [21, 22]. Ethos is the art of establishing authority on a topic,

logos is building logical argumentation and pathos is stating an opinion. For influential writing, authors use various structural story planning templates, e.g. Joseph Campbell's "The Hero's Journey" [23]. For news, the journalist ensures the presence of "Who, Where, What, When and Why" answers in the chronology of the news story [16, 17, 5]. In our demo, we experimentally mapped "Aristotle's Rhetoric" and inverted pyramids structure chronologically in a clockwise manner into the format of "The Hero's Journey" for news as shown in Figure 3(a). To establish "Ethos", news must answer "What", "Where", "Who" and "When" related to the story early in the document. For establishing "Logos", news must provide evidence for "Why" and "How". For enriching/educating the reader, an editorial conclusion attempts to establish "Pathos" at the end. In order to evaluate our assumptions, we carried out an experiment based on readers' experience during cognitive reading tasks in November - December 2021. We are still reviewing the results. To identify chronologically important terms of news we used the algorithm proposed by Sami and Farrahi (2017) [18]. We represented the terms in a clockwise manner to preserve relative positions, the bar represents the importance of the terms calculated by the algorithm and the colour represents various types (noun, verb, positive sentiment, negative) of the terms as show in 3(b).

4. Conclusion

"Story Analysis" is a tool to aid story planning for influential writing. The main contributions of this work are writing-focused interactive visualizations. Linking analytics with the interaction of the comprehensive visualization has the potential to guide our understanding of story planning and can lead to better automation of creative writing. Therefore, we are currently working on evaluating this approach via cognitive reading and writing experiment tasks.

Acknowledgments

References

[1] A Brief History of News, 2019. URL: https://schools.firstnews.co.uk/blog/journalistic-writing/a-brief-history-of-news/.

[2] N. Fenton, News in the digital age, Routledge, 2009.

[3] A. Russell, Networked: A contemporary history of news in transition, Polity, 2011.

[4] M. L. M. L. Spencer, News Writing The Gathering , Handling and Writing of News Stories, Project Gutenberg, 2007.

[5] A. McKane, News writing, SAGE, London, 2006.

[6] Listyani, Promoting academic writing students' skills through "process writing" strategy, Advances in language and literary studies 9 (2018) 173.

[7] M. Wang, M. J. Jacobson, Guest editorial - knowledge visualization for learning and knowledge management, Educational technology society 14 (2011) 1–3.

[8] P. Kraker, C. Kittel, A. Enkhbayar, Open Knowledge Maps: Creating a Visual Interface to the World's Scientific Knowledge Based on Natural Language Processing, 027.7 Zeitschrift für Bibliothekskultur 4 (2016) 98–103. URL: https://0277.ch/index.php/cdrs_0277/article/view/157. doi:10.12685/027.7-4-2-157.

[9] M. Dong, X. Cao, M. Liang, L. Li, H.-Y. Liang, G. Liu, Understand Research Hotspots Surrounding COVID-19 and Other Coronavirus Infections Using Topic Modeling, 2020. doi:10.1101/2020.03.26.20044164.

[10] L. McInnes, J. Healy, J. Melville, Umap: uniform manifold approximation and projection for dimension reduction (2020).

[11] C. Ordun, S. Purushotham, E. Raff, Exploratory Analysis of Covid-19 Tweets using Topic Modeling, UMAP, and DiGraphs, arXiv:2005.03082 [cs] (2020). URL: http://arxiv.org/abs/2005.03082, arXiv: 2005.03082.

[12] T. Dang, H. N. Nguyen, V. Pham, J. Johansson, F. Sadlo, G. Marai, Wordstream: Interactive visualization for topic evolution, in: EuroVis, 2019.

[13] Y. Yang, Q. Yao, H. Qu, Vistopic: A visual analytics system for making sense of large document collections using hierarchical topic modeling, Visual Informatics 1 (2017) 40–47.

[14] P. L. Bras, A. Gharavi, D. A. Robb, A. F. Vidal, S. Padilla, M. J. Chantler, Visualising COVID-19 Research, arXiv:2005.06380 [cs] (2020). URL: http://arxiv.org/abs/2005.06380, arXiv: 2005.06380.

[15] W. Cui, H. Qu, H. Zhou, W. Zhang, S. Skiena, Watch the story unfold with textwheel: Visualization of large-scale news streams, ACM transactions on intelligent systems and technology 3 (2012) 1–17.

[16] How to write a news article, BBC Bitesize (2021). URL: https://www.bbc.co.uk/bitesize/topics/zgqxwnb/articles/zbsbwty.

[17] Fact, opinion and report writing, 2021. URL: https://www.bbc.co.uk/bitesize/articles/zkvg47h.

[18] I. R. Sami, K. Farrahi, A simplified topological representation of text for local and global context, in: Proceedings of the 25th ACM international conference on Multimedia, 2017, pp. 1451–1456.

[19] M. Janah, IMPROVING STUDENTS' WRITING ABILITY THROUGH STORYBOARD, 2017. URL: https://doaj.org, iSSN: 2356-2048, 2356-203X Issue: 1 Publisher: Universitas Muhammadiyah Pringsewu Volume: 3.

[20] S. L. Harrington, An Author's storyboard technique as a prewriting strategy, The Reading Teacher 48 (1994) 283. URL: http://search.proquest.com/docview/203270492/?pq-origsite=primo, place: Newark Publisher: Blackwell Publishing Ltd.

[21] C. Rapp, Aristotle's Rhetoric, in: E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy, spring 2010 ed., Metaphysics Research Lab, Stanford University, 2010. URL: https://plato.stanford.edu/archives/spr2010/entries/aristotle-rhetoric/.

[22] M. Meyer, Aristotle's Rhetoric, Topoi 31 (2012) 249–252. doi:10.1007/s11245-012-9132-0, place: Dordrecht Publisher: Springer Nature BV.

[23] Y. Cao, R. Klamma, M. Jarke, The Hero's Journey - Template-Based Storytelling for Ubiquitous Multimedia Management, Journal of Multimedia 6 (2011) 156–169. doi:10.4304/jmm.6.2.156-169.