Comprehensive Terms Board Visualization for News Analysis and Editorial 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

Knowledge providers, such as authors, teachers, researchers and journalists rely on researching facts to convey evidence-driven information about a selected topic and story planning in the pre-writing phase enhances engagement and understanding of the audience through a better content organization. Typical search engines support finding relevant facts, but they do not aid an individual's metacognition process of a topic. In this demo, we introduce the concept of the Terms Board, a topic-driven comprehensive visualization for presenting terms to provide a cognitive guide for news analysis and formulating their plans for storytelling in editorial writing. Terms Board is composed of six cards reflecting the major storytelling aspects: what the story is about, who are the characters of the story, where the story is located, why there are challenges, what has been done to address the challenges and why the actions were effective. Each card shows three top terms based on three factual timeline aspects: historical, consistent and latest. For this demo, we extracted emphasised terms from a collection of documents in a news archive and produced a Terms Board for the most frequent topics which were then presented to a group of study participants. Participants' performances on several tasks have been measured and analysed. The study results are encouraging. The major contribution of this research is presenting a Terms Board visualisation approach as a cognitive guide for news analysis and editorial story planning and presenting an experimental evaluation of this approach via cognitive reading and writing user experiment tasks.

Keywords

Natural Language Processing, Visualization, Story Planning, News Writing

1. Introduction

Writing is a communication stream through which authors enrich, entertain and educate the audience about a specific topic. For ascertaining facts in the pre-writing researching phase, various search engines provide relevant results with titles, descriptions, links and tags along with a range of statistical charts and snippets. But these results do not comprehend underlying resources to aid individuals' metacognition. Metacognition (individual's reflections about their own knowledge [1]) can be improved by cognitive control which can be introduced by purposeful goal-oriented behaviour and decision-making process [2]. The "Terms Board" (TB) demo presented in this paper accommodates cognitive control from two perspectives: story planning and timeline. While story planning templates like Joseph Campbell's "The Hero's Journey" guide the authors in organizing their plans [3] for better writing, the timeline

relativity of the terms plays a major role in understanding their factual relativity. As the use of a storyboard improves individual/collective engagement and expressing ability [4], TB visualizes the terms as a board. Based on the "The Hero's Journey" template, TB displays six cards to reflect story planning aspects: what the story is about, who are the characters of the story, where the story is located, why there are challenges, what has been done to address the challenges and why the actions were effective. Each card contains three packed bubbles of three top terms in that category to reflect the timeline aspect: historical, consistent and latest. By doing this, TB is visualizing a palette of thoughts for metacognition during news analysis and editorial writing. Therefore, the major contributions of this article are: the introduction of a concept of comprehensive Terms Board (TB) visualization as a cognitive guide and reporting an experimental evaluation of TB via cognitive reading and writing user experiment tasks.

2. Related Work

Applications can facilitate information seekers by providing information through search facilities or automatically generated comprehensive categorization/visualization [5]. Global and professional domain-specific search engines play a vital role in knowledge-seeking by providing various advanced search options, categorical filters and statistical information [5]. On the other hand systems like Open Knowledge Maps (OKM) [6][7], WordStream [8] and VISTopic [9] comprehend metadata of the search result categorically using comprehensive visualization. Existing search and comprehensive technologies are focused on knowledge discovery, while TB is focused on analyzing and planning knowledge delivery, stimulating creativity and improving the engagement of the audience by aiding individuals to create a guided story plan.

3. Methodology

TB categorizes terms into the following story planning aspects. Who: role players in the documents are identified by the category "Person" or "Organization". Where: key locations in the documents are identified by the category "Location". They are represented in proper nouns. What (topics): main topics are identified by emphasized nouns and proper nouns. Why (negatives): challenges are identified by a static list for the prototype system. What (actions): verbs that reflect the actions in the documents. Why (positives): words identified by a static list for the prototype system. Each story planning category forms a card in TB. A card is further separated into three timeline-based aspects. Historical: if a term appears early in the given date range, it gets a higher weight. Latest: if a term appears later in the given date range, it gets a higher weight. Consistent: these are the terms that appear consistently both historically and in the latest documents. The weight is the average of historical and latest weight.

For displaying terms in TB, emphasized terms are extracted from a topic-based collection of documents and displayed as shown in Figure 1. Each document is analyzed in the archive using the topological approach presented in the paper "A simplified topological representation of text for local and global context" [10] to identify topics and actions (verbs) from a document. The identified terms are categorized into six story-planning aspects. For an extracted term, the classification type (Person/Organization/Location) of the term is identified by the categorical
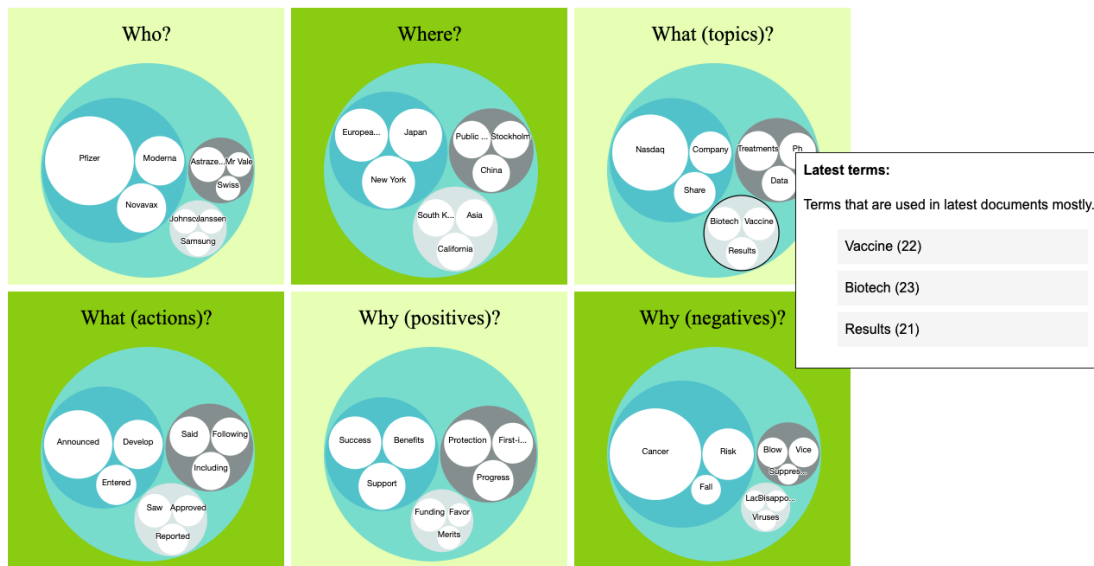
Figure 1: The TB visualization of the topic "Nasdaq". Hovering over a term displays the full term and the number of occurrences in the topic collection is displayed.

information provided by Google Knowledge Graph API [11]. The extracted emphasized term are combined as a list of highly frequent related terms using frequency analysis. The algorithm uses the category of terms to place them in appropriate cards and weights each term on the timeline-based aspects to select the top terms of that aspect. TB visualization is generated using the D3 library's packed bubble chart [12]. On hovering over each timeline bubble, full terms are displayed. Clicking a term opens a relational view modal with a related set of documents (title, meta-description and link to the main document). We applied the scanning approach to display related terms. Scanning helps to locate specific facts rapidly [13]. For a given set of terms, all the sentences in the relative context are scanned based on word co-occurrence, top 5 relations for each given term are displayed using the D3 library's Sankey [13] diagram when a timeline bubble is clicked. On clicking a relation, the set of documents (title, meta-description) that reflects the relation is displayed. When clicking a document, a user navigates to the original document/news.

4. Cognitive reading and writing experiment

We evaluated TB via an online participation-based experiment. We used a pharmaceutical news website to build a collection and produce the visualizations. Each participant from a homogeneous group was given two story writing tasks on pre-decided topics. The selected two topics were Pfizer and AstraZeneca. For both of the tasks, the participants were given

Table 1

Writing Performance.

| Criterion	Scale	List(mean)	TB and List(mean)	P-value	Hypothesis testing with p = 0.05
Quality of story	1(Poor) - 5 (Very good)	3.46875	3.515625	0.4974603636	Can not reject null hypothesis
Completion time	Suggested 20 mins	13.8 min	16.0 min	0.03805485437	Reject null hypothesis
Ease	1(Hard) - 5 (Very easy)	3.21875	2.84375	0.02596133121	Reject null hypothesis

a list of documents (title, introduction and link to full document text) to plan a story and write a minimum of 500 characters about the topic. But for one of the tasks, we presented TB visualization generated based on the documents having the topic before the list. The topics were randomly assigned to the control condition. The order of the tasks was generated using "Latin Square Design" to balance the systematic difference between successive conditions. We invited three academic reviewers to blindly score the writing tasks. They scored the quality of the story on a scale from 1 to 5, where 1 stands for poor and 5 stands for very good. We took the average score given by the reviewers for each task. We performed a paired t-test on the average scores of quality of writing, ease of use and completion time. We recruited 32 participants for this experiment.

4.1. Evaluation

The result is reported in Table 1. As shown in Table 1, we have achieved 95% confidence in the reported result on all criteria apart from "Quality of Story". According to the evaluation, the participants spent more time writing their stories and achieved a better quality of writing when they are assisted with TB visualization compared to the tasks where they were given only a list of documents. They found the task assisted by TB comparatively difficult as they had to process additional information than only scanning the documents.

4.2. Google Analytics Evaluation

Clearly, participants' thoughts can't be directly observed. Therefore, information about metacognition must be collected in indirect ways [1]. We used Google Analytics to record participants' activity while doing the task with their permission. We collected 17 participants' activities. We tracked clicks and hovers on cards, timeline bubbles and documents during reading and selecting documents in the story planning phase. We also tracked how long the selected document was open during the session. Using this information, we generated directed graphs. The red nodes represent the topic of the story (Pfizer or AstraZeneca)which is the root node. The amber nodes represent story planning aspects (whatTopicBox, whatActionBox, whoBox, whereBox, whyNegativeBox, whyPositive Box). The yellow nodes represent the timeline aspects (consistent, old_to_new/historical, new_to_old/latest). The green nodes represent documents. The size of the nodes represents the number of events/time associated with the node during the pre-writing story planning phase. Figure 2 (a) displays the story planning directed graph of a participant while writing a story using only a list of documents. The participant selected 6 documents from the list to plan their story and then gradually increased their focus on 2 documents to base the writing. Figure 2 (b) displays the story planning directed graph of a
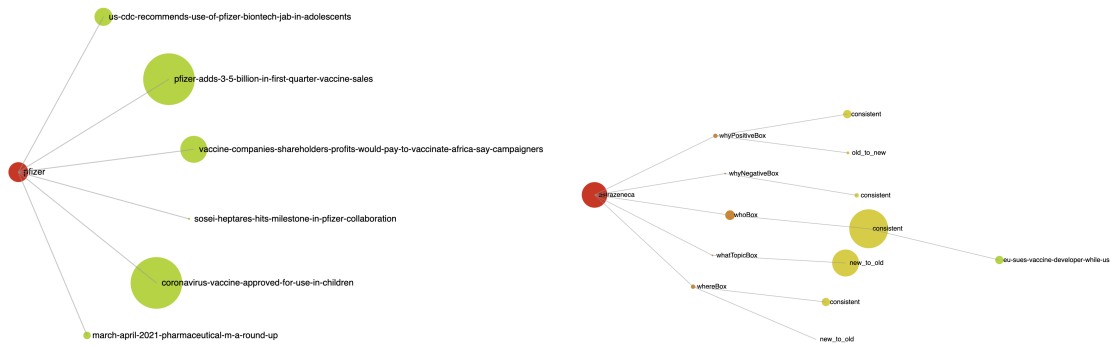
(a) Selection pattern during list of documents task (b)Selection pattern during TB based task

Figure 2: Directed graph representing strategy-making during reading and selecting documents. The bubble size represents the amount of time or event associated with the topic (red)/ reading documents (green).

participant while writing a story using TB. The figure shows that the participant investigated five cards to start with and then selected the Who card's consistent terms to select a document for reading in the planning phase. All collected observations from analytics show that the participants selected fewer documents via TB than the list of documents which refers to the fact that they selected the documents using a strategic planning process. Therefore, combining this observation along with the results reported in Table 1, we can state that for a group of participants, TB contributed to more metacognition (increased time for finishing the task) and better quality of writing by aiding planning before writing a story.

5. Conclusion

TB is designed to comprehend facts and aid in improving an individual's metacognition to create a plan for a designated task based on a strategy. TB has the potential to contribute to news analysis and editorial writing. TB can expose the metacognition process while writing. In this paper, we demonstrated TB using a news corpus and showed that TB contributed to metacognition and quality of writing by aiding planning before writing a story. TB can also be used for other text corpora such as articles and books as an aid for brainstorming in learning and teaching.

References

[1] A. Haukås, C. Bjørke, M. Dypedahl, Metacognition in Language Learning and Teaching, Routledge Studies in Applied Linguistics, 1 ed., Taylor Francis, 2018.

[2] M. S. Gazzaniga, R. B. Ivry, G. R. Mangun, Cognitive neuroscience: the biology of the mind, fourth edition ed., W. W. Norton & Company, Inc, New York, N.Y, 2014.

[3] 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`.

[4] M. Janah, Improving Students' Writing Ability Through Storyboard, 2017. ISSN: 2356-2048, 2356-203X Issue: 1 Publisher: Universitas Muhammadiyah Pringsewu Volume: 3.

[5] T. Russell-Rose, T. Tate, Designing the Search Experience, Elsevier, 2013. doi:`10.1016/C2011-0-07401-X`.

[6] O. K. Maps, Open Knowledge Maps - A visual interface to the world's scientific knowledge, 2021. URL: https://openknowledgemaps.org/index, last accessed on 09/03/23.

[7] 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. doi:`10.12685/027.7-4-2-157`.

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

[9] 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.

[10] 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.

[11] Google Knowledge Graph Search API, 2021. URL: https://developers.google.com/knowledge-graph, last accessed on 09/03/23.

[12] Y. Holtz, Circular Packing | the D3 Graph Gallery, 2018. URL: https://www.d3-graph-gallery.com/circularpacking, last accessed on 09/03/23.

[13] Skimming and Scanning - TIP Sheet - Butte College, 2021. URL: http://www.butte.edu/departments/cas/tipsheets/readingstrategies/skimming_scanning.html, last accessed on 09/03/23.