

Triangulating the News: Visualizing Commonality and Variation Across Many News Stories on the Same Event

Elena L. Glassman, Janet Sung, Katherine Qian, Yuri Vishnevsky, Amy Zhang

¹Harvard University, ²University of Washington

Cambridge, MA, USA

glassman@seas.harvard.edu, janetysung@gmail.com, qiank@college.harvard.edu, yurivish@gmail.com, axz@cs.uw.edu

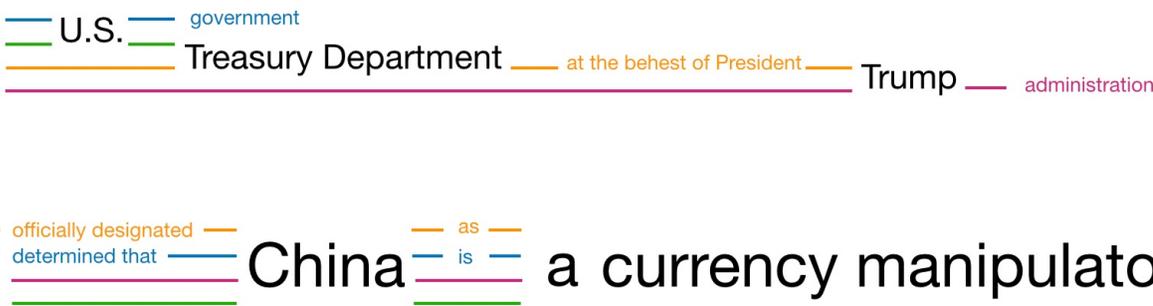
The  U.S. government
Treasury Department at the behest of President Trump administration
has officially designated determined that China is a currency manipulator
labeled

Figure 1: Prototype visualization of 4 lede sentences about a single event, manually aligned using a early version of our lede-alignment algorithm.

ABSTRACT

In today’s news environment, there are many articles written about the same event. Readers may intentionally diversify their news sources to get the full picture or *triangulate* the underlying truth, but they can only realistically read a very small number of articles. In the current reading environment, readers must read through many redundant passages from different articles to understand where they overlap and differ to get a comprehensive understanding of the event.

The differences between these articles can be explained by one or more of the following factors:

- information accessible to the article author
- perceived accuracy of the accessible information
- perceived relevance to the intended article audience
- perceived relationship to the audience
- stylistic expectations about formalism, emotionally charged words, etc.
- semantics-preserving or semantics-bending word choices
- the impartiality or open bias of the author

Given that small sample size of articles the typical reader has the patience to read, and the cognitive load of comparing across even a small number of articles, it is possible for readers to miss key details or presume unanimity of interpretation among those familiar with the situation. It is the commonalities and variations across these articles that we wish to computationally expose for readers’ inspection, to augment their ability to make their own judgements about the event being discussed.

In this on-going work, we are focusing on lede sentences that summarize the rest of the article and iterating on a novel multi-file soft differencing algorithm to merge lede passages from multiple (and eventually many) articles covering a single event, along with a novel visualization tool that renders these commonalities and differences in a way that lets the human eye quickly read the backbone as well as the additions, omissions, and variation in word choice of individual authors, highlighting where different articles use the same or similar language and where they diverge. We hope that this technique will eventually scale to many more articles than an individual would even consider reading on a common event unassisted.

In developing the algorithm, we are following the following axioms:

- employing pre-existing NLP techniques such as entity recognition and stemming
- preserving the readability of the original ledes within the visualization, i.e., as lossless a visualization as possible
- no summarization except for collapsing words within a user-controlled threshold of similarity.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Woodstock '18, June 03–05, 2018, Woodstock, NY

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-9999-9/18/06...\$15.00

<https://doi.org/10.1145/1122445.1122456>

An initial version is implemented in an ObservableHQ notebook where sets of lede sentences can be input. The output is a graph data structure capturing an alignment of the sentences for visualization similar to what was manually created as a gold standard for the non-interactive algorithm shown in Figure 1. We will present examples of sets of article ledes merged using our current version of the algorithm, allow attendees to play with sets of ledes input as well as any thresholds for collapsing degrees or types of variation, and discuss how this tool could be used by readers and journalists to interrogate coverage, emphasis, and bias across articles.

KEYWORDS

news, text visualization, variation visualization

ACM Reference Format:

Elena L. Glassman, Janet Sung, Katherine Qian, Yuri Vishnevsky, Amy Zhang. 2018. Triangulating the News: Visualizing Commonality and Variation Across Many News Stories on the Same Event. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/1122445.1122456>