Online Study of Word-Sized Visualizations in Social Media

F. Huth[1], M. Awad-Mohammed[1], J. Knittel[1], T. Blascheck[1] and P. Isenberg[2]
[1]University of Stuttgart, Germany
[2]Université Paris Saclay, CNRS, Inria, France

Abstract

We report on an online study that compares three different representations to show topic diversity in social media threads: a word-sized visualization, a background color, and a text representation. Our results do not provide significant evidence that people gain knowledge about topic diversity with word-sized visualizations faster than with the other two conditions. Further, participants who were shown word-sized visualizations performed tasks with equally few or only slightly fewer errors.

CCS Concepts

• ***Human-centered computing*** → ***Empirical studies in visualization; Visualization design and evaluation methods;***

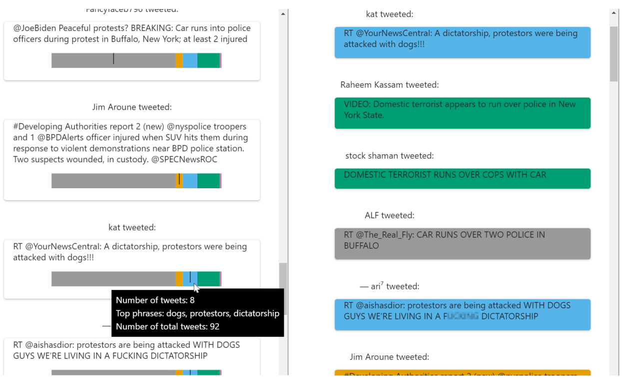
1. Introduction

On Twitter, single tweets can be shared and replied to with or without a comment. This often leads to long threads in which the meaning and content of the original tweet might be altered or completely changed. As a consequence, it can be difficult for people to understand the context or meaning of a single tweet or reply without additional information about the topic that is currently being discussed. While there are several approaches that visually analyze such text data [DL16, JFCS17, KK15], we use word-sized visualizations [Tuf06] since they are well suited to convey information about changing topics in short social media messages. They are small visualizations the size of a word or a paragraph, can be placed and designed flexibly [GWFI14, GBWI17, LB18], and have been used in social media before [HBK*21]. Here, we designed a stacked bar chart  as a word-scale visualization that represents the different topics discussed in a single Twitter thread about a common shared item.

2. User Study Design and Setup

To find out whether word-sized visualizations help to give an overview of topic diversity in a Twitter thread, we conducted an online study on Prolific [pro]. We hoped to find evidence for a clear advantage of word-sized visualizations and chose a quantitative study. As sample data we collected 92 unique tweets [twi] sharing a controversial video of a car driving into a group of police officers in Buffalo, NY, USA, and clustered those tweets with k-means on their tf-idf features [Llo82, SMR08], a suitable k being determined with the Davies-Bouldin index [DB79]. We wanted to find out whether people could answer questions about the topic clusters in the Twitter thread faster and more accurately with embedded word-sized visualizations than with two other encodings: a color encoding in the background of individual tweets representing the topic discussed in the tweet and a text label in the bottom right corner of the tweet showing the name of the topic.

Setup. We showed participants a re-implemented Twitter thread in which we encoded the topic of each tweet in one of the three conditions. In the first condition (*WSV*), we added a word-sized visualization to each tweet, which at its core is a stacked bar chart in which each area represents the number of tweets per topic in the thread. A small tick mark indicates the topic of the current tweet. On mouse-over, participants get information about the number of tweets, the three most salient words of this topic, and the total number of tweets in the thread (see Figure 1 (a)). The second condition (*color*) uses a background color for each tweet according to its topic as seen in Figure 1 (b). In the third condition (*text*), a text label in the bottom right corner of each tweet states its topic number such



(a) *WSV* (b) *Color*

Figure 1: *Example threads for tested conditions, (a) WSV with a word-sized visualization and mouse-over interaction, (b) color with colored background.*

DOI: 10.2312/evp.20211069

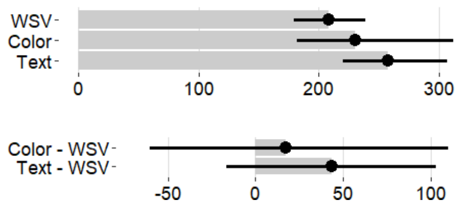


Figure 2: *Answering times in seconds per condition summarized for questions $Q1$ to $Q4$ on the real Twitter data and pairwise differences of conditions color and text compared to WSV with 95% confidence intervals.*

as "topic 3", for example. As such, the *WSV* condition summarizes data about the topics in the thread in one graphic, while the other conditions required scrolling to gauge the same information.

Tasks. We asked participants to answer six questions: how many different topics are in the Twitter thread ($Q1$), what percentage of tweets belongs to the largest topic cluster ($Q2$), which of the topic clusters are most different in content to each other ($Q3$) and, finally, a multiple choice question to decide which keyword describes one of the topics best ($Q4$). After each question, participants were asked to rate the difficulty of the task. In a final set of questions, we inquired how confident participants were in their answers ($Q5$) and how well they liked the presentation of the topics ($Q6$) – see OSF.

Study design. We used the same Twitter data for all three conditions and, therefore, used a between-subject study design. After their informed consent, participants were assigned to a condition randomly. On Prolific it is difficult to recruit the same number of participants for each condition, we recruited 19, 25 and 18 participants for the *WSV*, *color* and *text* condition, respectively, after excluding 4 participants for failing the training questions and finishing the study unreasonably fast. 50% of the participants used Twitter at least weekly and 77% were familiar or very familiar with visualizations in general or bar charts. It took participants on average 9 minutes to complete the study. Participants were given an introduction into their condition and the questions. They were shown an easy training data example, a thread with six tweets, and asked the same questions as for the real study, except for question $Q3$.

3. Study Results

We evaluated the time it took participants to answer the questions and the correctness of the answers. We plot the average answering times as well as effect sizes for each condition with 95% confidence intervals in Figure 2. With the help of word-sized visualizations, participants could answer questions about topic diversity quicker than with the other conditions. However, the confidence intervals do not allow us to conclude that there is a significant difference between the conditions (confidence intervals overlap with 0).

Table 1 compares the error rates of the answers for each condition. We see that in the *color* condition, participants' error rate for $Q1$ and $Q4$ was higher than for *text* and *WSV*, while $Q2$ has more correct answers with the word-sized visualizations (*WSV*) than the other conditions. Table 1 further shows the perceived difficulties

Table 1: *Error rates in percent for the real Twitter data for questions $Q1$, $Q2$ and $Q4$ across conditions, as well as average difficulty ratings for $Q1$ to $Q4$, 1 being easy and 5 being difficult.*

	Error Rates			Difficulty			
	$Q1$	$Q2$	$Q4$	$Q1$	$Q2$	$Q3$	$Q4$
WSV	21%	26%	5%	2.11	2.42	3.58	2.16
Color	50%	38%	22%	2.56	2.67	3.61	2.06
Text	24%	32%	4%	2.16	2.92	3.48	1.72

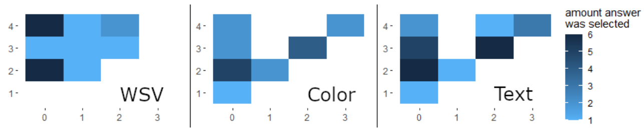


Figure 3: *Answer distribution of $Q3$, which topic cluster x is most different to another topic cluster y, contentwise, for each condition.*

of the tasks, which are similar between conditions, showing no advantage of one condition over another. The results of questions $Q5$ (*How confident were you in your answers?*) and $Q6$ (*How satisfied were you with the display of the different topics?*) are also similar between conditions, ranging between 3 and 4 with a small confidence interval, on a 1 (low) to 5 (high) scale (see OSF for more details). Question $Q3$, which asked about the two topic clusters that differ most, is more subjective and has no single correct answer. In *WSV*, there are two answers that are more common, i.e. topic 0 vs topic 2 and topic 0 vs topic 4 (see Figure 3 left, dark blue cells). In the conditions *color* and *text*, there is no clear distinction. This indicates that with *WSV* it might be easier to compare topics.

4. Discussion

We only saw slight improvements and could not confirm our assumptions that with word-sized visualizations people can answer questions about topic diversity in a Twitter thread quicker or more accurately than with the help of a colored background or text information. We expected that at least answering questions $Q1$, $Q2$ and $Q4$ would be easier with word-sized visualizations. Further, we assume that the topic label in the *text* condition made it easier to gauge the total number of topics for question $Q1$ than if we had shown the most salient words instead of a number. Interestingly, we saw that participants with word-sized visualizations found other types of topic differences ($Q3$) than participants that saw the *text* and *color* condition. Studying visualizations' contribution to content comprehension would be a worthwhile avenue for further research. Another interesting question for future studies is whether improvements to the word-sized visualizations yield better results, which kinds of word-sized visualizations perform best for which scenario, and which interactions are beneficial in a social media setting. Further, there are improvements to be made for text processing, as a clear distinction of topics in social media often is not possible.

5. Acknowledgements

This has been a joint effort funded by multiple agencies, including the University of Stuttgart, the DFG, ESF, MWK, and ANR.

References

[DB79] DAVIES D. L., BOULDIN D. W.: A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-1*, 2 (Apr. 1979), 224–227. doi:10.1109/TPAMI.1979.4766909. 1

[DL16] DOU W., LIU S.: Topic- and Time-Oriented Visual Text Analysis. *IEEE Computer Graphics and Applications 36*, 4 (July 2016), 8–13. doi:10.1109/MCG.2016.73. 1

[GBWI17] GOFFIN P., BOY J., WILLETT W., ISENBERG P.: An Exploratory Study of Word-Scale Graphics in Data-Rich Text Documents. *IEEE Transactions on Visualization and Computer Graphics 23*, 10 (Oct. 2017), 2275–2287. doi:10.1109/TVCG.2016.2618797. 1

[GWFI14] GOFFIN P., WILLETT W., FEKETE J.-D., ISENBERG P.: Exploring the Placement and Design of Word-Scale Visualizations. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (Dec. 2014), 2291–2300. doi:10.1109/TVCG.2014.2346435. 1

[HBK*21] HUTH F., BLASCHECK T., KOCH S., UTZ S., ERTL T.: Word-sized Visualizations for Exploring Discussion Diversity in Social Media. In *Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications* (Mar. 2021), IVAPP '21, SciTePress, pp. 256–265. doi:10.5220/0010328602560265. 1

[JFCS17] JÄNICKE S., FRANZINI G., CHEEMA M. F., SCHEUERMANN G.: Visual Text Analysis in Digital Humanities. *Computer Graphics Forum 36*, 6 (2017), 226–250. doi:https://doi.org/10.1111/cgf.12873. 1

[KK15] KUCHER K., KERREN A.: Text visualization techniques: Taxonomy, visual survey, and community insights. In *2015 IEEE Pacific Visualization Symposium* (Apr. 2015), PacificVis '15, IEEE, pp. 117–121. doi:10.1109/PACIFICVIS.2015.7156366. 1

[LB18] LATIF S., BECK F.: Visually Augmenting Documents With Data. *Computing in Science Engineering 20*, 6 (Nov. 2018), 96–103. doi:10.1109/MCSE.2018.2875316. 1

[Llo82] LLOYD S.: Least squares quantization in PCM. *IEEE Transactions on Information Theory 28*, 2 (Mar. 1982), 129–137. doi:10.1109/TIT.1982.1056489. 1

[pro] © 2021 Prolific. Last accessed 12.04.2021. URL: https://www.prolific.co/. 1

[SMR08] SCHÜTZE H., MANNING C. D., RAGHAVAN P.: Scoring, term weighting & the vector space model. In *Introduction to information retrieval*, vol. 39. Cambridge University Press, 2008, ch. 6. 1

[Tuf06] TUFTE E. R.: *Beautiful Evidence*, 1 ed. Graphics Press, 2006. 1

[twi] © 2021 Twitter, Inc. Last accessed 12.04.2021. URL: https://twitter.com/. 1