Text Visualization Revisited: The State of the Field in 2019

K. Kucher[1] & A. Kerren[1]

[1] Department of Computer Science and Media Technology, Linnaeus University, Sweden

Abstract

Text and document data visualization is an important research field within information visualization and visual analytics with multiple application domains including digital humanities and social media, for instance. During the past five years, we have been collecting text visualization techniques described in peer-reviewed literature, categorizing them according to a detailed categorization schema, and providing the resulting manually curated collection in an online survey browser. In this poster paper, we present the updated results of analyses of this data set as of spring 2019. Compared to the recent surveys and meta-analyses that mainly focus on particular aspects and problems related to text visualization, our results provide an overview of the current state of the text visualization field and the respective research community in general.

CCS Concepts

• ***Human-centered computing*** → ***Visualization techniques; Visualization theory, concepts and paradigms;***

1. Introduction

Text and document data visualization [KK15] is an important research field within information visualization and visual analytics with multiple application domains including digital humanities [BEAC*18, JFCS15, JFCS17] and social media [CLY17, WCG*16], for instance. Several recent survey articles and meta-analyses focus on particular aspects and problems related to this field, including the general approaches [CC16], text highlighting techniques [SOK*16], topic- and time-related text visualization [DL16], visual analysis of scientific literature [FHKM17], integration of word-sized graphics into documents [BW17], sentiment visualization [KPK18], visualization of word vector embeddings [HG18], visual text analytics [LWC*18], and even the existing text visualization surveys themselves [AL19].

During the past five years, we have been collecting text visualization techniques [KK15] described in peer-reviewed literature, categorizing them according to a detailed categorization schema, and providing the resulting manually curated collection in an online survey browser available at `http://textvis.lnu.se`. In this poster paper, we present the updated results of analyses of this data set that provide us with an overview of the text visualization field as of spring 2019, including the insights about the respective research community.

2. Text visualization techniques revisited

Our data set currently includes 430 techniques described in research publications, the majority of which have been published during the past 12–13 years, as displayed in Figure 1. The category statistics presented in Table 1 demonstrate the interest for tasks related to summarization and topics (66% of all entries), relations (48%), trends/patterns (43%), and sentiments and opinions (38%). They also demonstrate the continued interest for corpora/collections (80%) and time-dependent data (47%).

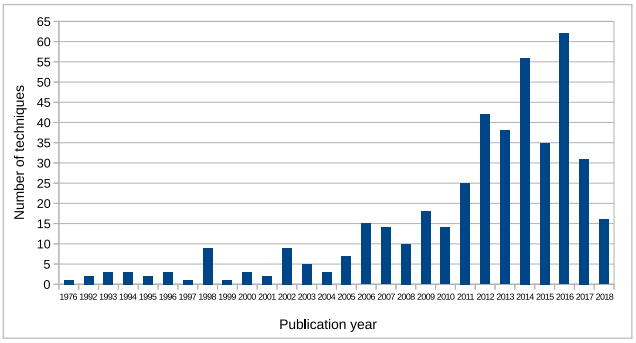

Figure 1: *Histogram of the collected text visualization techniques set (**430** techniques in total as of May 7, 2019) with regard to the publication year.*

The authorship statistics for the current data set reveal the top five authors with regard to number of techniques as Daniel A. Keim (28 entries), Shixia Liu (17 entries), Christian Rohrdantz (13 entries), Christopher Collins (12 entries), and Huamin Qu (12 entries). As seen from Table 2, besides a large number of authors with

DOI: 10.2312/eurp.20191138

Data Domain	338
Online Social Media	146
Communication	33
Patents	5
Reviews / (Medical) Reports	44
Literature/Poems	52
Scientific Articles/Papers	42
Editorial Media	89
Data Source	**425**
Document	89
Corpora	342
Streams	54
Data Properties	**266**
Geospatial	61
Time Series	203
Networks	102

Analytic Tasks	417
Text Summarization / Topic Analysis / Entity Extraction	284
Discourse Analysis	25
Stance Analysis	16
Sentiment Analysis	162
Event Analysis	48
Trend Analysis / Pattern Analysis	184
Lexical/Syntactical Analysis	48
Relation/Connection Analysis	208
Translation / Text Alignment Analysis	19
Visualization Tasks	**430**
Region of Interest	91
Clustering/Classification/Categorization	320
Comparison	352
Overview	403
Monitoring	45
Navigation/Exploration	282
Uncertainty Tackling	27

Visual Dimensionality	430
2D 2D	413
3D 3D	27
Visual Representation	**430**
Line Plot / River	139
Pixel/Area/Matrix	173
Node-Link	147
Clouds/Galaxies	169
Maps	93
Text	214
Glyph/Icon	121
Visual Alignment	**423**
Radial	87
Linear/Parallel	259
Metric	224

Table 1: *The complete categorization of text visualization techniques. Each row contains the number of corresponding techniques in our data set as of May 7, 2019. The percentage relative to the current total of **430** techniques is also illustrated by heatmap-style icons.*

#techniques	1	2	3	4	5	6	7	8	9	10	12	13	17	28
#authors	874	135	52	20	15	13	5	3	3	1	2	1	1	1

Table 2: *Authorship count distribution. The current data set includes **430** text visualization techniques and **1,126** authors in total.*

one or two contributions, the text visualization research community includes a core group of authors with three or more techniques.

Further insights about the relations within the community can be learned from the co-authorship network (1,126 nodes, 2,672 edges), which we have analyzed with Gephi [BHJ09]. Figure 2 reveals that the majority of author nodes are included into isolated connected components of small sizes (less than 15 nodes) while there is a giant connected component with 315 nodes and 1,095 edges present in the graph, which includes the aforementioned authors as cluster center nodes. The largest *betweenness* centrality [LLY13, New10] values in the network are associated with Jaegul Choo, Niklas Elmqvist, Shixia Liu, Daniel A. Keim, and Ross Maciejewski, who are all active contributors in the text visualization community. A GMap [GHK10] based on this co-authorship network is available for the interested readers online at http://gmap.cs.arizona.edu/map/9143/ (last accessed in May 2019).

In this poster paper, we have provided a brief overview of the updated analyses of the text visualization techniques data set maintained by us and provided via an online survey browser. Text visualization remains an active research field and we expect to witness further publications on novel techniques addressing, for instance, tight integration of visualization and machine learning approaches

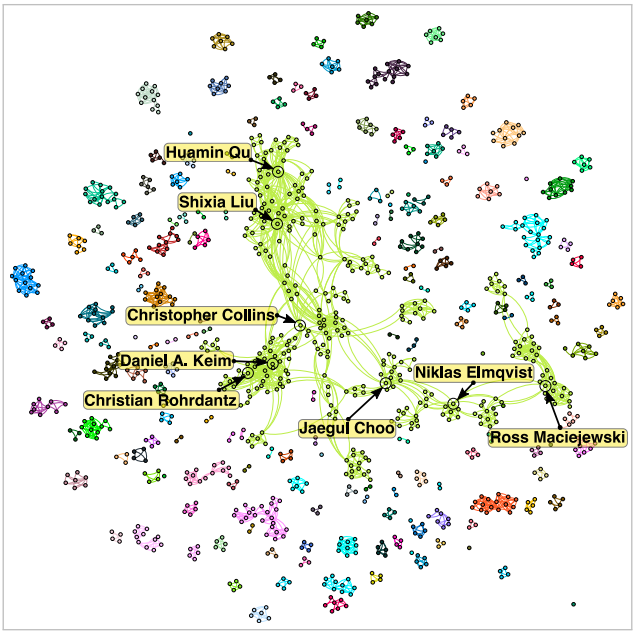

Figure 2: *Co-authorship network for our current data set visualized in Gephi. Note the giant connected component in the center (in green) containing **315** author nodes.*

involving textual data [ERT*17, HG18, LGH*17, LWC*18]; thus, we intend to continue maintaining and extending our data set and the survey browser accordingly.

References

[AL19] ALHARBI M., LARAMEE R. S.: SoS TextVis: An extended survey of surveys on text visualization. *Computers 8*, 1 (2019). doi:10.3390/computers8010017. 1

[BEAC*18] BRADLEY A. J., EL-ASSADY M., COLES K., ALEXANDER E., CHEN M., COLLINS C., JÄNICKE S., WRISLEY D. J.: Visualization and the digital humanities: Moving toward stronger collaborations. *IEEE Computer Graphics and Applications 38*, 6 (Nov. 2018), 26–38. doi:10.1109/MCG.2018.2878900. 1

[BHJ09] BASTIAN M., HEYMANN S., JACOMY M.: Gephi: An open source software for exploring and manipulating networks. In *Proceedings of the 3rd International AAAI Conference on Weblogs and Social Media* (2009), ICWSM '09, AAAI Press, pp. 361–362. URL: https://aaai.org/ocs/index.php/ICWSM/09/paper/view/154. 2

[BW17] BECK F., WEISKOPF D.: Word-sized graphics for scientific texts. *IEEE Transactions on Visualization and Computer Graphics 23*, 6 (June 2017), 1576–1587. doi:10.1109/TVCG.2017.2674958. 1

[CC16] CAO N., CUI W.: *Introduction to Text Visualization*, vol. 1 of *Atlantis Briefs in Artificial Intelligence*. Atlantis Press, 2016. doi:10.2991/978-94-6239-186-4. 1

[CLY17] CHEN S., LIN L., YUAN X.: Social media visual analytics. *Computer Graphics Forum 36*, 3 (June 2017), 563–587. doi:10.1111/cgf.13211. 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

[ERT*17] ENDERT A., RIBARSKY W., TURKAY C., WONG B. W., NABNEY I., BLANCO I. D., ROSSI F.: The state of the art in integrating machine learning into visual analytics. *Computer Graphics Forum 36*, 8 (2017), 458–486. doi:10.1111/cgf.13092. 3

[FHKM17] FEDERICO P., HEIMERL F., KOCH S., MIKSCH S.: A survey on visual approaches for analyzing scientific literature and patents. *IEEE Transactions on Visualization and Computer Graphics 23*, 9 (Sept. 2017), 2179–2198. doi:10.1109/TVCG.2016.2610422. 1

[GHK10] GANSNER E. R., HU Y., KOBOUROV S.: GMap: Visualizing graphs and clusters as maps. In *Proceedings of the IEEE Pacific Visualization Symposium* (2010), PacificVis '10, IEEE, pp. 201–208. doi:10.1109/PACIFICVIS.2010.5429590. 2

[HG18] HEIMERL F., GLEICHER M.: Interactive analysis of word vector embeddings. *Computer Graphics Forum 37*, 3 (2018), 253–265. doi:10.1111/cgf.13417. 1, 3

[JFCS15] JÄNICKE S., FRANZINI G., CHEEMA M. F., SCHEUERMANN G.: On close and distant reading in digital humanities: A survey and future challenges. In *Proceedings of the EG/VGTC Conference on Visualization — STARs* (2015), EuroVis '15, The Eurographics Association. doi:10.2312/eurovisstar.20151113. 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:10.1111/cgf.12873. 1

[KK15] KUCHER K., KERREN A.: Text visualization techniques: Taxonomy, visual survey, and community insights. In *Proceedings of the 8th IEEE Pacific Visualization Symposium* (2015), PacificVis '15, IEEE, pp. 117–121. doi:10.1109/PACIFICVIS.2015.7156366. 1

[KPK18] KUCHER K., PARADIS C., KERREN A.: The state of the art in sentiment visualization. *Computer Graphics Forum 37*, 1 (Feb. 2018), 71–96. doi:10.1111/cgf.13217. 1

[LGH*17] LU Y., GARCIA R., HANSEN B., GLEICHER M., MACIEJEWSKI R.: The state-of-the-art in predictive visual analytics. *Computer Graphics Forum 36*, 3 (2017), 539–562. doi:10.1111/cgf.13210. 3

[LLY13] LI E. Y., LIAO C. H., YEN H. R.: Co-authorship networks and research impact: A social capital perspective. *Research Policy 42*, 9 (2013), 1515–1530. doi:10.1016/j.respol.2013.06.012. 2

[LWC*18] LIU S., WANG X., COLLINS C., DOU W., OUYANG F., EL-ASSADY M., JIANG L., KEIM D. A.: Bridging text visualization and mining: A task-driven survey. *IEEE Transactions on Visualization and Computer Graphics* (2018). doi:10.1109/TVCG.2018.2834341. 1, 3

[New10] NEWMAN M. E. J.: *Networks: An Introduction*. Oxford University Press, 2010. 2

[SOK*16] STROBELT H., OELKE D., KWON B. C., SCHRECK T., PFISTER H.: Guidelines for effective usage of text highlighting techniques. *IEEE Transactions on Visualization and Computer Graphics 22*, 1 (Jan. 2016), 489–498. doi:10.1109/TVCG.2015.2467759. 1

[WCG*16] WU Y., CAO N., GOTZ D., TAN Y.-P., KEIM D. A.: A survey on visual analytics of social media data. *IEEE Transactions on Multimedia 18*, 11 (Nov. 2016), 2135–2148. doi:10.1109/TMM.2016.2614220. 1