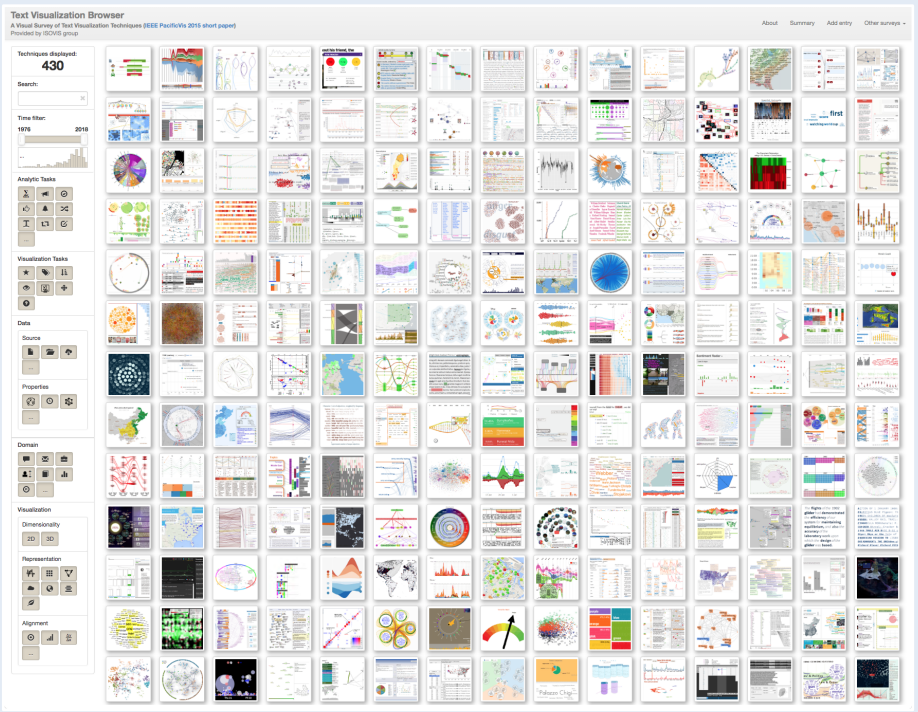
Text Visualization Revisited: The State of the Field in 2019

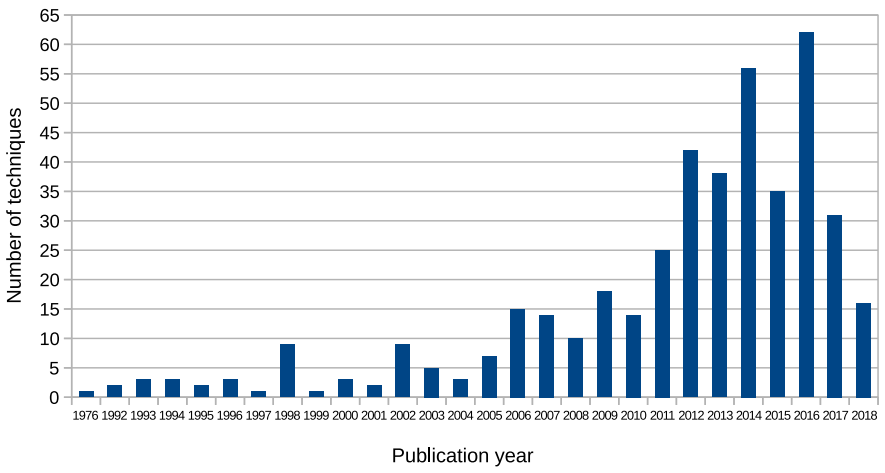
Kostiantyn Kucher* and Andreas Kerren
Linnaeus University, Sweden

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.

The online survey browser based on our manually curated collection of text visualization techniques is available at

http://textvis.lnu.se





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

#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

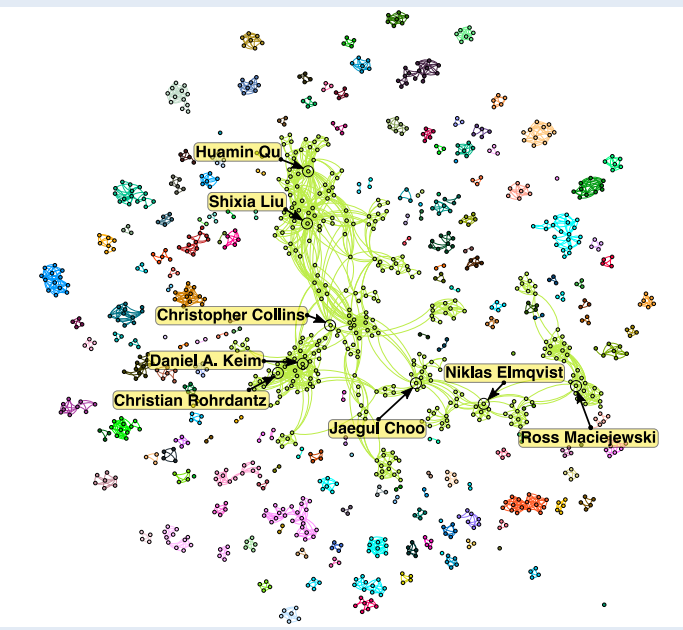
Authorship count distribution based on the corresponding publications. The current data set includes 1,126 authors. Besides a large number of authors with 1 or 2 contributions, the text visualization research community includes a core group of authors with 3+ techniques.

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

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. The category statistics 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%).



The co-authorship network based on the corresponding publications includes 1,126 nodes and 2,672 edges. The giant connected component in the center (in green) consists of 315 nodes and 1,095 edges, including the authors with largest numbers of techniques and largest betweenness centrality values (more details in the poster paper).

*Contact: kostiantyn.kucher@lnu.se

Linnæus University

ISOVIS http://cs.lnu.se/isovis/

EuroVis '19, 3–7 June 2019, Porto, Portugal