Text Visualization Browser:
A Visual Survey of Text Visualization Techniques

Kostiantyn Kucher*   Andreas Kerren†

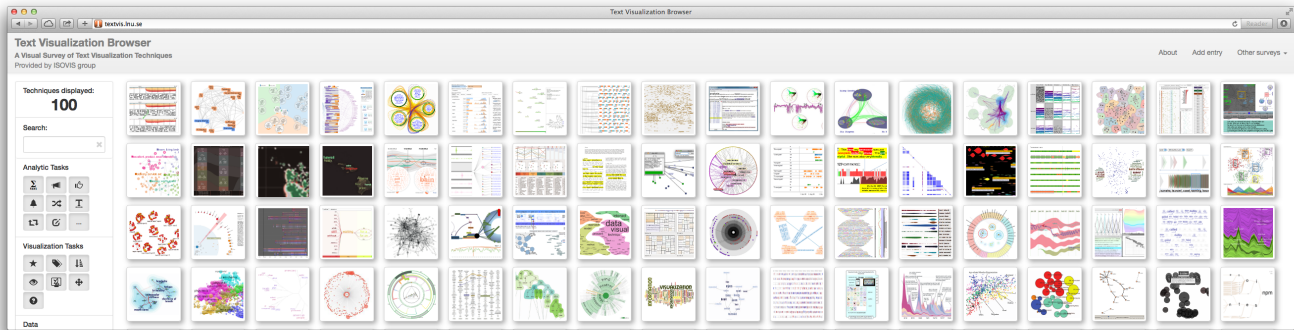ISOVIS Group, Department of Computer Science, Linnaeus University, Växjö, Sweden

Figure 1: The web-based user interface of our visual survey. By using the interaction panel on the left hand side, researchers can look for specific visualization techniques and filter out entries with respect to a set of categories (cf. the taxonomy given in Sect. 2). Details for a selected entry are shown by clicking on a thumbnail image in the main view. The survey contains 100 categorized visualization techniques by June 24, 2014.

ABSTRACT

Text visualization has become a growing and increasingly important subfield of information visualization. Thus, it is getting harder for researchers to look for related work with specific tasks or visual metaphors in mind. In this poster, we present an interactive visual survey of text visualization techniques that can be used for the purposes of search for related work, introduction to the subfield and gaining insight into research trends.

Keywords: Visualization, text visualization, survey, interaction, web-based systems

1 INTRODUCTION

The interest for text visualization and visual text analytics has been increasing for the last ten years. The reasons for this development are manifold, but for sure the availability of large amounts of heterogeneous text data (caused by the popularity of online social media) and the adoption of text processing algorithms (e.g., for topic modeling) by the InfoVis and Visual Analytics communities are two possible explanations. Inspired by the TreeVis.net [3] and TimeVis [4] projects, we propose an interactive visual survey of text visualization techniques that can be used for getting an overview of the field, teaching purposes, and finding related work based on various categories defined in a survey taxonomy. Our web-based survey browser is available at:

<div align="center">

http://textvis.lnu.se/

</div>

The term 'text visualization' is typically used for information visualization techniques which in some cases focus on raw textual data, in other cases on results of text mining algorithms. In the same way, they can be rather general or very specialized and dedicated to specific analytic tasks or application domains. This is the reason

*e-mail: kostiantyn.kucher@lnu.se
†e-mail: andreas.kerren@lnu.se

why we have decided to construct a taxonomy with numerous categories and subcategories that is exploited by the survey browser in order to facilitate the interactive exploration of the current set of entries. Our visual survey has been implemented as an interactive web page and includes 100 techniques at present originating from peer-reviewed work in InfoVis, Visual Analytics and other relevant research fields. After a short discussion on relevant surveys in the following, we highlight the taxonomy used by our survey browser as well as some implementation details.

Related Surveys There are a number of survey papers in the literature that focus on text visualization or its specific subproblems. Šilić and Bašić [5] classify text visualization methods with regard to data source, underlying text representation and processing method, temporal aspects, and supported user interactions. Alencar et al. [1] describe techniques by means of data source, underlying text representation, visual metaphor, layout, and supported user tasks. Gan et al. [2] categorize techniques on the basis of data source, user tasks, visual representation, and supported interactions. The recent work of Wanner et al. [6] on event detection in texts classifies visualization approaches with regard to data source, text processing methods, event detection methods, visualization representations, and tasks.

Finally, the aforementioned visual survey projects use dimensionality, visualization metaphor and alignment to classify tree-oriented techniques [3], and data properties, temporal properties, visual representation (dimensionality and animation support) to classify time-oriented techniques [4].

2 SURVEY TAXONOMY

We have arranged a taxonomy (cf. Fig. 2) with multiple categories in order to classify the techniques with fine granularity. While we cannot claim that our classification is absolutely definite (numerous techniques have been ambiguous, especially in case of hybrid approaches), we have tried to base the choice of categories for particular entries on the description and claims of the original author(s). For example, certain techniques could be easily applied to domains other than originally described, but we do not reflect that in our
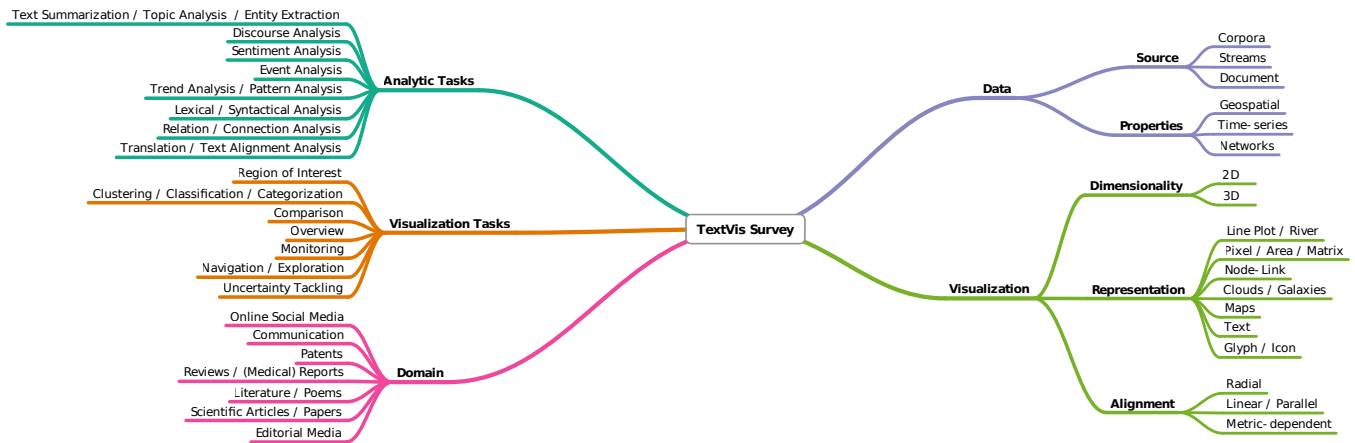
Figure 2: The taxonomy of text visualization techniques used in our visual survey.

choice of categories for those techniques. On the other hand, some papers mentioned specific domains only for the sake of giving examples, though the corresponding techniques were not tailored for those domains. In such cases, we have not assigned entries to those domain categories. Our top-level categories include the following:

Analytic Tasks This category describes high-level analytic tasks that are facilitated by corresponding techniques: these items are critical to the main analysis goals that users expect to achieve when employing a visualization technique.

Visualization Tasks This category describes lower-level representation and interaction tasks that are supported by the techniques, for instance, clustering or the identification of a region of interest.

Domain This category describes the dedicated text domains a technique was developed for.

Data This category comprises two subcategories, *Source* and *Properties*, that describe the type of data source and—if available—special properties (e.g., time-series) of data used by the corresponding techniques.

Visualization This category consists of three subcategories: *Dimensionality*, *Representation*, and *Alignment* (i.e., the layout). By using them, we can describe the corresponding properties of visual representations used by the techniques.

3 INTERACTIVE BROWSER

We have implemented our visual survey as an interactive HTML/JavaScript page that merely requires a modern web browser for access, see Fig. 1 for a screenshot. The survey browser has a main view with a collection of thumbnails (ordered by time) that represent the individual visualization techniques as well as filter controls that comprise text search field and category radio buttons.

Our interface does not provide dedicated controls for selecting a time range, but the publication year is available for every technique reference. So, it is possible to use the text search field to filter entries for a specific year. After clicking on an entry's thumbnail image, the corresponding details are displayed in a dialog box. Here, a slightly larger thumbnail, a complete list of assigned category tags, a bibliographical reference, a URL (optional), and a link to a BibTeX file (if available) are displayed (see Fig. 3).

We have also provided an additional form for authors who wish to add a new entry to our survey. The form generates a JSON entry that can be sent to us via email to prevent direct-manipulation of the survey browser content. Finally, we visualize some basic statistics about the current entry set in the "About" dialog. Since the techniques can be assigned with multiple category tags, the sets of corresponding techniques overlap for sibling categories—therefore,
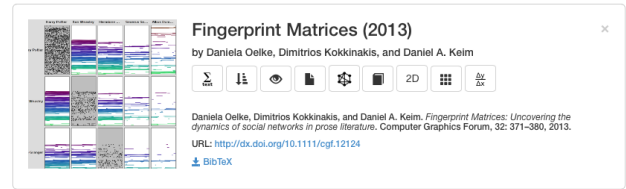

Figure 3: Details of a survey entry.

we use simple bar charts for showing the statistics. The results indicate that currently all entries are assigned with category tags pertaining to visualization tasks, data source, visual dimensionality, and representation; 95 entries are relevant to specific analytic tasks, and 74 are relevant to particular data domains.

4 CONCLUSION

In this poster abstract, we presented a visual survey of text visualization techniques. The main contributions of this work are the survey taxonomy and the interactive web-based browser that currently includes 100 techniques. In the future, we plan to continue including new entries and create additional visualizations based on the data the entries provide.

REFERENCES

[1] A. B. Alencar, M. C. F. de Oliveira, and F. V. Paulovich. Seeing Beyond Reading: A Survey on Visual Text Analytics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(6):476–492, 2012.

[2] Q. Gan, M. Zhu, M. Li, T. Liang, Y. Cao, and B. Zhou. Document Visualization: An Overview of Current Research. *Wiley Interdisciplinary Reviews: Computational Statistics*, 6(1):19–36, 2014.

[3] H.-J. Schulz. TreeVis.net: A Tree Visualization Reference. *Computer Graphics and Applications, IEEE*, 31(6):11–15, Nov 2011.

[4] C. Tominski and W. Aigner. The TimeVis Browser. Available at http://survey.timeviz.net/. Accessed: 2014-06-18.

[5] A. Šilić and B. D. Bašić. Visualization of Text Streams: A Survey. In R. Setchi, I. Jordanov, R. Howlett, and L. Jain, editors, *Knowledge-Based and Intelligent Information and Engineering Systems*, volume 6277 of *Lecture Notes in Computer Science*, pages 31–43. Springer Berlin Heidelberg, 2010.

[6] F. Wanner, A. Stoffel, D. Jäckle, B. C. Kwon, A. Weiler, and D. A. Keim. State-of-the-Art Report of Visual Analysis for Event Detection in Text Data Streams. *Computer Graphics Forum*, 33(3), 2014.