

Hierarchical Topic Maps for Visual Exploration and Comparison of Documents

M. Tytarenko¹, L. Shao^{1,2}, T.W. Rutar¹, M.A. Bedek³, C. Krenn⁴, S. Lengauer¹, and T. Schreck¹

¹Institute of Computer Graphics and Knowledge Visualization, Graz University of Technology

²Fraunhofer Austria Center for Data Driven Design

³Institute of Psychology, University of Graz

⁴Institute of General Practice and Evidence-based Health Services Research, Medical University of Graz

Abstract

Information visualization nowadays provides a large amount of different text visualization techniques that help to summarize and present textual information in an intuitive and comprehensible manner. Despite many advancements, there remains a gap in effectively illustrating the thematic and structural distinction between similar documents in a hierarchical and interactive manner. We present the Hierarchical Topic Maps (HTM), an innovative approach, inspired by Tile Bars, that addresses this gap by illustrating the content distribution across a document hierarchically. Our model incorporates a multi-resolution display feature, enabling users, in particular curators of large document collections, with the need to quickly obtain text document structure, to delve deeper and draw more meaningful conclusions, to assess thematic similarities at multiple levels of detail, as well as facilitate nuanced comparison of textual documents. We demonstrate the effectiveness of both our approach's document exploration and document comparison potential by two exemplary use case scenarios. Our findings suggest that HTM not only simplifies the document overview process but also provides a practical solution for comparing thematic structures, thereby offering contributions to the field of text visualization and visualization analytics.

CCS Concepts

• **Human-centered computing** → **Treemaps**; Visualization systems and tools; • **Information systems** → Digital libraries and archives; • **Applied computing** → Document searching;

1. Introduction

With the rapid growth of online text documents distributed by digital libraries and specialized systems, it is becoming increasingly difficult for the users, e.g., researchers or document curators, to keep track, compare different documents and obtain specific information quickly. This means that humans cannot process large corpus (thousands, hundreds of thousands of documents) in a realistic time frame, especially when the consumed information has to be correlated with information from other documents. Tailored visualizations, relying on computerized methods for text abstractions, content modelling, similarity computations, etc., facilitate the search for relevant information on a contextual and structural level.

A very common use case scenario involves users wanting to quickly evaluate whether a document is qualified to fulfill their information need and – if this is the case – where within the document and in which context the respective information resides. Such a *query-to-document* [LX14] matching requires that the information need can be accurately described by query, in most cases a set of keywords, which describe a semantically coherent concept. While it is possible to compile such a set manually, for practical

applications it is advantageous to automatically determine the different concepts of a document through a so-called *topic modeling* (TM) [DC14], a concept which captures said concepts unsupervised as collections of related terms. Such topic models are also used to efficiently compare documents on a semantic level. That is, document comparison is another common use case as users want to know which document from a collection is the most appropriate one to address their information need or review where the structural difference between two similar documents are.

As semantic concepts are usually defined by multiple keywords, illustrating the complexity and diversity of topics within a document. Such topics can be compiled either manually, which demands extensive domain knowledge and time, or also automatically through TM approaches [BNJ03] – the automatic structuring of a document's content into sets of descriptive keywords.

In response to the common challenge of comparing topic distributions across documents for better selection, curation, and understanding, we present Hierarchical Topic Maps (HTM). This visualization tool enhances interactive document exploration with keyword-based queries, enabling more effective visual compari-

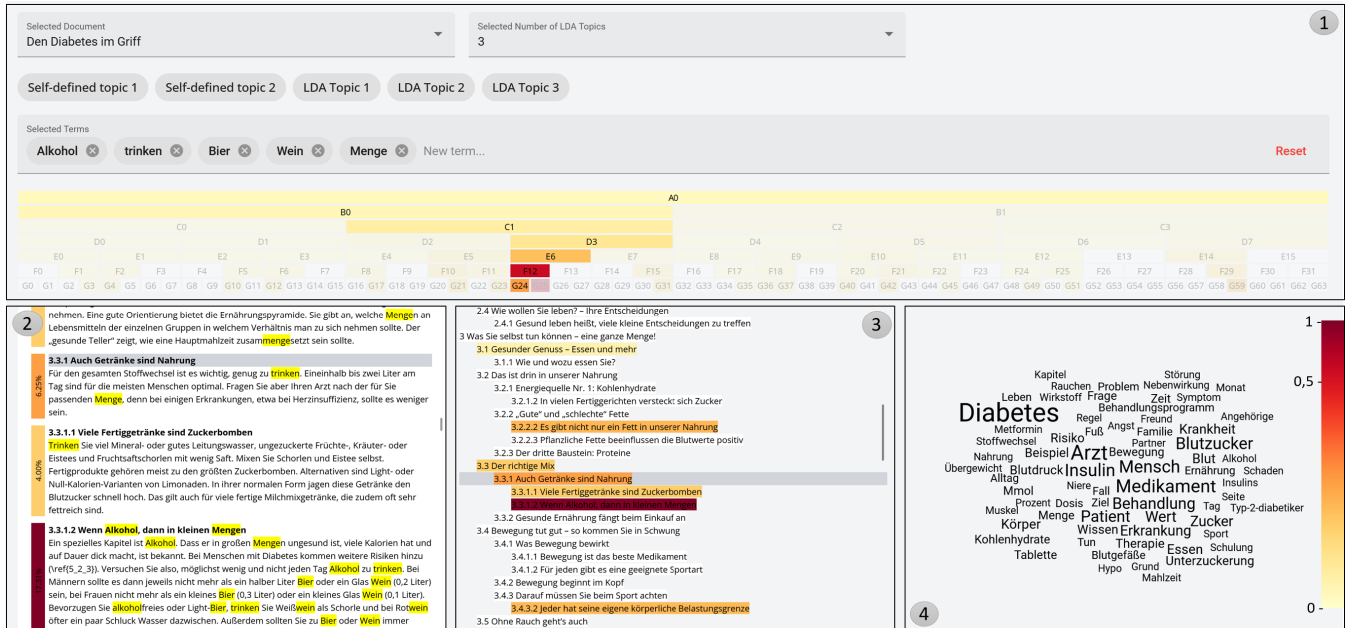


Figure 1: The HTM shows the presence and distribution of a manually or automatically generated topic within a document through a balanced binary-tree layout ①. Linked views, showing the document’s full-text ② and table-of-content ③ allow to review and explore the textual content associated with a cell. A word cloud ④, comprising the document’s most frequently terms allows to easily compile a query vector.

son of topics and deeper understanding of contextual relationships within the data. Our approach uses a balanced binary tree layout to represent text segments and allows the user to navigate through layers of increasingly fine-grained text passages in a divide-and-conquer manner. The visual design of our HTM represents the text segments with a tile analogy, with each level of the tree having twice as many tiles as the previous level. To detect the correlation of a given topic within the tree structure, we use a straightforward frequency measure and color code the relevance for each segment. This method not only assist with deeper understanding of complex datasets but also allows users to uncover hidden topics and relationships between topics in their data. Existing approaches, such as the Tile Bar [Hea95] analogy (the conceptually most similar method), commonly suffer from discretization to a fixed segment size for the visual representation – i.e., concepts stretching over longer passages or concepts appearing only briefly within a short passage remain undetected. Previous research in psychology has shown that visual representations, such as node-link diagrams, can help to understand content-based relations by specifying the relationship that exists between the content [SAH07]. The advantage of our hierarchical representation is that such macro- and micro structure are also visible. For the HTM, the query can be defined by a user or be generated automatically with a TM approach.

In the following, we describe how the HTM compares to other existing techniques (Sec. 2), before our design is covered in depth (Sec. 3). To demonstrate the usefulness of our approach, we compare our technique to the classic Tile Bar approach show two exemplary use case scenarios for the mentioned tasks of document exploration and document comparison (Sec. 4). Finally, we discuss the advantages and possible extension of our approach (Sec. 5).

2. Related Work

There are a large variety of existing visualization techniques that can be challenging to choose for specific data. Depending on the nature of the data and visualization goals, techniques can be selected from a collection of known approaches, including but not limited to: Theme River [HHN00] that illustrates the temporal progression of topics; Word Cloud [HLE14] an unsupervised document visualization approach, based on the *Wordle word cloud algorithm* [SI10]; Topic Islands [Roc89] uses wavelets to summarize the thematic characteristics of an individual document; FacetAtlas [CSL*10] that connects global and local text patterns, providing comprehensive visualization of document collections, etc. For textual information in particular, there are also many know surveys [LWC*19, AL19, AdOP12] that act like a lighthouses, highlighting key trends and methodologies as well as classifying them and helping to chose the right method for specific data.

One way to speed up the exploration in large document collections could be the use of distant reading and document visualization techniques that allow users to make sense of the large volumes of text, as well as analyse, classify, compare and discern key ideas and topics from all of those texts. To this end, visualization techniques for TM, such as LDAvis [SS14] and Hiérarchie [SHM14], can be used to obtain a global view of the corpus, the main topics it contains, and how they differ from each other. However, to extract the topics, most approaches rely on probabilistic models, such as LDA [BNJ03], which are not always easily interpreted by humans. Even though ArchiText [KDEP20] deal with LDA imperfection, by allowing a scalability in hierarchical topic modeling, enabling users interactivity and better understanding of the data corpus by

representing information at multiple levels of detail. All these approaches attempt to provide a good overview of the corpus as a whole, but do not support the exploration of individual documents. Our visualization approach helps the user to understand the distributions of topics within a single document and compare multiple documents for relevance.

Termite [CMH12] goes further, allowing analysts to reveal topical overlap and identify coherent and significant themes within a document, it compares terms both within and across LDA using a matrix layout, where rows correspond to terms and columns to topics. It provides a broad understanding of the model's overall properties, however, this approach restricts its ability to thoroughly examine individual topics, by depicting a potentially different set of terms tailored for each individual topic. Instead, in our approach we consider this limitation by dynamically adjusting displayed terms through a hierarchical binary tree structure, allowing for nuanced interaction and direct document comparisons.

Another possibility would be the use of the classic Tile Bars visualization [Hea95] – the direct predecessor of our design – that shows the distribution of relevant sections or tiles in full-text documents for a specific topic defined by keywords. This visualization technique uses a multi-paragraph segmentation that splits a document into adjacent and non-overlapping segments, called TextTiles. More recent approaches like Reordered TileBars [TH09] and Deep-TileBars [TY19] provide a better focus and navigation through document exploration. The method by [TH09], depicts a document as a large rectangle of small squares, with rows indicating query terms and columns representing text positions, it helps to intuitively understand the document's length, as well as the frequency and distribution of each term. However, a known problem of these approaches is to determine where one set of terms ends and the next begins. Furthermore, the results of the algorithm are difficult to evaluate and comparisons to human judgments show that the results are imperfect. On the other hand, [TY19] build an interaction matrix by comparing a word and a topical segment, it matches queries to documents at the topic level, rather than at the word level. It gives a more detailed view of the document's content and provides complex user interface design, which makes it more difficult to intuitively understand and navigate through the visualization, especially when dealing with lengthy or complex documents.

The landscape of text visualization is rich with diverse techniques, each suited to specific aspects of document analysis but often limited when addressing complex or large-scale datasets. Traditional methods such as Tree Map [JS91], Sunburst [SZ00], and Icicle Plots [KL83] have also provided inspiration for our approach. Even though these are effective visualization techniques, they work mainly for hierarchical data and, depending on the data, may show limitations including depth representation, overcrowding, lack of detail and it may be hard to compare different hierarchical structures. To address these issues, our HTM allow users to manually adjust the depth representation, as well as employed color-coding that reflects the similarity computations of thematic clusters, enhance both the granularity and clarity of the visual analysis, making it easier to compare diverse document structures effectively. Luboschik et al. [LMS*12] showed that such hierarchical representations can also be used to guide the exploration of multiscale time

series data, where the data deviate between subsequent scales. In contrast to their approach, we visualize the heterogeneity of topics between different scales within a document.

Although all the techniques that we mentioned, have effectively covered the topic of document visualization, but the crucial factor of context is not considered, leading to potential misinterpretations of textual content. Our approach emphasizes this context, bridging the gap by revealing hidden meanings, facilitating document comparisons, enhancing the overall depth and clarity of visualization.

3. Hierarchical Topic Maps

The basis of our proposed concept is a visual tree-structure where we illustrate the similarity of branches to a given topic using a color coding. To quantify the similarity between a topic and a text passage, we employ a custom error function, inspired by established text processing principles. The topic, which serves as a query, can be defined by a user or be generated automatically with a TM approach. To this end we subject the document's tokenized full text to a *topic modeling* algorithm.

Visual Encoding and Interactivity The visual design is inspired from well-known visualization techniques, such as Tile Bars [Hea95], and reveals the occurrences of multiple keywords within a given document segment. The necessary segmentation is achieved by recursively splitting the text into two equally long segments (Figure 1①). Hence the uppermost level represents an entire document, which is further divided into two halves at the second level, four quarters at the third level, and so on. This allows text documents to be represented in a hierarchical fashion and enables the exploration of topics at different levels of granularity. The depth τ of the tree – its number of levels – is determined from the document's length l (in number of words) by $\tau = \lfloor \alpha \log l \rfloor$, with the hyperparameter α ($\alpha = 0.75$ used in all experiments) governing the granularity. Each cell is colorized with an established sequential *YlOrRd* color scheme [HB03], reflecting the results of the cell-wise similarity computation. To visually magnify thematic clusters a log-scale mapping to the color space is applied. Cell labels ($A0, \dots, G63$) are a combination of a capital letter, reflecting the level in the map and a number for the offset within this level.

To support the exploration of individual segments, the cells of the HTM are linked to two detail views that show both the respective document's full-text (Figure 1②) and its Table of Contents (ToC) (Figure 1③). Clicking a cell filters the ToC entries and texts, such that only associated elements are displayed. Additionally those remaining elements are similarly colorized according to their similarity to the query. This allows a user to drill-down and evaluate the textual embedding of thematic section.

Similarity Computation For the intensity colorization of the HTM's cells, we need an objective quantization for the similarity of a topic to the associated text of a cell. This cell-wise similarity is given by $|\mathcal{T}|^{-1} \sum_{q \in \mathcal{Q}} \text{count}(q \in \mathcal{T})$, with \mathcal{Q} as the set of query terms and \mathcal{T} as the tokenized cell text, after stop-word removal and stemming/lemmatization.

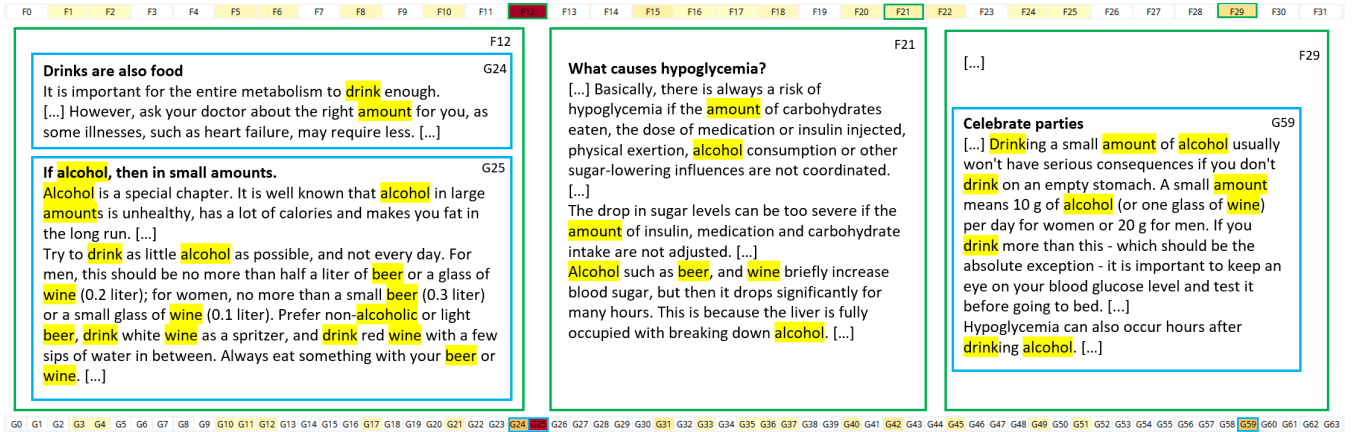


Figure 2: Tile bars visualizations for the AOK [BVP21] brochure with 2^5 and 2^6 tiles and the keyword query $\{‘alcohol’, ‘beer’, ‘wine’, ‘drinking’, ‘amount’\}$. The contents of the top-3 results of both configurations are displayed together with the respective labels F and G and color-coded in green and blue, respectively. Please note that the original text is in German and has been translated into English by us.

Querying The HTM supports two different options for generating a query, constituted by a set of descriptive keywords. Firstly, a user can manually specific those keywords manually in a multiselect form (Figure 1①). A Word Cloud (Figure 1④), for showing the document’s most prominent tokens, gives the user a first impression regarding said document’s thematic content. In our design it also serves as an intuitive interface to add keywords via clicking. While this ensures that a chosen keywords also appears in the text, additionally keywords can also be added via typing. Secondly, queries can be selected from an automatically generated selection of meaningfully queries. To this end, we employ a *topic modeling* algorithm Latent Dirichlet Allocation (LDA) [BNJ03] for single documents and a non-parametric Hierarchic Dirichlet Process (HDP) [TJBB06] for document comparison, to obtain sets of thematically homogeneous keywords. Of course both querying modalities can be used interchangeable and can also be combined – e.g., a topic selected from the topic models can be edited and extended with typed keywords or keywords from the word cloud.

4. Exemplary Use Case Scenarios

We implemented the above described design in a working prototype, using a web stack (an Angular[†]-based front end, paired with a back end in Python), which we use to evaluate the concept’s effectiveness for exemplaric use cases. The two main applications of the HTM are the evaluation of a single document and the content-wise comparison of two (similar) documents. Both usages which will be discussed in the following on the basis of meaningful examples.

Interactive Document Exploration First, we demonstrate how our visualization technique can be used to visually explore interesting topics in the health domain and compare our HTM with the most closely related visualization – the Tile Bars [Hea95] representation. To this end, we look at a fictional scenario in which a pa-

tient, who has recently been diagnosed with Type 2 Diabetes Mellitus (T2DM) wants to learn more about the disease. We assume a patient who enjoys drinking alcohol and would like to learn more about the effects of alcohol in with regards to T2DM. Above all, the patient wants to know whether it is possible to continue drinking alcohol with this disease. As information source, we used a T2DM information brochure from the German health insurance company AOK [BVP21]. The text document is available in PDF format and contains over 130 pages of comprehensive health information, including figures, tables, and infographics. To extract the underlying health information, we focus exclusively on the full text, which we extract using Apache PDFBox[‡].

The patient starts the exploration with a tilebar (similar to a single level of our HTM) and searches for the keywords ‘alcohol’, ‘beer’, ‘wine’, ‘drink’, and ‘amount’. We evaluate two different tiling configurations, splitting the given text into 2^5 and 2^6 equal-sized tiles (corresponding to the HTM’s G and F row in Figure. 1).

By looking solely at the top three results of these tilings (Fig. 2), the patient may think that alcohol is not that problematic in the context of diabetes as long as it is not too much. The HTM, however, reveals that F12/G25 are part of the superior level $D3$, which forks into the two thematic clusters $F12/G25$ and $G31$. While $D3$ deals with the right mixture of food and how exercise can help, $G31$ reveals that the risk of hypoglycemia increases if you drink alcohol after exercise. This small yet important piece of information could go unnoticed in the tilebar view, since many other unimportant tiles achieve a similar relevance score.

Document Comparison Next, we demonstrate how our technique can be used to compare different documents. To this end, we look into two related survey papers on interactive lenses in visualization by Tominski et al. [TGK*14, TGK*17] and try to find similarities and differences. The latter paper, published in 2017, is a revised

[†] <https://angular.io/>

[‡] <https://pdfbox.apache.org/>

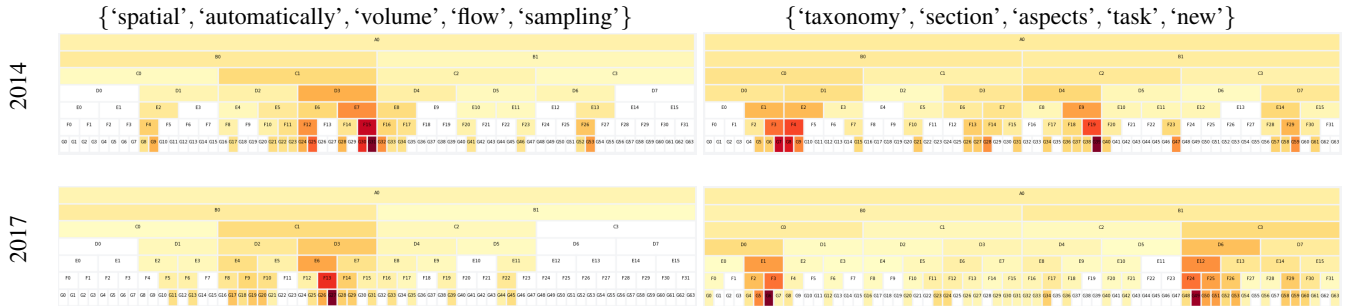


Figure 3: HTMs for topics t_1 (left) and t_2 (right) over two different versions (from 2014 and 2017) of a survey on interactive lenses [TKG*14, TKG*17]. The same color scheme as in Fig. 1 is used.

and improved version of the earlier 2014 survey. We are interested in understanding how the surveys relate and where they differ, to unveil recent developments in the state-of-the-art.

To identify what exactly has changed or expanded in the new edition, we employ the HTM in combination with queries, stemming from the topics obtained through the HDP. A total of 5 topics are identified this way. We use the first two of them ($t_1 = \{\text{'spatial', 'automatically', 'volume', 'flow', 'sampling'}\}$ and $t_2 = \{\text{'taxonomy', 'section', 'aspects', 'task', 'new'}\}$) to investigate the 2014-to-2017 changes (Fig. 3). Regarding t_1 , we observe that the intensities of the 2014 survey (Fig. 3, top left) are slightly shifted to the left in the 2017 survey (Fig. 3, bottom left) due to additional new content (C_3) towards the end of the latter. In general, we were able to identify the same pattern shifts for multiple topics, indicating the new sections at the end of the document. Checking the distribution for t_2 (Fig. 3, right) confirms that these newly added parts mostly relate to t_2 thematically. A closer look at the papers confirms that two new sections on the taxonomy of lenses and detailed examples had been added at the end of the 2017 edition.

5. Limitations and Future Work

Our HTM concept builds upon various text processing techniques for which we rely on existing algorithms. Each of those come with their respective advantages and limitations. Importantly, these elements are interchangeable, meaning extensions or changes in this area can further enhance the concept's potential.

One example is the function for the text similarity, which quantifies the overall similarity of a paragraph to the query terms. But it does not take into account how many of the query terms actually appear. On the contrary, a large quantity of even a single of the terms will result in a high similarity. For some applications, it could, however, be essential that the query terms appear jointly. In future work we plan to define an alternative similarity function which also considers both the quantity and variety of terms.

On a similar note, the employed similarity mapping operates on a per-word-stem-level, meaning that similarity on a conceptual-level (e.g., synonyms like 'diet', 'food', 'sustenance', etc. for the concept 'nutrition') will not be recognized as such. There are existing lemmatization [KGJ23] and tokenization [GBG*18] approaches to this end, which can be used if this behavior is desired. In addition,

open source libraries for natural language processing and dictionaries such as *Gensim*, *FastText*, and *NLTK*, could be used to augment user queries with similar words, synonyms and contextual terms.

For document comparison, the granularity of tiles and the number of hierarchical levels significantly impact the outcome, especially when we dealing with documents of varying lengths. Different lengths of documents result in slightly different binnings, meaning that even parts which are verbatim across documents exhibit (slightly) different intensities distributions (Fig. 3). Finer granularity and more levels allow for detailed analysis but may lead to information overload and difficulty in navigating between topics in longer documents. In the future we want to investigate two orthogonal approaches to address this issue. Firstly, the documents' structures (chapters, sections, subsection, etc.) can be used for the upper levels of the HTM instead of the binary split. This would allow to compare, for example, on a per-chapter basis. Secondly, discretization errors can be smoothed by a sliding-window filter for the different levels.

Lastly, we want to extend our concept for comparison of a larger number of documents. To this end, we intend to display the documents' HTMs in a vertical list, ranked by the respective overall similarities to the query terms. To support the comparison of substructures within documents user interactions could be integrated to search for similar subtrees in the document collection.

6. Conclusion

We presented a new document visualization tool – Hierarchical Topic Maps (HTM) designed to enhance the exploration of single document, reveal hidden topics and facilitate the comparison between similar documents, that demonstrates significant potential in improving document visualization and exploration. Our approach follows this idea with a hierarchic design concept, inspired by the Tile Bar concept. It allows a user to assess thematic similarities across a document in an intuitive and simple manner. Also – as opposed to many other visualizations – our design does not suffer from discretization errors, due to its hierarchic nature. I.e., with a fixed sampling the presence of a topics could remain unnoticed if the bin-size is chosen either too large or small. We investigate the mentioned benefits of our concept through a set of carefully selected use cases, which confirm its potential for interactive document exploration.

Acknowledgements

This work was funded by the Austrian Science Fund (FWF) as part of the project 'Human-Centered Interactive Adaptive Visual Approaches in High-Quality Health Information' (A+CHIS; Grant No. FG 11-B).

References

- [AdOP12] ALENCAR A. B., DE OLIVEIRA M. C. F., PAULOVICH F. V.: Seeing beyond reading: a survey on visual text analytics. *WIREs Data Mining Knowl. Discov.* 2, 6 (2012), 476–492. doi:10.1002/WIDM.1071. 2
- [AL19] ALHARBI M., LARAMEE R. S.: SoS TextVis: An Extended Survey of Surveys on Text Visualization. *Computers* 8, 1 (2019). doi:10.3390/computers8010017. 2
- [BNJ03] BLEI D. M., NG A. Y., JORDAN M. I.: Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022. 1, 2, 4
- [BVP21] BAUMGART J., VIEGENER U., POHL C.: Den Diabetes im Griff: Ein Handbuch für Patientinnen und Patienten mit Diabetes mellitus Typ 2. *AOK-Bundesverband, Berlin* (2021). URL: <https://www.aok.de/pk/magazin/cms/fileadmin/pk/pdf/patientenhandbuch-diabetes.pdf>. 4
- [CMH12] CHUANG J., MANNING C. D., HEER J.: Termite: Visualization Techniques for Assessing Textual Topic Models. In *Proceedings of the international working conference on advanced visual interfaces* (2012), AVI '12, Association for Computing Machinery, pp. 74–77. doi:10.1145/2254556.2254572. 3
- [CSL*10] CAO N., SUN J., LIN Y.-R., GOTZ D., LIU S., QU H.: FacetAtlas: Multifaceted Visualization for Rich Text Corpora. *IEEE Transactions on Visualization and Computer Graphics* 16, 6 (2010), 1172–1181. doi:10.1109/TVCG.2010.154. 2
- [DC14] DING W., CHEN C.: Dynamic topic detection and tracking: A comparison of HDP, C-word, and cocitation methods. *Journal of the Association for Information Science and Technology* 65, 10 (2014), 2084–2097. doi:https://doi.org/10.1002/asi.23134. 1
- [GBG*18] GRAVE E., BOJANOWSKI P., GUPTA P., JOULIN A., MIKOLOV T.: Learning word vectors for 157 languages. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018)* (2018). doi:10.48550/arXiv.1802.06893. 5
- [HB03] HARROWER M., BREWER C. A.: ColorBrewer.org: an online tool for selecting colour schemes for maps. *The Cartographic Journal* 40, 1 (2003), 27–37. doi:10.1179/000870403235002042. 3
- [Hea95] HEARST M. A.: TileBars: visualization of term distribution information in full text information access. In *International Conference on Human Factors in Computing Systems* (1995). URL: <https://api.semanticscholar.org/CorpusID:2152596>. 2, 3, 4
- [HHN00] HAVRE S., HETZLER B., NOWELL L.: ThemeRiver: visualizing theme changes over time. In *IEEE Symposium on Information Visualization 2000. INFOVIS 2000. Proceedings* (2000), pp. 115–123. doi:10.1109/INFVIS.2000.885098. 2
- [HLE14] HEIMERL F., LOHMANN S., LANGE S., ERTL T.: Word Cloud Explorer: Text Analytics Based on Word Clouds. In *2014 47th Hawaii International Conference on System Sciences* (2014), pp. 1833–1842. doi:10.1109/HICSS.2014.231. 2
- [JS91] JOHNSON B., SHNEIDERMAN B.: Tree-Maps: a space-filling approach to the visualization of hierarchical information structures. In *Proceedings of the 2nd Conference on Visualization '91* (Washington, DC, USA, 1991), VIS '91, IEEE Computer Society Press, p. 284–291. doi:10.1109/VISUAL.1991.175815. 3
- [KDEP20] KIM H., DRAKE B. L., ENDERT A., PARK H.: ArchiText: Interactive Hierarchical Topic Modeling. *IEEE Transactions on Visualization and Computer Graphics* 27, 9 (2020), 3644–3655. doi:10.1109/TVCG.2020.2981456. 2
- [KJG23] KAVANAGH J., GREENHOW K., JORDANOUS A.: Assessing the effects of lemmatisation and spell checking on sentiment analysis of online reviews. In *2023 IEEE 17th International Conference on Semantic Computing (ICSC)* (2023), pp. 235–238. doi:10.1109/ICSC56153.2023.00046. 5
- [KL83] KRUSKAL J. B., LANDWEHR J. M.: Icicle Plots: Better Displays for Hierarchical Clustering. *The American Statistician* 37, 2 (1983), 162–168. doi:10.2307/2685881. 3
- [LMS*12] LUBOSCHIK M., MAUS C., SCHULZ H.-J., SCHUMANN H., UHRMACHER A.: Heterogeneity-based guidance for exploring multi-scale data in systems biology. In *2012 IEEE Symposium on Biological Data Visualization (BioVis)* (2012), pp. 33–40. doi:10.1109/BioVis.2012.6378590. 3
- [LWC*19] 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* 25, 7 (2019), 2482–2504. doi:10.1109/TVCG.2018.2834341. 2
- [LX14] LI H., XU J.: Semantic matching in search. *Foundations and Trends® in Information Retrieval* 7, 5 (2014), 343–469. URL: <http://dx.doi.org/10.1561/15000000035>, doi:10.1561/15000000035. 1
- [Roc89] ROCHEMONT M.: Topic Islands and the Subjacency Parameter. *Canadian Journal of Linguistics/Revue canadienne de linguistique* 34, 2 (1989), 145–170. doi:10.1017/S0008413100013293. 2
- [SAH07] STEINER C. M., ALBERT D., HELLER J.: Concept mapping as a means to build e-learning. *Advanced principles of effective e-learning* (2007), 59–111. 2
- [SHM14] SMITH A., HAWES T., MYERS M.: Hiérarchie: Visualization for Hierarchical Topic Models. In *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces* (June 2014), Association for Computational Linguistics, pp. 71–78. doi:10.3115/v1/W14-3111. 2
- [SI10] STEELE J., ILHINSKY N.: *Beautiful visualization: Looking at data through the eyes of experts*. "O'Reilly Media, Inc.", 2010. 2
- [SS14] SIEVERT C., SHIRLEY K.: LDavis: A method for visualizing and interpreting topics. In *Proceedings of the workshop on interactive language learning, visualization, and interfaces* (2014), pp. 63–70. doi:10.13140/2.1.1394.3043. 2
- [SZ00] STASKO J. T., ZHANG E.: Focus+context display and navigation techniques for enhancing radial, space-filling hierarchy visualizations. In *IEEE Symposium on Information Visualization 2000 (INFOVIS'00), Salt Lake City, Utah, USA, October 9-10, 2000* (2000), Mackinlay J. D., Roth S. F., Keim D. A., (Eds.), IEEE Computer Society, pp. 57–65. doi:10.1109/INFVIS.2000.885091. 3
- [TGM*14] TOMINSKI C., GLADISCH S., KISTER U., DACHSELT R., SCHUMANN H.: A Survey on Interactive Lenses in Visualization. In *EuroVis - STARs* (2014), The Eurographics Association. doi:10.2312/eurovisstar.20141172. 4, 5
- [TGM*17] TOMINSKI C., GLADISCH S., KISTER U., DACHSELT R., SCHUMANN H.: Interactive Lenses for Visualization: An Extended Survey. *Computer Graphics Forum* 36, 6 (2017), 173–200. doi:10.1111/cgfm.12871. 4, 5
- [TH09] THAI V., HANDSCHUH S.: Reordered tilebars for visual text exploration. In *2009 IEEE Symposium on Visual Analytics Science and Technology* (2009), pp. 225–226. doi:10.1109/VAST.2009.5333436. 3
- [TJBB06] TEH Y., JORDAN M., BEAL M., BLEI D.: Hierarchical Dirichlet Processes. *Machine Learning* (12 2006), 1–30. doi:10.1198/01621450600000302. 4
- [TY19] TANG Z., YANG G. H.: DeepTileBars: Visualizing Term Distribution for Neural Information Retrieval. In *Proceedings of the AAAI Conference on Artificial Intelligence* (2019), AAAI Press, pp. 289–296. doi:10.1609/AAAI.V33I01.3301289. 3