Secondary Publication



Liebers, Carina; Agarwal, Shivam; Beck, Fabian

CohExplore : Visually Supporting Students in Exploring Text Cohesion

Date of secondary publication: 25.01.2024 Version of Record (Published Version), Conferenceobject Persistent identifier: urn:nbn:de:bvb:473-irb-930514

Primary publication

Liebers, Carina; Agarwal, Shivam; Beck, Fabian (2023): "CohExplore : Visually Supporting Students in Exploring Text Cohesion". In: EuroVis Posters 2023, The Eurographics Association, doi: 10.2312/evp.20231058.

Legal Notice

This work is protected by copyright and/or the indication of a licence. You are free to use this work in any way permitted by the copyright and/or the licence that applies to your usage. For other uses, you must obtain permission from the rights-holder(s).

This document is made available under a Creative Commons license.



The license information is available online: https://creativecommons.org/licenses/by/4.0/legalcode

Poster

CohExplore: Visually Supporting Students in Exploring Text Cohesion

C. Liebers^{†1}, S. Agarwal², and F. Beck² ¹University of Duisburg-Essen, Germany ²University of Bamberg, Germany Document Overview R Text Legend D С B_1 Ba Is7 Also, the zooming adds or removes events in regards to their roduction Cohesio calculated relevance A cohesive text allows readers to follow the ideas and events shows the algorithm's process of aligning the elements to avoid clutter Overview: Visu.. described. We integrate three Movement Behavio cohesion metrics: semanti Is9 Movement is a change of the spatial dimension over time nilarity, connectives, and co-160 It occurs as part of an event or event chain, which can be aligned to reference cohesion. specific persons or groups. Visualizations . ——Biograp.. Trajectories Words, which often appear In Trajectories display the continuous movement behavior of an entity over together, are grouped in one topic. We color-code the topic space and time Mostly, they are visualized by three - dimensional Space - Time Cubes categories mantic Simila 163 displays the movement behavior during a trip with a Space - Time antic similarity betwee Relations Cube, where its path and stations are highlighted text segments are shown by Furthermore, the path can be enriched with qualitative and quantitative lines and arcs. High opacity information by coloring or varying the width indicates a high similarity. 65 It **also** can be enhanced by linking multimedia objects, whose details are displayed on demand 166 Space - Time Cubes are **also** able to display multiple trajectories Connective words signal a relation between parts of text. ¹⁶⁷ Kraak et al. e.g., because and how -Analysis .. • 168 show this by comparing the Russian and French troops movements in They are encoded with **bold** 1812 within a single Space - time Cube () text Discontinuous Movement Data Co-references Co-references (in ou ⁶⁹ In the context of movement <u>data</u>, **often** space **and** time are only mpact on HifiBo available at discrete points in a person's history implementation) are word ¹⁷⁰ For this discontinuous data, trajectories may not be the best repetitions in consecutive sentences. They are shown by visualization possibility Words of Topic 5 underlining the first occurre A person's location at an event can, for exa while its repetition has a grey time space historical cube maps, as previously introduced in (see). Conclusion figures cubes dim backaround Similar processes, which have a temporal ar therefore show movement behavior, can be visualizations А Topics in the Text 4 nodesedgeseve

Figure 1: *CohExplore shows (A) a colored bar for detected topics, (B) the document structure with similarity lines between (B*₁*) the text headings and (B*₂*) paragraphs, (C) the raw text with decorations and similarity arcs, and (D) an interactive legend.*

Abstract

A cohesive text allows readers to follow the described ideas and events. Exploring cohesion in text might aid students enhancing their academic writing. We introduce CohExplore, which promotes exploring and reflecting on cohesion of a given text by visualizing computed cohesion-related metrics on an overview and detailed level. Detected topics are color-coded, semantic similarity is shown via lines, while connectives and co-references in a paragraph are encoded using text decoration. Demonstrating the system, we share insights about a student-authored text.

1. Introduction

Students with limited experience and skills in writing face challenges in authoring a text that is easy to follow for the readersthe text might lack cohesion. Text cohesion is a property of the text that involves features that guide readers in interpreting substantial ideas [GMLC04, GMK11, GM11]. Topics reflect the overall text structure and semantics. Semantic similarity, based on topic detection, is significantly higher for high-cohesion than lowcohesion texts [MLMG10] and can be calculated using Latent Se-

© 2023 The Authors.



[†] e-mail: carina.liebers@uni-due.de

Proceedings published by Eurographics - The European Association for Computer Graphics. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

- **e** 263 Historical⁵ items¹ have similar⁷ complex² data² dimensions² as biographies⁶, as only the
- viewpoint is switched to objects⁷ [instead of persons². It reaches from combining³ maps¹ and timelines⁷ to specialized⁶ Space⁵- Time⁵ Cubes⁵ 264
- As before , interactions⁴ like⁴ linking⁴, filtering⁴, and selecting⁴ and showing¹ details⁴ on 265 nd⁴ plau⁴ an important role

Figure 2: Selecting a paragraph shows the topic words in their respective color. Hovering the connectives tile in the legend highlights connectives through borders.

mantic Analysis (LSA), word2vec, and Latent Dirichlet Allocation (LDA) [CKD19]. Referential cohesion [GMLC04], the references between words, contributes on a lower abstraction level. Further, cohesive texts incorporate more co-references, where a noun or pronoun refers back to another element in the text [MLMG10]. Connectives, such as "because", "while", or "however", improve comprehension of a text by contributing to local cohesion [KPMS19].

2. Visualization Approach

We present CohExplore (Fig. 1), an approach that visualizes cohesion-related metrics to aid reflection. It consists of an overview showing similarities of chapters and paragraphs and a reading view. It does not judge the text quality or provide scores but invites exploring text features that contribute to its cohesion. CohExplore aims to support analyzing texts by visualizing: (1) semantic text similarity, (2) connectives, and (3) co-references, covering global (1) and local (2, 3) features of text cohesion.

Regarding global cohesion, we use Latent Semantic Analysis (LSA) [LMDK07, Dos26] with a computed topic number [Nav09] to identify topics in a text by grouping frequently co-occurring words. CohExplore leverages these topics to compute the semantic similarity between text segments: pairs of sentences, paragraphs, and chapters. We compute the cosine similarity of their vectors, containing their impact values per topic. The results range between 0 (lowest) and 1 (highest).

Topic Bars The discovered topics are color-coded consistently and displayed at the bottom of the interface as a horizontal stacked bar (Fig. 1A). Their width indicates the topic's share in the text. Hovering reveals a list of words, ordered by their impact value for their topic. The font size encodes the impact value.

Document Overview The Document Overview on the left (Fig. 1B) is split into two columns and was inspired by VarifocalReader [KJW*14]. The headings column (Fig. 1B1) shows the hierarchical structure of the text via indentation. In the paragraphs column (Fig. 1B₂), colored stacked bars show the topic distribution of paragraphs. Users can scroll or click on headings, or drag a frame to navigate the document. The other columns follow. Additional lines depict semantic similarity among chapters, subchapters, and paragraphs, to support analyzing if similarities are expected or desired, ordered hierarchically to minimize overlap. Their opacity encodes their similarity.

Text Cohesion in Paragraphs To show aspects of local cohesion, the text panel (Fig. 1C, Fig. 2) highlights cohesion between sentences. Pie charts 🌖 show topic distribution of sentences to spot unusual patterns like multiple discussed topics. Arcs depict the semantic similarity within a paragraph, their opacity encodes the similarity value. We chose slightly different visualizations for topic distribution and semantic similarity to visually discern them from the overview, which has more text and topic words. Text decorations display local features: co-references, connectives, and a word's topic. Co-references, pointing to the previous or following sentence, are detected using word lemmas (stem), and are underlined in the first and given a grey background on subsequent occurrences. Commonly used connectives, indicating relations between arguments, are emphasized in **bold** to not interfere with the coreferences encoding. When selecting a paragraph, words are colored and marked with superscript numbers of their topic (Fig. 2).

Legend and Interactions The legend on the right (Fig. 1D) explains cohesion features and visual encodings. Hovering highlights the feature, like emphasizing connectives with borders (Fig. 2). Clicking a bar or a pie chart reveals links connecting the corresponding segment, while hiding unrelated lines and arcs. A slider enables setting a minimal threshold for similarity links.

3. Application Example

We explore a seminar report's cohesion (attached as supplement), written by one of the co-authors two years ago as a student. The Overview Panel (Fig. 1B) shows an imbalanced structure, where only one chapter, Visualizations of Historical Figures and Events, exhibits subchapters, contributing to over half of the text. Although it contains the report's main parts, its subchapters show an imbalance of similarity lines. Analyzing the high similarity of Movement Behavior and Analysis of Historical Items, we observe common themes in topic one and five
. Their word clouds of the topic bar (Fig. 1A) indicate overlap of visualization techniques, temporal dimension, and space-time cubes. Hence, re-thinking the structure could be promising. Further, the Conclusion section might be suboptimal: While its paragraphs exhibit similarity, the sentence similarity varies a lot. For instance, only the first two sentences have a similarity of >0.2 in the shown paragraph (Fig. 2), indicating a flow shift of discussed topics. Varying pie chart colors indicate abrupt topic shifts (l. 263 🌏, l. 264 🌎, and l. 265 🌒), and the absence of co-references might make the section harder to follow.

4. Conclusion and Future Work

We introduced a visualization approach that enables exploring text cohesion. It uses different cohesion metrics such as semantic similarity, connectives, and co-references. Future work includes adding co-reference cohesion metrics [MLMG10]; incorporating measures such as syntactic complexity, readability, and style word usage in its visualizations [GMLC04]; and applying alternative topic detection methods [ASFS18, BZSA18] to improve the topic quality, reduce their overlap, and refine the visualized text cohesion.

Acknowledgments

We thank Andrea Horbach and Torsten Zesch for valuable feedback on the approach.

References

- [ASFS18] ALTSZYLER E., SIGMAN M., FERNÁNDEZ SLEZAK D.: Corpus specificity in LSA and word2vec: The role of out-of-domain documents. In Proceedings of The Third Workshop on Representation Learning for NLP (2018), pp. 1–10. doi:10.18653/v1/w18-3001.2
- [BZSA18] BERNARD J., ZEPPELZAUER M., SEDLMAIR M., AIGNER W.: VIAL: A unified process for visual interactive labeling. *The Visual Computer: International Journal of Computer Graphics* 34, 9 (2018), 1189–1207. doi:10.1007/s00371-018-1500-3.2
- [CKD19] CROSSLEY S. A., KYLE K., DASCALU M.: The Tool for the Automatic Analysis of Cohesion 2.0: Integrating semantic similarity and text overlap. *Behavior Research Methods* 51, 1 (2019), 14–27. doi: 10.3758/s13428-018-1142-4.2
- [Dos26] DOSHI S.: Latent Semantic Analysis Deduce the hidden topic from the document, 2020-02-26. Accessed: March 2022. URL: https://towardsdatascience.com/latentsemantic-analysis-deduce-the-hidden-topic-fromthe-document-f360e8c0614b. 2
- [GM11] GRAESSER A. C., MCNAMARA D. S.: Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science* 3, 2 (2011), 371–398. doi:10.1111/j.1756-8765.2010.01081.x. 1
- [GMK11] GRAESSER A. C., MCNAMARA D. S., KULIKOWICH J. M.: Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher* 40, 5 (2011), 223–234. doi:10.3102/0013189X11413260.1
- [GMLC04] GRAESSER A. C., MCNAMARA D. S., LOUWERSE M. M., CAI Z.: Coh-Metrix: Analysis of text on cohesion and language. *Behav*ior Research Methods, Instruments, & Computers 36, 2 (2004), 193–202. doi:10.3758/BF03195564.1,2
- [KJW*14] KOCH S., JOHN M., WÖRNER M., MÜLLER A., ERTL T.: VarifocalReader – in-depth visual analysis of large text documents. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 1723–1732. doi:10.1109/TVCG.2014.2346677. 2
- [KPMS19] KLEIJN S., PANDER MAAT H. L., SANDERS T. J.: Comprehension effects of connectives across texts, readers, and coherence relations. *Discourse Processes* 56, 5-6 (2019), 447–464. doi:10.1080/ 0163853X.2019.1605257. 2
- [LMDK07] LANDAUER T. K., MCNAMARA D. S., DENNIS S., KINTSCH W. (Eds.): Handbook of Latent Semantic Analysis, 1 ed. Psychology Press, 2007. 2
- [MLMG10] MCNAMARA D. S., LOUWERSE M. M., MCCARTHY P. M., GRAESSER A. C.: Coh-Metrix: Capturing linguistic features of cohesion. *Discourse Processes* 47, 4 (2010), 292–330. doi:10.1080/ 01638530902959943. 1, 2
- [Nav09] NAVLANI A.: Latent Semantic Analysis using Python. determining optimum number of topics, 2018-10-09. Accessed: March 2022. URL: https://www.datacamp.com/community/tutorials/ discovering-hidden-topics-python. 2