EG UK Computer Graphics & Visual Computing (2023)
D. Hunter and P. Vangorp (Editors)

Less is more: Focused Design and Problem Framing in Visualisation – Developing the ColloCaid Collocation Editor

J. C. Roberts† [1], P. W. S. Butcher‡ [1], G. Rees§ [2,3], R. Lew¶ [4], N. Sharma‖ [1,5] and A. Frankenberg-Garcia** [2]

[1]Bangor University, UK
[2]University of Surrey, UK
[3]Universitat Rovira i Virgili, Spain
[4]Adam Mickiewicz University, Poland
[5]Open University, UK

Abstract

One of the challenges when developing a visualisation tool, especially at the start of a research project, is to amalgamate numerous requirements and various possibilities and decide what to create. With software development, it is too easy to incorporate all ideas, but quickly the tool becomes unusable, with feature overload. We reflect on designing and building the ColloCaid collocation visualisation editor, especially our conceptual focus on simplicity. We were inspired by Hemingway's iceberg theory of deliberate omission, to help frame the visualisation challenge and achieve clarity and focused design. The ColloCaid tool enables people to discover collocations, to help people improve vocabulary and fluency as they write. It was developed by a multidisciplinary team of applied linguists, lexicographers, human-computer interaction and visualisation experts. We promote focused design and problem solving, in visualisation, highlight concepts, including parti, design essence, and simplification. We provide a collection of insights that hold potential to evolve into a structured set of design guidelines, offering valuable direction to researchers.

1. Introduction

We have been developing, as a multidisciplinary team of linguists, lexicographers, human-computer interaction and visualisation experts, the ColloCaid collocation visualisation editor [FRL*19; SRB*19; RBL*20]. This online platform provides an environment for writing and exploration, offering real-time suggestions of academic English. Our goal is to help individuals enhance the lexicon and fluency of their written content, while unveiling potential collocations that they may use in the future. In this work, we explain focused design and problem framing strategies when developing the ColloCaid collocation editor and visualisation application.

When starting a visualisation project – such as our digital humanities visualisation project, ColloCaid – there are many challenges; not only with difficulties over handling extremely large data sets, but different (often conflicting) requirements and goals of the project. In fact, researchers should recognise the importance of dedicating effort to comprehend, analyse and handle data, because it forms the foundation of the project. Furthermore, it takes much time to thoughtfully design new visualisation solutions, integrate user feedback throughout the project, and decide on the best visual designs for the use-case, application users, deployment setting, that can display the data effectively and efficiently. In addition to determining the project's outcome and vision, the process must involve honing the focus of the project, in order to establish well-defined objectives for the work. Consequently, it is necessary to have a clear design philosophy, carefully understand user needs, gather feedback, and meticulously understand the impact of different (potential) features.

2. ColloCaid development & visualisation design philosophy

Our vision was to develop a tool that will help people to write better by choosing strong collocations – "lexical items occurring ... with a greater frequency than the law of averages" [Kri87]. For example, when an individual seeks a new computer, they desire a *powerful computer*, rather than referring to it as a **strong computer*. Similarly in the context of visualisation, the phrase *bar chart* is more common than **bar plot* [RAB*19]. ColloCaid is an online text editor that extends the open-source TinyMCE editor, incorporates dynamic text suggestions, in-situ visualisations, and is underpinned by an expert curated a dataset of over 30,000 collocations.

Our design philosophy drew inspiration from Hemingway, especially his iceberg theory: purposeful omission can greatly enhance the overall clarity [Smi83]. The visible part of the iceberg, above the waterline, represents only a fraction of the entirety. Similar ideas can be applied to a data visualisation project. For example,

† j.c.roberts@bangor.ac.uk, ‡ p.butcher@bangor.ac.uk,
§ geraintpaul.rees@urv.cat, ¶ rlew@amu.edu.pl,
‖ nirwan.sharma@open.ac.uk, ** a.frankenberg-garcia@surrey.ac.uk

© 2023 The Authors.
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.

DOI: 10.2312/cgvc.20231191

delivered by EUROGRAPHICS DIGITAL LIBRARY www.eg.org diglib.eg.org

frequently there exists an excess of data, needing its reduction, processing or selection. Project development time is limited, requiring focused effort. Each objective takes time to implement, therefore developers must select the most important aspects to focus on. In ColloCaid, we made intentional choices to reduce the quantity of the data, specify examples, conscious selections over visualisation design, and thoughtful choices on functionality of the interface. In this paper we focus on design goals, data, and interface design.

3. 'Less is more' and examples from ColloCaid

Understanding **design goals** is important, especially at the start of the project. This can be achieved by refining the problem *frame*, and deeply considering every aspect of the challenge to determine a list of priorities [Ent93]. Framing involves creating a framework that shapes the approach. We had a clear design philosophy: to help users write better and use appropriate collocations. This clear *design essence*, in architecture, is known as the *parti* [Fre07]. Indeed, to design the visualisations we used the Five Design-Sheet (FdS) methodology, to sketch many alternative (low-fidelity [Ret94]) visualisation concepts, which incorporates the design parti as a fundamental principle [RHR16]. Because various perspectives (frames) can be constructed for a single challenge, each emphasising distinct facets, it is crucial to recognise that each perspective significantly impacts, and shapes the ultimate form of, the end result. Subsequently, frames can change, as knowledge and experience grow, and are 'elaborated' over time [KMH06; PC05]. The idea of shifting design objectives aligns effectively with the process of *prototyping*. In ColloCaid, we implemented several prototypes, tested each with real users, improved the version, and so on. In fact, we started with a Powerpoint implementation — a *pretotype* [Sav19] (a quickly developed solution to test the concept) — which we demonstrated at conferences, and for which gained useful and constructive feedback.

One significant challenge inherent in any project involving **data** utilisation and visualisation is the organisation and management of reference data. High quantities of data quickly become unwieldy, are slow to analyse, create unfocused tools and confusing results. Demonstration data must be focused, to help answer the specific question. We employed various strategies to focus the project goals. We chose Academic English as it is the lingua franca of the academic world. This choice fitted well with us being academics in fields involving lexicography, linguistics, and computing, because it gave us access to a broad set of writers in the English language, and provided us with appropriate participants for our evaluation studies. We reduced the data further by including only frequent words (Zipf distribution [Fra20]) and specifically used three vocabulary lists as detailed by Frankenberg-Garcia et al. [FLR*19]. Using SketchEngine [KBB*14; KK17], logDice [KBB*14] and frequency statistics from the corpora of published academic writing (e.g. OCAE [WSS19], PICAE [ADKT11; AC13]) we curated collocation association scores, example sentences, storing collocate nodes as headwords (which makes the database smaller), and expanding the words to their full lexical set when required [RBL*20].

The notion of uncomplicated **interfaces** is deeply ingrained in human-computer interaction. One of Jakob Nielsen's principles is toward "aesthetic and minimalist design"; Shneiderman recommends maintaining simplicity in design [SP10], while Baldonado et al. [WWK00] emphasise the need for economical view use (view parsimony) in multiple coordinated view systems [Rob07], as supported by quantification analysis [CZL*20]. Rather than using hundreds of visualisations, we focused on a few specific well-crafted visual depictions, and on text visualisation [KK15; CC16; LWC*18; AL19]. However, crafting effective interfaces is not a matter of making things simple – a point that Donald Norman underlined [Nor10] – instead, designers should ensure "complexity is tamed". For example, it would be confusing to display hundreds of examples. We decided to show the eight most relevant examples, motivated by research in cognitive psychology [Mil56].

4. Discussion

Design simplification and the concept of 'less is more' hold substantial power. They prompt us, as designers of tools, to consider what is essential. However, comprehending which elements are necessary, or deserving of emphasis, is not always straightforward. With ColloCaid, our decisions were helped through early prototypes and ongoing user involvement throughout the development; from early prototypes, PowerPoint pretype, to visualisation sketches with the FdS [RHR17], ongoing evaluation by workshop participants and language learners. Importantly, we had a strong guiding essence, of developing a user-focused tool that helped users improve their written English. Indeed, by the end of 2022 we hit over 10,000 ColloCaid users. Choosing to omit certain elements, in favour of highlighting others, and simplifying and refining individual facets, has not necessarily meant less work. On the one hand, we invested more time and energy in specific project components, such as getting the interface right, testing it with users, and so forth. On the other, we saved time, by focusing on specific data, and the development of bespoke and specific visualisations.

Finally, we emphasise the importance of understanding the concept *essence* (the parti), in our case, an 'environment for writing and exploration, offering real-time suggestions of academic English using collocations'. This was important, throughout the project, but especially when designing the visualisations. Furthermore, on any team, when determining to include a function, someone can quickly ask 'does it fit with the parti?' However, this is not necessarily a quick action, as it often requires developers to place additional effort and time, make redesigns, re-implement functions that users have suggested can be improved, and so on. A visualisation project can be likened to a swan gracefully gliding on water. At a glance, users perceive a streamlined, refined system. However, beneath the exterior lies a substantial amount of dedicated work, careful strategising, data processing and manipulation, choice of effective colour schemes, all supported by a foundation of scientific rigour. Data experts and visualisation designers must make numerous decisions, on data storage, selection and processing, data mapping, colour schemes, interface controls and so forth. To guarantee the effective communication of the intended message through visualisation and the creation of a dependable and user-friendly interface.

Acknowledgements

ColloCaid was funded by AHRC grant AH/P003508/1. We acknowledge every user and their feedback.

© 2023 The Authors.
Proceedings published by Eurographics - The European Association for Computer Graphics.

References

[AC13] ACKERMANN, KIRSTEN and CHEN, YU-HUA. "Developing the Academic Collocation List (ACL)–A corpus-driven and expert-judged approach". *Journal of English for Academic Purposes* 12.4 (2013), 235–247. DOI: 10.1016/j.jeap.2013.08.002.

[ADKT11] ACKERMANN, KIRSTEN, DE JONG, JHAL, KILGARRIFF, ADAM, and TUGWELL, DAVID. "The Pearson international corpus of academic English (PICAE)". *Proceedings of Corpus Linguistics*. 2011.

[AL19] ALHARBI, MOHAMMAD and LARAMEE, ROBERT. "SoS TextVis: An Extended Survey of Surveys on Text Visualization". *Computers* 8 (Feb. 2019), 17. DOI: 10.3390/computers8010017.

[CC16] CAO, NAN and CUI, WEIWEI. *Introduction to Text Visualization*. 1st. Atlantis Publishing Corporation, 2016. ISBN: 9462391858. DOI: 10.5555/3029322.

[CZL*20] CHEN, XI, ZENG, WEI, LIN, YANNA, et al. "Composition and configuration patterns in multiple-view visualizations". *IEEE Transactions on Visualization and Computer Graphics* 27.2 (2020), 1514–1524. DOI: 10.1109/TVCG.2020.3030338.

[Ent93] ENTMAN, ROBERT M. "Framing: Toward Clarification of a Fractured Paradigm". *Journal of Communication* 43.4 (1993), 51–58. DOI: 10.1111/j.1460-2466.1993.tb01304.x.

[FLR*19] FRANKENBERG-GARCIA, ANA, LEW, ROBERT, ROBERTS, JONATHAN C, et al. "Developing a writing assistant to help EAP writers with collocations in real time". *ReCALL* 31.1 (2019), 23–39. DOI: 10.1017/S0958344018000150.

[Fra20] FRANKENBERG-GARCIA, ANA. "Combining user needs, lexicographic data and digital writing environments". *Language Teaching* 53.1 (2020), 29–43. DOI: 10.1017/S0261444818000277.

[Fre07] FREDERICK, MATTHEW. *101 Things I Learned in Architecture School*. MIT Press, Sept. 2007.

[FRL*19] FRANKENBERG-GARCIA, ANA, REES, GERAINT, LEW, ROBERT, et al. "ColloCaid: a tool to help academic English writers find the words they need". *CALL and complexity – short papers from EUROCALL 2019*. Ed. by MEUNIER, FANNY, VAN DE VYVER, JULIE, BRADLEY, LINDA, and THOUËSNY, SYLVIE. Research-publishing.net, 2019. ISBN: 978-3-03868-106-9. DOI: 0.14705/rpnet.2019.38.1000.

[KBB*14] KILGARRIFF, ADAM, BAISA, VÍT, BUŠTA, JAN, et al. "The Sketch Engine: ten years on". *Lexicography* (2014), 7–36. DOI: 10.1007/s40607-014-0009-9.

[KK15] KUCHER, KOSTIANTYN and KERREN, ANDREAS. "Text visualization techniques: Taxonomy, visual survey, and community insights". *PacificVis*. IEEE. 2015, 117–121. DOI: 10.1109/PACIFICVIS.2015.7156366.

[KK17] KUNILOVSKAYA, MARIA and KOVIAZINA, MARINA. "Sketch Engine: A Toolbox for Linguistic Discovery". *Journal of Linguistics* 68.3 (2017), 503–507. DOI: 10.2478/jazcas-2018-0006.

[KMH06] KLEIN, G., MOON, B., and HOFFMAN, R.R. "Making Sense of Sensemaking 2: A Macrocognitive Model". *IEEE Intelligent Sys.* 21.5 (Sept. 2006), 88–92. ISSN: 1541-1672. DOI: 10.1109/MIS.2006.100.

[Kri87] KRISHNAMURTHY, RAMESH. "The process of compilation". *Looking up: An account of the COBUILD project in lexical computing* (1987), 62–85.

[LWC*18] LIU, SHIXIA, WANG, XITING, COLLINS, CHRISTOPHER, et al. "Bridging text visualization and mining: A task-driven survey". *IEEE TVCG* 25.7 (2018), 2482–2504. DOI: 10.1109/TVCG.2018.2834341.

[Mil56] MILLER, GEORGE A. "The magical number seven, plus or minus two: Some limits on our capacity for processing information". *Psychological review* 63.2 (1956), 81.

[Nor10] NORMAN, DONALD A. *Living with Complexity*. Cambridge, MA: MIT Press, 2010. ISBN: 978-0-262-01486-1.

[PC05] PIROLLI, P. and CARD, S. "The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis". *Proc. Intelligence Analysis*. Vol. 5. 2005, 2–4.

[RAB*19] ROBERTS, JONATHAN C., AL-MANEEA, HAYDER MAHDI ABDULLAH, BUTCHER, PETER W. S., et al. "Multiple Views: different meanings and collocated words". *Computer Graphics Forum* (Mar. 2019). ISSN: 0167-7055. DOI: 10.1111/cgf.13673.

[RBL*20] ROBERTS, JONATHAN C., BUTCHER, PETER W. S., LEW, ROBERT, et al. "Visualising Collocation for Close Writing". *EuroVis 2020 - Short Papers*. Ed. by KERREN, ANDREAS, GARTH, CHRISTOPH, and MARAI, G. ELISABETA. The Eurographics Association, 2020. ISBN: 978-3-03868-106-9. DOI: 10.2312/evs.20201069.

[Ret94] RETTIG, MARC. "Prototyping for Tiny Fingers". *Commun. ACM* 37.4 (Apr. 1994), 21–27. ISSN: 0001-0782. DOI: 10.1145/175276.175288.

[RHR16] ROBERTS, J. C., HEADLEAND, C., and RITSOS, P. D. "Sketching Designs Using the Five Design-Sheet Methodology". *Transactions on Visualisation and Computer Graphics* 22.1 (2016), 419–428. DOI: 10.1109/TVCG.2015.2467271.

[RHR17] ROBERTS, JONATHAN C., HEADLEAND, CHRISTOPHER J., and RITSOS, PANAGIOTIS D. *Five Design-Sheets – Creative design and sketching in Computing and Visualization*. Springer, 2017. ISBN: 978-3-319-55626-0. DOI: 10.1007/978-3-319-55627-7.

[Rob07] ROBERTS, JONATHAN C. "State of the Art: Coordinated & Multiple Views in Exploratory Visualization". *CMV2007*. Ed. by ANDRIENKO, GENNADY, ROBERTS, JONATHAN C., and WEAVER, CHRIS. 2007, 61–71. DOI: 10.1109/CMV.2007.20.

[Sav19] SAVOIA, ALBERTO. *The Right It – Why So Many Ideas Fail and How to Make Sure Yours Succeed*. HarperOne, 2019.

[Smi83] SMITH, PAUL. "Hemingway's Early Manuscripts: The Theory and Practice of Omission". *Journal of Modern Literature* 10.2 (1983), 268–288. ISSN: 0022281X, 15291464. DOI: 10.2307/3831126.

[SP10] SHNEIDERMAN, B. and PLAISANT, C. *Designing the User Interface: Strategies for Effective Human-computer Interaction*. Addison-Wesley, 2010. ISBN: 9780321601483.

[SRB*19] SHARMA, N., REES, G., BUTCHER, P., et al. "Co-creating an online learning environment to support academic writing: Lessons learnt in an interdisciplinary setting". *4th Workshop on Visualization for the Digital Humanities (VIS4DH)*. IEEE, Oct. 2019.

[WSS19] WANG, S, SOLADI É, C., and S ÉGUIER, R. "OCAE: Organization-Controlled Autoencoder for Unsupervised Speech Emotion Analysis". *2019 5th International Conference on Frontiers of Signal Processing (ICFSP)*. 2019, 72–76. DOI: 10.1109/ICFSP48124.2019.8938073.

[WWK00] WANG BALDONADO, MICHELLE Q., WOODRUFF, ALLISON, and KUCHINSKY, ALLAN. "Guidelines for Using Multiple Views in Information Visualization". *Proceedings of the Working Conference on Advanced Visual Interfaces*. AVI '00. Palermo, Italy: ACM, 2000, 110–119. ISBN: 1-58113-252-2. DOI: 10.1145/345513.345271.

© 2023 The Authors.
Proceedings published by Eurographics - The European Association for Computer Graphics.