



## UvA-DARE (Digital Academic Repository)

### [Review of: J. Guldi (2023) The Dangerous Art of Text Mining: A Methodology for Digital History]

Wevers, M.

**DOI**

[10.1353/tech.2024.a933123](https://doi.org/10.1353/tech.2024.a933123)

**Publication date**

2024

**Document Version**

Final published version

**Published in**

Technology and Culture

**License**

Article 25fa Dutch Copyright Act (<https://www.openaccess.nl/en/policies/open-access-in-dutch-copyright-law-taverne-amendment>)

[Link to publication](https://doi.org/10.1353/tech.2024.a933123)

**Citation for published version (APA):**

Wevers, M. (2024). [Review of: J. Guldi (2023) The Dangerous Art of Text Mining: A Methodology for Digital History]. *Technology and Culture*, 65(3), 1033-1035.  
<https://doi.org/10.1353/tech.2024.a933123>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

*UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)*

---

## The Dangerous Art of Text Mining: A Methodology for Digital History

By Jo Guldi. Cambridge: Cambridge University Press, 2023. Pp. 465.



In *The Dangerous Art of Text Mining*, historian Jo Guldi explores the application of text mining in historical research. Text mining, a method for quantitatively analyzing digitized text, is utilized by Guldi to examine British parliamentary records. The approach is portrayed as a dual-edged sword, embodying both art and hazard. Guldi characterizes text mining as an art that demands specialized expertise and flexible methodologies, aligning with historians' heuristic and hermeneutic techniques. At the same time, she warns of its dangers, such as the potential for algorithms to foster overgeneralizations, amplify biases in data, and yield conclusions that overlook historical complexities and the nuanced interpretations of past actors. Despite the book's ostensibly alarming title, Guldi's critique of text mining is constructive, underscoring its capacity to enhance historical research if used with discernment, thereby avoiding the pitfalls of technological solutionism that mar some big data and early digital humanities projects.

The book is structured into three main sections. The first, "Towards a Smarter Data Science," outlines Guldi's methodological approach, advocating for a seamless integration of data science with historical research's nuanced source criticism. The book adopts a somewhat antagonistic stance toward data science, portraying it as naive regarding data and algorithmic bias.

JULY  
2024  
VOL. 65

Guldi posits that historians are uniquely positioned to prevent the uncritical application of algorithms to historical data. While data science studies often overlook such biases (as pointed out by R. Benjamin, *Race after Technology*, 2019; C. O’Neil, *Weapons of Math Destruction*, 2016), similar issues are evident in digital humanities and digital history. Think, for example, of the frequent use in digital history of topic modeling algorithms as exploratory tools without inspecting the degree of robustness of the models. Historians could also benefit from work in areas such as explainable AI, where initiatives are being advanced to augment the transparency of AI models, ensuring that users not only trust but also comprehend the underlying mechanisms and rationale of the algorithms employed in their research (for example, Molnar, *Interpretable Machine Learning*, 2022). Concurrently, information retrieval scholars have developed metrics to evaluate search strategy efficacy. Such metrics deserve more attention in Guldi’s book.

Guldi presents the strategy of critical search, which relies on the critical use of multiple algorithms to guide inquiries into large collections of historical materials. “Critical” here refers to the awareness of possible biases introduced by the data and the algorithms at every step of the process. Her language effectively resonates with historians, potentially increasing the method’s adoption. However, a more integrated approach, combining data science methods and validation strategies, would have facilitated a more cohesive amalgamation of the two disciplines. The book adeptly discusses why prediction is a contentious and difficult concept for many historians. Given the contingency of history and the sparsity of data, making predictions is notoriously difficult. Nonetheless, by narrowing down modeling as mostly a predictive tool, Guldi overlooks how modeling could be beneficial for historians. A broader exploration of modeling’s other applications, such as guiding explanations and understanding data biases, would have been beneficial (for example, J. Epstein, *Why Model?*, 2008). The breadth in examples might have been balanced with a broader overview of methodological interventions, encompassing machine learning, sampling, and time series analysis. The focus on text mining is not necessarily a flaw, but many of the techniques are repeated throughout the book, making it somewhat repetitive. Generally, the book suffers from editing issues, including errors in footnotes, inconsistencies, and repetition of content.

The second part, the book’s strongest, presents case studies addressing temporal experience themes like memory, periodization, and event. Guldi adeptly uses historical theory as a conduit between text mining methods, data, and historical inquiry. She demonstrates how theories from scholars like Reinhardt Koselleck, William Sewell, and Astrid Erll can translate historical data analysis goals into specific methodological applications. As such, the book effectively argues for a stronger integration of theory and method in advancing (digital) history.

The final part speculates on digital history's future and its impact on history departments, emphasizing the need for collaboration and the hiring of digital historians within these departments. Others have made similar points concerning digital history (A. Crymble, *Technology and the Historian*, 2022). Guldi's book, however, stands out for its robust grounding in historical theory and its clear demonstration of how text mining can assist historians in exploring temporal experiences. Its focus on parliamentary data provides tangible examples, inspiring future scholars to further explore how digital methods can augment historical research.

MELVIN WEVERS

Melvin Wevers is assistant professor in digital history at the University of Amsterdam. His research interests include the modeling of cultural-historical processes, multimodality, and the role of events in public discourse. His most recent publication (with Alexandra Barancová and Nanne van Noord) is "Blind Dates: Examining the Expression of Temporality in Historical Photographs," in *Proceedings of the Computational Humanities Research Conference 2023* (CEUR Workshop Proceedings, 2023).

Citation: Wevers, Melvin. "Review of *The Dangerous Art of Text Mining: A Methodology for Digital History* by Jo Guldi." *Technology and Culture* 65, no. 3 (2024): 1033–35.

---