*DIVE-IN* **2024**, **4**(**2**): 185–188
https://doi.org/10.6092/issn.2785-3233/21613

DIVE-IN
AN INTERNATIONAL JOURNAL ON DIVERSITY & INCLUSION

# Felix Hamborg, *Revealing Media Bias in News Articles: NLP Techniques for Automated Frame Analysis*, Cham: Springer Nature, 2023, 238 pp., e-ISBN 978-3-031-17693-7.

**Reviewed by** Beatrice Fiumanò
*Alma Mater Studiorum – Università di Bologna*

Serving as a primary source of information on societal and political issues, news articles play the crucial role of shaping and disseminating narratives, influencing public perceptions of topics, events and social actors. Since the early days of the press and especially in contemporary society, characterised by an abundance of different news outlets, biased news coverage has posed a significant problem, holding the power to guide public opinion and affect collective decision-making.

Over the past decades, the study of media bias has been largely explored in the humanities and social sciences. More recently, this area of research has gained traction also in the field of computer science, particularly within the domain of Natural Language Processing (NLP). However, as Felix Hamborg notes, a significant gap persists between these two domains. Hamborg argues that while bias research in the social sciences has led to a comprehensive set of theories and methodologies, it is bound to remain limited in scope due to its reliance on traditional methods that are time-consuming and require significant manual efforts and expertise. In contrast, NLP approaches have the potential to improve scalability and overcome the barriers of manual studies by leveraging sophisticated models and techniques. Nevertheless, by neglecting well-established methods and resources from other domains, computer science researchers tend to address bias as a rather vague phenomenon, resulting in simpler and more superficial methodological approaches.

Hamborg's work *Revealing Media Bias in News Articles: NLP Techniques for Automated Frame Analysis* tackles this issue and bridges the gap between

decade-long research in the social sciences and advanced computer science techniques by proposing an interdisciplinary, automated approach to uncover bias in news articles. Informed by Entman's theory of political framing, Hamborg introduces a framework to extract and analyse frames in news media, exploring how different news sources assemble and promote narratives that shape our interpretation of events.

The book - a result of Hamborg's extensive research on media bias - is organised into seven chapters. After a brief introduction to research questions and gaps addressed by the work, the second chapter delves into the theoretical and methodological foundations of bias research. Drawing upon a well-established tradition in the social sciences, the chapter defines media bias and its effects, explores the process of news production and identifies nine bias forms that arise at each production stage, i.e. event selection, source selection, commission and omission of information, labelling and word choice, story placement, size allocation, picture selection, picture explanation and spin.

Building on this categorisation, Hamborg provides an in-depth review of manual and automated approaches addressing each bias form, extensively referencing prior work in both social sciences and computer science. Besides outlining methodological characteristics, strengths and limits of these approaches, the chapter also highlights different NLP techniques that can be leveraged to analyse each bias type. The section thus proves particularly helpful for those who are new to slanted media analysis, as well as for researchers seeking to adopt new methodologies and explore the possibilities of an automated approach.

To bridge the theoretical and literature framework outlined in the previous chapter with the detailed methodological structure illustrated in the following ones, Chapter 3 introduces person-oriented framing analysis (PFA), an innovative approach for the analysis of political frames and media bias forms. Integrating decade-long bias research in the social sciences with NLP techniques, this automated approach aims at uncovering how individual persons are portrayed across different news sources and communicating these findings to news consumers. After a brief (re)introduction to the concept of media bias and to bias detection methods, the chapter offers a clear overview of PFA, explaining its objectives and the three main stages of the analysis, i.e. preprocessing, target concept analysis and frame analysis. Attention is also given to the dataset construction phase, which is often a complex task when working with (multiple) online sources. In particular, Hamborg introduces *news-please*, a high-quality tool for online news crawling and information

extraction, designed for adapting to different news outlets. Combining three different state-of-the-art extractors, *news-please* outperforms them in terms of extraction quality and is openly available on GitHub.

Moving to target concept analysis (TCA) - the second stage of PFA - in Chapter 4 Hamborg explores two NLP techniques that may enable such analysis, namely event extraction and context-driven cross-document coreference resolution (CDCDCR). As for the first task, the author proposes a new open-source system for extracting an article's main event based on the journalistic 5W1H questions, i.e. who does what, when, where, why and how. The *Giveme5W1H* method, evaluated against a manually annotated dataset of news articles, achieves higher precision scores compared to the few similar approaches available in computer science literature and establishes itself as a valuable tool for various types of text mining analyses. As a downside, the tool is only able to detect a single, main event. The second viable technique for TCA is cross-document coreference resolution, which enables the identification of the same semantic concept across multiple texts. In particular, the CDCDCR method proposed by the author leverages a rule-based, multi-sieve approach to merge candidates and resolve coreferences, enabling more flexible and context-aware results.

The last component of PFA, namely frame analysis, is concerned with biases deriving from slanted portrayals of individuals. In this regard, Hamborg notes that devising an automated approach based on social sciences best practices, e.g. on the analysis of Entman's political frames is impractical due to the high degree of interpretation and manual effort involved. Instead, he suggests shifting the focus to frame polarity, exploring whether news coverage contains positive or negative assessments of target individuals. To achieve this, the proposed NLP technique is target-dependent sentiment classification (TSC). After testing the performance of three state-of-the-art TSC models on different datasets, Hamborg introduces a BERT model fine-tuned on political news articles, as well as a new manually annotated dataset for TSC applied to the news domain. One last notable contribution of Hamborg's work is the presentation of a prototype system for evaluating PFA's performance on real-world news and illustrating how insights derived from the analysis can be communicated to non-expert news consumers, enabling bias-aware news consumption.

To conclude, the book stands as a valuable resource for NLP researchers working on media bias analysis, offering methodological insights and new open-access tools for bias detection that are also suitable for various text mining

applications. As a downside, some of the proposed methods do not harness the full potential of computer science technologies nor reflect the latest advancements in NLP techniques. Additionally, the volume requires a solid understanding of computer science methodologies, making it less accessible for readers without a technical background.

Overall, the book's most significant contribution lies in redirecting attention to the rich array of resources, methods, and theories developed through decade-long bias research, countering the tendency in computer science to overlook foundational work in the social sciences. At the same time, the volume successfully underscores the importance of coupling these methodologies with the advantages of NLP technologies, illustrating how an interdisciplinary approach can transcend the limitations of more traditional methods.