Text and Data Mining: The theory and practice of using TDM for scholarship in the humanities

Paul Verhaar

List price **£64.95**
Product Details

Format: Paperback
ISBN: 9781783304196
Published: 15th Jul 2025
Publisher: Facet Publishing
Dimensions: 234mm x 156mmmm

Description

This book offers a broad and accessible introduction to research based on text and data mining (TDM), focusing specifically on the ways in which TDM has been applied within the humanities.

TDM is a collection of computational and algorithmic methods that enable researchers to extract information from large collections of machine-readable texts. As is the case in many other academic disciplines, a growing number of scholars in the humanities are trying to harness the numerous innovate possibilities that can emanate from TDM. While there is a clear uptake of TDM within the humanities, it is relatively difficult for scholars who are new to the field to find books which explain in understandable terms what TDM actually entails. This book offers a accessible and comprehensive overview of the methodology and the theory of TDM, concentrating on applications within the humanities.

The book firstly discusses TDM on a practical level. It defines central terms and concepts, and it characterises the tools and the algorithms which have been used most commonly. The purposes and the contexts of these techniques are clarified using a generic description of the workflow that is followed during research projects. The book additionally contains chapters about the various ways in which academic libraries are organising their support for TDM, and about some of the obstacles posed by legislation in the field of intellectual property rights.

Based on a thorough scrutiny of existing critical debates about computer-assisted textual research, this book also characterises the possibilities and the limitations of TDM on a more conceptual level. The main objective of the book is to develop a theoretical framework which can help to clarify aspects of research based on TDM and to describe the general ways in which TDM may affect and transform traditional scholarship in the humanities. Supported by international case studies, coverage in the book includes:

- pre-processing operations
- data analysis

- obstacles posed by Intellectual Property Rights
- text and data mining on a conceptual level
- visualisation
- tools criticism
- library support for text and data mining.

The book will be essential reading for humanities scholars interested in getting started in TDM and those who aim to develop their understanding of TDM on a more theoretical level. It will also be a must-read for academic librarians and information professionals who seek to develop services to support scholarship based on TDM and students interested in digital humanities.

Author

Peter Verhaar is a lecturer at the MA programme Book and Digital Media Studies at Leiden University. He has taught several courses about text encoding, database theory, the digital humanities and media theory, among many other topics. In the fall of 2016, he defended his PhD dissertation, *Affordances and Limitations of Algorithmic Criticism*, which focuses on applications of Text and Data Mining within the field of literary criticism. This dissertation was awarded the 2016 Victorine van Schaik Prize, for the best publication in the field of Library and Information Science.