Text Mining for Information Professionals

Manika Lamba • Margam Madhusudhan

Text Mining for Information Professionals

An Uncharted Territory

Springer

Manika Lamba
Library and Information Science
University of Delhi
Delhi, India

Margam Madhusudhan
Library and Information Science
University of Delhi
Delhi, India

ISBN 978-3-030-85084-5 ISBN 978-3-030-85085-2 (eBook)
https://doi.org/10.1007/978-3-030-85085-2

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To my grandparents and brother for their endless support and constant encouragement
– Manika

Preface

Machine learning and artificial intelligence are the futuristic approaches dominating all disciplines currently. Data analytics, data mining, and data science are some of the significant sub-domains with a strong market of research and jobs at the moment in the world. With a long leap from paper to digitization, the burden on libraries and librarians has increased to manage, organize, and generate knowledge from such a massive amount of data stored in their repositories/databases/websites. As libraries deal with a higher percentage of textual data daily, this book focuses primarily on textual data and presents various text mining approaches through a new lens. Text mining is a very efficient, fast, and effective way of managing and extracting knowledge from existing data stored in the archives of libraries. This book will make every library and information professional competent to use text mining in their daily life and get the best out of it by serving their patrons, researchers, faculty, or scientists with new services. Text mining techniques can be applied to any library type, be it a school, university, or special library by the librarians. It will help to provide the *right information* to the *right user* at the *right time* by providing services like recommendation services, current awareness services, or selective dissemination services to its users.

From understanding different types and forms of data to case studies covering primary research showing the application of each text mining approach on data retrieved from various resources, this book will be a must-read for all library professionals interested in text mining and its application in libraries. Additionally, this book will be helpful to archivists, digital curators, or any other humanities and social science professionals who want to understand the basic theory behind text data, text mining, and various tools and techniques available to solve and visualize their research problems.

Key points of the book

1. Contains 14 demonstrative step-by-step case studies which show how to conduct 8 different text mining and visualization approaches on 9 distinct data type sources
2. Provides case studies demonstrating the use of five open-source software for both non-programmers and programmers

3. Reproduces six case studies using R programming in the cloud without having to install any software
4. Story section presenting 17 real-life experiences of the application of text mining methods and tools by 24 librarians/researchers/faculty/publishers
5. Elucidates 19 open-source text mining and visualization tools with their advantages and disadvantages
6. Illustrates various use cases that show how text mining strategies have been used in different ways in libraries across the globe

The book contains *11 chapters*, *14 case studies* showing 8 different text mining and visualization approaches, and *17 stories*. A website (https://textmining-infopros.github.io/) and a GitHub account (https://github.com/textmining-infopros) are also maintained for the book. They contain the code, data, and notebooks for the case studies; a summary of all the stories shared by the librarians/faculty; and hyperlinks to open an interactive virtual RStudio/Jupyter Notebook environment. The interactive virtual environment runs case studies based on the R programming language for hands-on practice in the cloud without installing any software. Text mining is a topic of international interest, and this book has been written to meet the reading interests of both national and international audiences. It will be appropriate for both beginners and advanced-level readers as it has been written keeping their information needs in mind.

Many books in the market are written to meet the need of computer science professionals on text mining, whereas there are very few books on text mining for librarians. Also, the books present in the market on this topic are very difficult for non-programmers to understand. They may contain lots of jargon, which may not be easily understood by a library professional. In contrast, this book focuses on a basic theoretical framework dealing with the problems, solutions, and applications of text mining and its various facets in the form of *case studies*, *use cases*, and *stories*.

Delhi, India — Manika Lamba
Delhi, India — Margam Madhusudhan
June 2021

Acknowledgments

The authors are immensely grateful to Alex Wermer-Colan, Amy J. Kirchhoff, Anne M. Brown, C. Cozette Comer, Carady DeSimone, Chreston Miller, Cody Hennesy, Issac Williams, Jacob Lahne, Jonathan Briganti, Jordan Patterson, Karen Harker, Leah Hamilton, Lighton Phiri, Manisha Bolina, Manoj Kumar Verma, Marcella Fredriksson, Maya Deori, Michael J. Stamper, Nathaniel D. Porter, Parthasarathi Mukhopadhyay, Rachel Miles, Sephra Byrne, and Vinit Kumar for sharing their personal experience of using different text mining approaches for the *story section* of the book.

Contents