

Unveiling Language: A Comprehensive Guide to Natural Language Processing

First Edition

Authors

Dr. Nisha Varghese Prof. Dr. M. Punithavalli



Title of the Book: Unveiling Language: A Comprehensive Guide to Natural Language Processing

First Edition - 2024

Copyright 2024 © Authors

Dr. Nisha Varghese, Assistant Professor, Department of Computer Science, CHRIST (Deemed to be University), Hosur Road, Bengaluru, Karnataka, India.

Prof. Dr. M. Punithavalli, Professor and Head, Department of Computer Applications, Bharathiar University, Coimbatore, Tamilnadu, India.

No part of this book may be reproduced or transmitted in any form by any means, electronic or mechanical, including photocopy, recording or any information storage and retrieval system, without permission in writing from the copyright owners.

Disclaimer

The authors are solely responsible for the contents published in this book. The publishers don't take any responsibility for the same in any manner. Errors, if any, are purely unintentional and readers are requested to communicate such errors to the editors or publishers to avoid discrepancies in future.

E-ISBN: 978-93-5747-667-6

MRP Rs. 280/-

Publisher, Printed at & Distribution by:

Selfypage Developers Pvt Ltd., Pushpagiri Complex, Beside SBI Housing Board, K.M. Road Chikkamagaluru, Karnataka.

Tel.: +91-8861518868 E-mail:info@iipbooks.com

IMPRINT: I I P Iterative International Publishers

For Sales Enquiries:

Contact: +91-8861511583 E-mail: sales@iipbooks.com

Foreword

In this era of information and digital communication, the realm of Natural Language Processing (NLP) stands as a pillar for understanding and extracting meaning from the vast universe of human language. This book is designed as a guide that unlocks the detailed landscape of NLP, providing the reader with the knowledge, tools, and techniques in the area.

The initial chapter delves into the history of language processing and underscores the significance of NLP in contemporary times. It explores various trends and applications of NLP, understanding different levels and facilitating the extraction of meaningful insights. Moreover, it addresses the critical issue of nuances in processing text data and emphasizes the importance of preprocessing to ensure clarity and accuracy in analyses. Following the history of NLP, the book guides readers through the gateway to explore the profusion of unstructured and semi-structured data, along with various methods and tools for normalization. The book explains in detail the techniques in data processing/cleaning and how to transform unstructured text data into a structured and analyzable format.

The book also unveils the worlds of text similarity by exploring various similarity metrics and their applications, ranging from plagiarism checking to text categorization, question answering, and more. As a case study, the book delves into the semantic similarities within three Holy books: The Bible, The Quran, and the Tanakh, unveiling the insights they provide. Furthermore, the book takes you on a journey through the realms of classification and prediction in machine learning and data science. These techniques are indispensable for analyzing and categorizing data to make informed decisions and predictions. Readers will be able to unravel the mysteries of classification and clustering algorithms, exemplifying their use in solving real-world problems. The book also introduces sentiment analysis in NLP, a powerful technique for understanding and extracting sentiment from text data. This helps us learn how businesses and organizations can leverage sentiment analysis to gain valuable insights and make data-driven decisions by gaining a deeper understanding of the human psyche.

Continuing through the journey of NLP, readers will delve deeper into the treasure trove of traditional Question Answering and Neural Machine Reading Comprehension. This is a journey through the technology that enables machines to understand, process, and respond to human queries, and explore the architectural aspects and case studies that make this field fascinating. You'll find a comprehensive roadmap that covers modules, tasks, baseline models, types of

questions and answers, and recent trends in the field. Furthermore, the book explores two core NLP tasks: Topic modeling and Text Summarization, discovering how topic modeling can automatically identify and extract topics from a collection of documents, and exploring the art of condensing text into concise and coherent summaries using extractive and abstractive techniques.

The book opens the door to text generation, an exciting NLP task where it explores various techniques, including rule-based, statistical, and deep learning-based models. It covers dialogue systems and machine translation, revolutionizing the ways we communicate and understand languages and explains the techniques for generating human-like text and bridging language barriers.

Towards the end of the journey, the book introduces the crucial component of data analysis, i.e., data visualization in NLP. The book explores the power of visualizing data using Python and tracks down the libraries that enable us to create stunning and informative data visualizations tailored to our specific needs.

This book is more than a guide; it is a gateway to a world of possibilities. It offers an opportunity to transform the complexity of human language into insights, applications, and innovation. Providing a deeper understanding of the intricate world of NLP and data analysis, each chapter is designed to equip you with the knowledge and tools needed to navigate the world of textual data effectively. Whether you are a seasoned data scientist or a curious beginner, this book has something to offer. Turn the page, and let's delve into the realm of Natural Language Processing together.

Dr. Rajeev R R Associate Professor and Head International Centre for Free and Open-Source Software (ICFOSS) Kerala - 695581

Preface

Welcome to the world of Natural Language Processing (NLP), a field at the intersection of linguistics, computer science, and artificial intelligence. In this book, Unveiling Language: A Comprehensive Guide to Natural Language Processing, we embark on a journey to explore the fascinating realm of human language and the myriad ways in which machines can understand, interpret, and generate it.

Language, as a fundamental medium of human communication, has long been a source of intrigue and study. With the advent of computers and the digital age, the quest to bridge the gap between human language and machines has given rise to NLP. This field encompasses a broad range of challenges, from extracting meaning from raw text to enabling machines to engage in nuanced conversations.

In these pages, we aim to demystify the complexities of NLP, making it accessible to both the curious novice and the seasoned practitioner. We will delve into the core concepts, methodologies, and applications that define the landscape of NLP. From the basics of tokenization and part-of-speech tagging to advanced topics like sentiment analysis, machine translation, and dialogue systems, each chapter is crafted to build a comprehensive understanding of the subject.

As we navigate through the intricacies of NLP, we will encounter the power and limitations of current technologies, explore the ethical considerations that arise when dealing with language data, and envision the future possibilities that lie ahead. The field of NLP is dynamic, evolving rapidly with each technological breakthrough, and this book aims to equip you with the knowledge and skills to navigate this ever-changing landscape.

Whether you are a student, a researcher, or a practitioner in the field of artificial intelligence, we hope this book serves as a valuable resource in your quest to unlock the potential of natural language understanding. Join us on this expedition, where words meet algorithms, and together, we unravel the secrets of human language through the lens of Natural Language Processing.

Happy Reading.....!

Dr. Nisha Varghese, Prof. Dr. M. Punithavalli

Benefits to Users

In this digital era the communication with machines (Computers, Digital Gadgets/wearable devices) are inevitable. Human-Computer Interaction plays a key role in industry 5.0. As the result of Industrialization human are collaborated with the computers in all the areas of Industry. Consequently, machines that interacting with human should have the potential to understand the input in a comprehensive way. Otherwise, the communication won't be effective. Teaching machines to read, comprehend and seek answers from natural language documents is an arduous task because it involves language understanding. The noisy and unstructured form of text document (corpus) also increases the complexity of extracting the relevant information from the natural language documents. The rapid advancement of Natural Language Processing (NLP) has transformed the way humans interact with technology and communicate with each other.

Language Processing book ("Unveiling The Natural Language: Comprehensive Guide to Natural Language Processing") extends comprehensive invitation to readers seeking to harness the transformative power of language in the digital age. Languages are the vital and fundamental component of humankind for communication. "Unveiling Language" is a comprehensive and accessible guide that delves into the intricate world of NLP, offering both novices and experts an in-depth understanding of the field's foundational concepts, methodologies, and cutting-edge applications. This book starts by unraveling the History and fundamental principles of NLP, providing readers with a solid groundwork for comprehending the complexities of NLP. It explores the text understanding and normalization (pre-processing) techniques. Through clear explanations and real-world examples, the book clarifies NLP level concepts such as tokenization, syntactic analysis, semantic representation, and discourse modeling.

By delving into the intricate realms of NLP, readers will not only acquire a solid foundation in fundamental concepts like text processing and language modeling but also gain practical insights into real-world applications. The book serves as a practical guide, offering hands-on examples and case studies that bridge theory with implementation, empowering readers to navigate the complexities of natural language understanding. Whether you are an aspiring data scientist, a software engineer, or a seasoned AI enthusiast, the benefits are manifold. You'll develop the skills to design sophisticated language models, extract meaningful information from vast text datasets, and contribute to cutting-edge advancements in fields such as sentiment analysis, machine translation, and chatbot development. With ethical considerations woven into the narrative,

readers are equipped not just with technical proficiency but also a nuanced understanding of the societal implications of NLP. Now the book to you......for unlocking the doors of innovation in language technology, enabling you to explore, create, and contribute to the exciting future of Natural Language Processing.

NB: All the references of the chapters and the links of essential tools and techniques are included in the QR code at the end of every chapter.

Acknowledgments

I would like to extend my deepest gratitude to the **Almighty**, whose unwavering guidance and blessings have been the source of strength throughout this literary journey. Your divine presence has illuminated my path, providing inspiration, perseverance, and wisdom. In the moments of doubt, it was the faith in your plan that kept me going. I am profoundly thankful for the spiritual support that has fueled the creation of this work. To the Almighty, I offer my heartfelt thanks for being the silent co-author of this endeavor.

I am thankful to **Prof. Dr. M. Punithavalli**, My Co-Author, the Head of the Department of Computer Applications, Bharathiar University, for her constant motivation and support.

I am immensely grateful to **Mr. Shafi Shereef**, for your unwavering support and invaluable assistance throughout the journey of completing my book. Your encouragement, insightful feedback, proof reading and willingness to lend a helping hand have been instrumental in bringing this book to fruition. I couldn't have asked for a better friend and collaborator. Thank you for being a constant source of inspiration and for sharing in the joy of this accomplishment.

As I stand on the threshold of completing my book, I am filled with gratitude for the incredible support I've received from each one of you. **Dr. Rajeev R R**, for the Foreword and guidance, **Dinesh Paranthagan** (CEO, Hackup Technology) for your motivation, my friends and Colleagues in CHRIST (Deemed to be University). Your encouragement, thoughtful insights, and shared enthusiasm have been the driving force behind this creative endeavor.

I am indebted to my family members and there are no words to express my gratitude and gratefulness to them. Prayers and encouragement of my caring father, **Varghese Xavier** and spiritual and emotional support of my compassionate mother, **Alphonsa Thomas** led to the successful completion of my Book. Most of all I value my partner, **Abhilash Paul** and my kids **Aqueena** and **Aaron** for their endless patience, kindness, understanding and support.

I look forward to sharing the finished work with all of you and am deeply appreciative of the role each of you has played in making my dream a reality.

With heartfelt thanks **Dr. Nisha Varghese**

Contents

Chapter No.	Chapter Name	Page. No	
Chapter 1	An Introduction to Natural Language	1-20	
	Processing	_	
1.1	History of Natural Language Processing	2	
1.2	Components of Natural Language Processing	4	
1.3	Applications of Natural Language Processing	6	
1.4	Text Analysis vs Text Analytics	11	
1.5	Levels of Natural Language Processing	13	
1.6	Natural Language Processing Roadmap	18	
1.7	Summary	19	
1.8	Glossary	19	
Chapter 2	Text Understanding and Normalization	21-53	
$\overline{2.1}$	Natural Language Processing Libraries	22	
2.2	Text Understanding and Text Normalization	28	
2.3	Unveiling of Text	37	
2.4	Word Vector Representation	41	
2.5	Sentence Embedding	50	
2.6	Summary	52	
2.7	Glossary	52	
Chapter 3	Text Similarity	54-76	
3.1	Distance Measures and Distance Metrics	56	
3.2	Family of Similarity Measures and Distances	56	
3.3	Types of Similarity Measures	64	
3.4	Types of Similarities	68	
3.5	Semantic Similarity Approaches	69	
3.6	Datasets for Text Similarity	74	
3.7	Summary	76	
3.8	Glossary	76	
Chapter 4	Text Classification and Clustering	77-120	
4.1	Challenges in Classification and Prediction	80	
4.2	Comparison of Classification and Prediction	81	
4.3	Classification and Regression Algorithms	81	
4.4	Case Study - Text Classification	91	
4.5	Clustering	97	
4.6	Case Study - Clustering	108	
4.7	Text Classification Dataset Repositories	117	

Chapter 5 Sentiment Analysis and Applications 121-142 5.1 Steps of Sentiment Analysis 122 5.2 Strategies for Sentiment Analysis 125 5.3 Tools for Sentiment Analysis 131 5.4 Case Study 132 5.5 Datasets for Sentiment Analysis 137 5.6 Summary 140 5.7 Glossary 141 Chapter 6 Question Answering and Machine Reading Comprehension 143-183 Reading Comprehension 146 Comprehension 146 Comprehension 146 6.2 Machine Reading Comprehension Tasks 147 6.3 Attribute-Based Classification 149 6.4 Recent Trends and Challenges in Reading Comprehension 150 6.5 Benchmarked Datasets 153 6.6 Baseline Models of Machine Reading Comprehension 171 6.8 Datasets for Machine Reading Comprehension 173 6.9 Tools for MRC 174 6.10 Case Study o	4.8	Summary	118
5.1 Steps of Sentiment Analysis 122 5.2 Strategies for Sentiment Analysis 125 5.3 Tools for Sentiment Analysis 131 5.4 Case Study 132 5.5 Datasets for Sentiment Analysis 137 5.6 Summary 140 5.7 Glossary 141 Chapter 6 Question Answering and Machine Reading Comprehension 142 6.1 Architecture of Machine Reading Comprehension Tasks 146 6.1 Architecture of Machine Reading Comprehension Tasks 147 6.2 Machine Reading Comprehension Tasks 147 6.3 Attribute-Based Classification 149 6.4 Recent Trends and Challenges in Reading Comprehension 150 6.5 Benchmarked Datasets 153 6.6 Baseline Models of Machine Reading Comprehension 156 6.7 Performance Evaluation Metrics 171 6.8 Datasets for Machine Reading Comprehension 173 6.9 Tools for MRC 174 6.10	4.9	•	
5.1 Steps of Sentiment Analysis 122 5.2 Strategies for Sentiment Analysis 125 5.3 Tools for Sentiment Analysis 131 5.4 Case Study 132 5.5 Datasets for Sentiment Analysis 137 5.6 Summary 140 5.7 Glossary 141 Chapter 6 Question Answering and Machine Reading Comprehension 142 6.1 Architecture of Machine Reading Comprehension Tasks 146 6.1 Architecture of Machine Reading Comprehension Tasks 147 6.2 Machine Reading Comprehension Tasks 147 6.3 Attribute-Based Classification 149 6.4 Recent Trends and Challenges in Reading Comprehension 150 6.5 Benchmarked Datasets 153 6.6 Baseline Models of Machine Reading Comprehension 156 6.7 Performance Evaluation Metrics 171 6.8 Datasets for Machine Reading Comprehension 173 6.9 Tools for MRC 174 6.10	Chanter 5	Sentiment Analysis and Annlications	121-142
5.2 Strategies for Sentiment Analysis 125 5.3 Tools for Sentiment Analysis 131 5.4 Case Study 132 5.5 Datasets for Sentiment Analysis 137 5.6 Summary 140 5.7 Glossary 141 Chapter 6 Question Answering and Machine Reading Comprehension 143-183 6.1 Architecture of Machine Reading Comprehension Tasks 147 6.1 Architecture of Machine Reading Comprehension Tasks 147 6.2 Machine Reading Comprehension Tasks 147 6.3 Attribute-Based Classification 149 150 6.4 Recent Trends and Challenges in Reading Comprehension 150 6.5 Benchmarked Datasets 153 6.6 Baseline Models of Machine Reading 156 153 6.7 Performance Evaluation Metrics 171 171 6.8 Datasets for Machine Reading 173 173 Comprehension 173 174 174 6.10 Case Study of MRC 175 175 6.11 S	_	· · · · · · · · · · · · · · · · · · ·	
5.3 Tools for Sentiment Analysis 131 5.4 Case Study 132 5.5 Datasets for Sentiment Analysis 137 5.6 Summary 140 5.7 Glossary 141 Chapter 6 Question Answering and Machine Reading Comprehension 143-183 Reading Comprehension 6.1 Architecture of Machine Reading Comprehension Tasks 147 6.2 Machine Reading Comprehension Tasks 147 6.3 Attribute-Based Classification 149 6.4 Recent Trends and Challenges in Reading Comprehension 150 6.5 Benchmarked Datasets 153 6.6 Baseline Models of Machine Reading Comprehension 156 6.7 Performance Evaluation Metrics 171 6.8 Datasets for Machine Reading Comprehension 173 6.9 Tools for MRC 174 6.10 Case Study of MRC 175 6.11 Summary 182 Chapter 7 Topic Modeling and Text Summarization 186 7.3<		•	
5.4 Case Study 132 5.5 Datasets for Sentiment Analysis 137 5.6 Summary 140 5.7 Glossary 141 Chapter 6 Question Answering and Machine Reading Comprehension 143-183 Reading Comprehension 146 Comprehension 6.1 Architecture of Machine Reading Comprehension Tasks 147 6.2 Machine Reading Comprehension Tasks 147 6.3 Attribute-Based Classification 149 6.4 Recent Trends and Challenges in Reading Comprehension 150 6.5 Benchmarked Datasets 153 6.6 Baseline Models of Machine Reading Comprehension 153 6.7 Performance Evaluation Metrics 171 6.8 Datasets for Machine Reading Comprehension 173 6.9 Tools for MRC 174 6.10 Case Study of MRC 175 6.11 Summary 182 Chapter 7 Topic Modeling and Text Summarization 184-210 7.1 <td></td> <td>· ·</td> <td></td>		· ·	
5.5 Datasets for Sentiment Analysis 137 5.6 Summary 140 5.7 Glossary 141 Chapter 6 Question Answering and Machine Reading Comprehension 143-183 Reading Comprehension 146 Comprehension 6.1 Architecture of Machine Reading Comprehension Tasks 147 6.2 Machine Reading Comprehension Tasks 147 6.3 Attribute-Based Classification 149 6.4 Recent Trends and Challenges in Reading Comprehension 150 6.5 Benchmarked Datasets 153 6.6 Baseline Models of Machine Reading Comprehension 156 6.7 Performance Evaluation Metrics 171 6.8 Datasets for Machine Reading Comprehension 173 6.9 Tools for MRC 174 6.10 Case Study of MRC 175 6.11 Summary 182 Chapter 7 Topic Modeling and Text Summarization 184-210 7.1 Stages of Topic Modeling 186 7.2 <td></td> <td>•</td> <td></td>		•	
5.6 Summary 140 5.7 Glossary 141 Chapter 6 Question Answering and Machine Reading Comprehension 143-183 6.1 Architecture of Machine Reading Comprehension Tasks 146 6.2 Machine Reading Comprehension Tasks 147 6.3 Attribute-Based Classification 149 6.4 Recent Trends and Challenges in Reading Comprehension 150 6.5 Benchmarked Datasets 153 6.6 Baseline Models of Machine Reading Comprehension 156 6.7 Performance Evaluation Metrics 171 6.8 Datasets for Machine Reading Comprehension 173 6.9 Tools for MRC 174 6.10 Case Study of MRC 175 6.11 Summary 182 Chapter 7 Topic Modeling and Text Summarization 184-210 7.1 Stages of Topic Modeling 185 7.2 Applications of Topic Modeling 186 7.3 Algorithms and Models for Topic Modeling 188 7.4 Text Summarization		•	
5.7 Glossary 141 Chapter 6 Question Answering and Machine Reading Comprehension 143-183 6.1 Architecture of Machine Reading Comprehension Tasks 146 6.2 Machine Reading Comprehension Tasks 147 6.3 Attribute-Based Classification 149 6.4 Recent Trends and Challenges in Reading Comprehension 150 6.5 Benchmarked Datasets 153 6.6 Baseline Models of Machine Reading Comprehension 156 6.7 Performance Evaluation Metrics 171 6.8 Datasets for Machine Reading Comprehension 173 6.9 Tools for MRC 174 6.10 Case Study of MRC 175 6.11 Summary 182 Chapter 7 Topic Modeling and Text Summarization 184-210 7.1 Stages of Topic Modeling 185 7.2 Applications of Topic Modeling 186 7.3 Algorithms and Models for Topic Modeling 188 7.4 Text Summarization 200 7.5		•	
Reading Comprehension 6.1 Architecture of Machine Reading Comprehension 6.2 Machine Reading Comprehension Tasks 6.3 Attribute-Based Classification 6.4 Recent Trends and Challenges in Reading Comprehension 6.5 Benchmarked Datasets 6.6 Baseline Models of Machine Reading Comprehension 6.7 Performance Evaluation Metrics 6.8 Datasets for Machine Reading Comprehension 6.9 Tools for MRC 6.10 Case Study of MRC 6.11 Summary 6.12 Glossary Chapter 7 Topic Modeling and Text Summarization 7.1 Stages of Topic Modeling 7.2 Applications of Topic Modelling 7.3 Algorithms and Models for Topic Modeling 7.4 Text Summarization 7.5 Text Summarization Applications 7.6 Tools for Text Summarization Using the Summy 7.8 Summary 208		•	
6.1 Architecture of Machine Reading Comprehension 6.2 Machine Reading Comprehension Tasks 6.3 Attribute-Based Classification 6.4 Recent Trends and Challenges in Reading Comprehension 6.5 Benchmarked Datasets 6.6 Baseline Models of Machine Reading Comprehension 6.7 Performance Evaluation Metrics 6.8 Datasets for Machine Reading Comprehension 6.9 Tools for MRC 6.10 Case Study of MRC 6.11 Summary 6.12 Glossary Chapter 7 Topic Modeling and Text Summarization 7.1 Stages of Topic Modelling 7.2 Applications of Topic Modelling 7.3 Algorithms and Models for Topic Modeling 7.4 Text Summarization 7.5 Text Summarization Applications 7.6 Tools for Text Summarization Using the Summy 7.8 Summary 208	Chapter 6		143-183
6.2 Machine Reading Comprehension Tasks 6.3 Attribute-Based Classification 6.4 Recent Trends and Challenges in Reading Comprehension 6.5 Benchmarked Datasets 6.6 Baseline Models of Machine Reading Comprehension 6.7 Performance Evaluation Metrics 6.8 Datasets for Machine Reading Comprehension 6.9 Tools for MRC 6.10 Case Study of MRC 6.11 Summary 6.12 Glossary 182 Chapter 7 Topic Modeling and Text Summarization 7.1 Stages of Topic Modeling 7.2 Applications of Topic Modeling 7.3 Algorithms and Models for Topic Modeling 7.4 Text Summarization 7.5 Text Summarization Applications 7.6 Tools for Text Summarization Using the Summy 7.8 Summary 208	6.1	Architecture of Machine Reading	146
6.3 Attribute-Based Classification 6.4 Recent Trends and Challenges in Reading Comprehension 6.5 Benchmarked Datasets 6.6 Baseline Models of Machine Reading Comprehension 6.7 Performance Evaluation Metrics 6.8 Datasets for Machine Reading Comprehension 6.9 Tools for MRC 6.10 Case Study of MRC 6.11 Summary 6.12 Glossary 182 Chapter 7 Topic Modeling and Text Summarization 7.1 Stages of Topic Modeling 7.2 Applications of Topic Modeling 7.3 Algorithms and Models for Topic Modeling 7.4 Text Summarization 7.5 Text Summarization Applications 7.6 Tools for Text Summarization Using the Summy 7.8 Summary 208		-	
6.4 Recent Trends and Challenges in Reading Comprehension 6.5 Benchmarked Datasets 6.6 Baseline Models of Machine Reading Comprehension 6.7 Performance Evaluation Metrics 6.8 Datasets for Machine Reading Comprehension 6.9 Tools for MRC 6.10 Case Study of MRC 6.11 Summary 6.12 Glossary Chapter 7 Topic Modeling and Text Summarization 7.1 Stages of Topic Modeling 7.2 Applications of Topic Modeling 7.3 Algorithms and Models for Topic Modeling 7.4 Text Summarization 7.5 Text Summarization Applications 7.6 Tools for Text Summarization Using the Summy 7.8 Summary 208			
Comprehension 6.5 Benchmarked Datasets 6.6 Baseline Models of Machine Reading Comprehension 6.7 Performance Evaluation Metrics 6.8 Datasets for Machine Reading Comprehension 6.9 Tools for MRC 6.10 Case Study of MRC 6.11 Summary 6.12 Glossary Chapter 7 Topic Modeling and Text Summarization 7.1 Stages of Topic Modeling 7.2 Applications of Topic Modelling 7.3 Algorithms and Models for Topic Modeling 7.4 Text Summarization 7.5 Text Summarization Applications 7.6 Tools for Text Summarization Using the Summy 7.8 Summary 208			
6.5 Benchmarked Datasets 6.6 Baseline Models of Machine Reading Comprehension 6.7 Performance Evaluation Metrics 6.8 Datasets for Machine Reading Comprehension 6.9 Tools for MRC 6.10 Case Study of MRC 6.11 Summary 6.12 Glossary 182 Chapter 7 Topic Modeling and Text Summarization 7.1 Stages of Topic Modeling 7.2 Applications of Topic Modeling 7.3 Algorithms and Models for Topic Modeling 7.4 Text Summarization 7.5 Text Summarization Applications 7.6 Tools for Text Summarization Using the Summy 7.8 Summary 208	6.4	•	150
6.6 Baseline Models of Machine Reading Comprehension 6.7 Performance Evaluation Metrics 6.8 Datasets for Machine Reading Comprehension 6.9 Tools for MRC 6.10 Case Study of MRC 6.11 Summary 6.12 Glossary Chapter 7 Topic Modeling and Text Summarization 7.1 Stages of Topic Modeling 7.2 Applications of Topic Modeling 7.3 Algorithms and Models for Topic Modeling 7.4 Text Summarization 7.5 Text Summarization Applications 7.6 Tools for Text Summarization 7.7 Case Study: Text Summarization Using the Summy 7.8 Summary 208	6.5	<u> </u>	152
Comprehension 6.7 Performance Evaluation Metrics 6.8 Datasets for Machine Reading Comprehension 6.9 Tools for MRC 6.10 Case Study of MRC 6.11 Summary 6.12 Glossary Chapter 7 Topic Modeling and Text Summarization 7.1 Stages of Topic Modeling 7.2 Applications of Topic Modeling 7.3 Algorithms and Models for Topic Modeling 7.4 Text Summarization 7.5 Text Summarization Applications 7.6 Tools for Text Summarization 7.7 Case Study: Text Summarization Using the Summy 7.8 Summary 208			
6.7 Performance Evaluation Metrics 6.8 Datasets for Machine Reading Comprehension 6.9 Tools for MRC 6.10 Case Study of MRC 6.11 Summary 6.12 Glossary Chapter 7 Topic Modeling and Text Summarization 7.1 Stages of Topic Modeling 7.2 Applications of Topic Modeling 7.3 Algorithms and Models for Topic Modeling 7.4 Text Summarization 7.5 Text Summarization Applications 7.6 Tools for Text Summarization 7.7 Case Study: Text Summarization Using the Summy 7.8 Summary 208	0.0		130
6.8 Datasets for Machine Reading Comprehension 6.9 Tools for MRC 6.10 Case Study of MRC 6.11 Summary 6.12 Glossary Chapter 7 Topic Modeling and Text Summarization 7.1 Stages of Topic Modeling 7.2 Applications of Topic Modeling 7.3 Algorithms and Models for Topic Modeling 7.4 Text Summarization 7.5 Text Summarization Applications 7.6 Tools for Text Summarization 7.7 Case Study: Text Summarization Using the Summy 7.8 Summary 208	6.7	-	171
Comprehension 6.9 Tools for MRC 6.10 Case Study of MRC 6.11 Summary 6.12 Glossary Chapter 7 Topic Modeling and Text Summarization 7.1 Stages of Topic Modeling 7.2 Applications of Topic Modelling 7.3 Algorithms and Models for Topic Modeling 7.4 Text Summarization 7.5 Text Summarization Applications 7.6 Tools for Text Summarization 7.7 Case Study: Text Summarization Using the Summy 7.8 Summary 208			
6.9 Tools for MRC 6.10 Case Study of MRC 6.11 Summary 6.12 Glossary Chapter 7 Topic Modeling and Text Summarization 7.1 Stages of Topic Modeling 7.2 Applications of Topic Modelling 7.3 Algorithms and Models for Topic Modeling 7.4 Text Summarization 7.5 Text Summarization Applications 7.6 Tools for Text Summarization 7.7 Case Study: Text Summarization Using the Summy 7.8 Summary 208		_	
6.11 Summary 6.12 Glossary Chapter 7 Topic Modeling and Text Summarization 7.1 Stages of Topic Modeling 7.2 Applications of Topic Modelling 7.3 Algorithms and Models for Topic Modeling 7.4 Text Summarization 7.5 Text Summarization Applications 7.6 Tools for Text Summarization 7.7 Case Study: Text Summarization Using the Summy 7.8 Summary 208	6.9	-	174
6.11 Summary 6.12 Glossary Chapter 7 Topic Modeling and Text Summarization 7.1 Stages of Topic Modeling 7.2 Applications of Topic Modelling 7.3 Algorithms and Models for Topic Modeling 7.4 Text Summarization 7.5 Text Summarization Applications 7.6 Tools for Text Summarization 7.7 Case Study: Text Summarization Using the Summy 7.8 Summary 208	6.10	Case Study of MRC	175
Chapter 7 Topic Modeling and Text Summarization 7.1 Stages of Topic Modeling 7.2 Applications of Topic Modelling 7.3 Algorithms and Models for Topic Modeling 7.4 Text Summarization 7.5 Text Summarization Applications 7.6 Tools for Text Summarization 7.7 Case Study: Text Summarization Using the Summy 7.8 Summary 208	6.11	•	182
7.1 Stages of Topic Modeling 185 7.2 Applications of Topic Modelling 186 7.3 Algorithms and Models for Topic Modeling 188 7.4 Text Summarization 200 7.5 Text Summarization Applications 202 7.6 Tools for Text Summarization 205 7.7 Case Study: Text Summarization Using the Summy 7.8 Summary 208	6.12	•	182
7.2 Applications of Topic Modelling 7.3 Algorithms and Models for Topic Modeling 7.4 Text Summarization 7.5 Text Summarization Applications 7.6 Tools for Text Summarization 7.7 Case Study: Text Summarization Using the 8 Summary 7.8 Summary 8 208	Chapter 7	Topic Modeling and Text Summarization	184-210
7.3 Algorithms and Models for Topic Modeling 7.4 Text Summarization 7.5 Text Summarization Applications 7.6 Tools for Text Summarization 7.7 Case Study: Text Summarization Using the Summy 7.8 Summary 7.8 Summary 7.8 Summary 7.8 200 7.7 Text Summarization Using the Summarization Using the Summarization Using the Summary	7.1	Stages of Topic Modeling	185
7.4 Text Summarization 200 7.5 Text Summarization Applications 202 7.6 Tools for Text Summarization 205 7.7 Case Study: Text Summarization Using the 207 Summy 7.8 Summary 208	7.2	Applications of Topic Modelling	186
 7.5 Text Summarization Applications 7.6 Tools for Text Summarization 7.7 Case Study: Text Summarization Using the Summy 7.8 Summary 202 205 207 Summary 208 	7.3	Algorithms and Models for Topic Modeling	188
7.6 Tools for Text Summarization 205 7.7 Case Study: Text Summarization Using the 207 Summy 7.8 Summary 208	7.4	Text Summarization	200
 7.7 Case Study: Text Summarization Using the Summy 7.8 Summary 207 208 	7.5	Text Summarization Applications	202
Summy 7.8 Summary 208	7.6	Tools for Text Summarization	205
7.8 Summary 208	7.7	•	207
•	7 &		208
/ 9 Cilossary 200	7.8 7.9	Glossarv	208

Chapter 8	pter 8 Text Generation, Machine Translation				
_	and Dialogue Systems				
8.1	Stages of Natural Language Generation	213			
8.2	Applications of Natural Language	214			
	Generation				
8.3	Chatbot or Virtual Assistants	214			
8.4	Types of Chatbots	218			
8.5	Working of a Chatbot	219			
8.6	Tools for Developing Chatbots	220			
8.7	Machine Translation	221			
8.8	Summary	228			
8.9	Glossary	228			
Chapter 9	Data Visualization	230-267			
9.1	Data Visualization Timeline	232			
9.2	Visualization Libraries	234			
9.3	Types of Visualization	240			
9.4	Summary	266			
9.5	Glossary	266			

AUTHORS PROFILE

Dr. Nisha Varghese received her MPhil. and Ph.D. degree in Computer Science from Bharathiar University. She is currently working as an Assistant Professor in the Department of Computer Science in Christ (Deemed to be University), Bangalore, India. She has more than 10 years of Academic and Research experience. Her research interests include Natural Language Processing and Information Retrieval, Data Mining, Machine Learning and Cyber Security. She holds funded projects.

Prof. Dr. M. Punithavalli received her Ph.D. degree in Computer Science from Alagappa University. She is currently Professor and Head of the Department of Computer Applications, Bharathiar University, Coimbatore, Tamilnadu. India. She has 27 years of Academic and Research experience. Her research interests include Information Security, Machine Learning, Data Mining, and Software Engineering. She has authored three books and published more than 100 research articles in international journals. She also holds funded projects.

