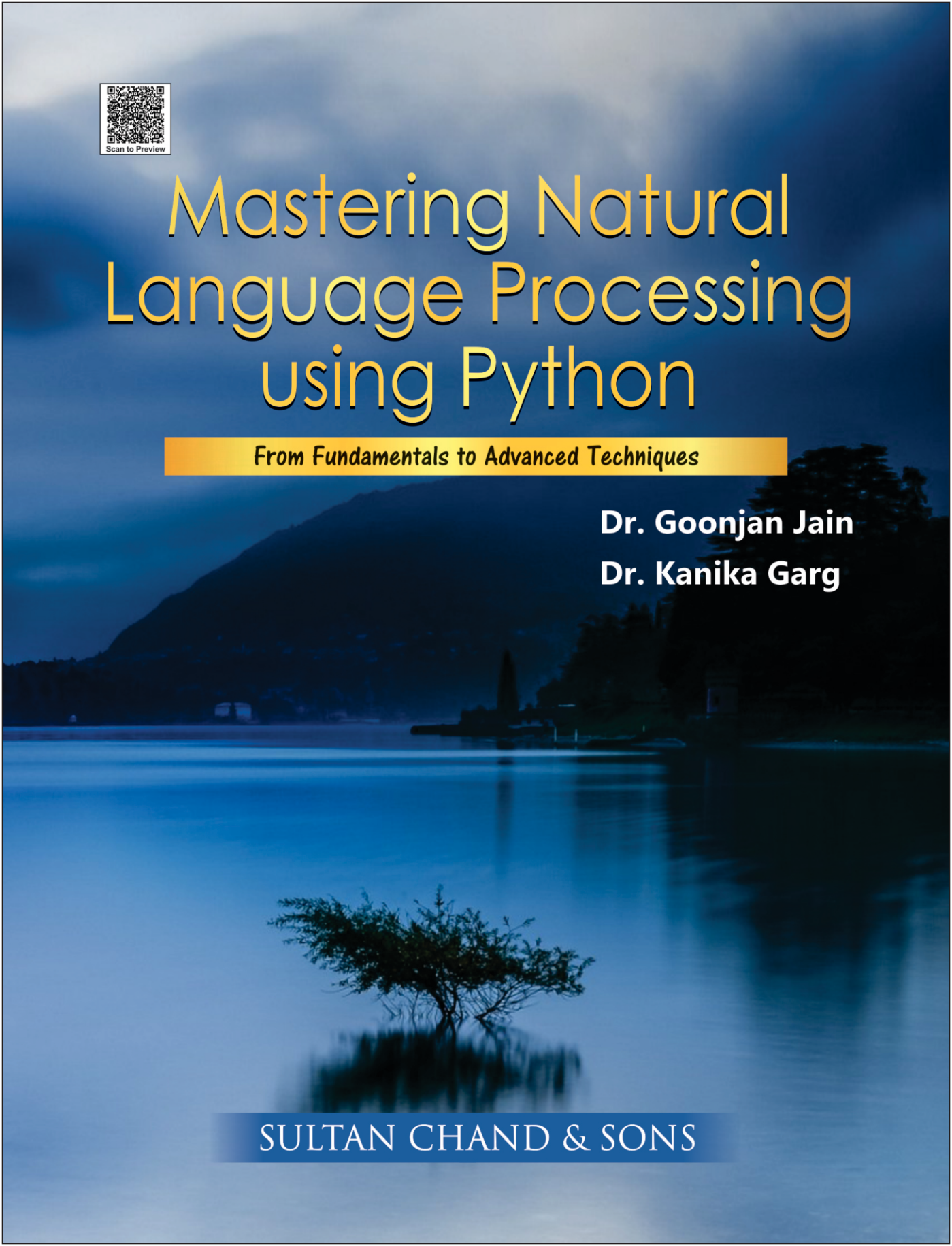
Scan to Preview
Mastering Natural Language Processing using Python
From Fundamentals to Advanced Techniques
Dr. Goonjan Jain
Dr. Kanika Garg
SULTAN CHAND & SONS

Mastering Natural Language Processing using Python

From Fundamentals to Advanced Techniques

Dr. Goonjan Jain
Assistant Professor
Mathematics and Computing
Delhi Technological University (DTU)
Delhi

Dr. Kanika Garg
Associate Professor
Computer Science Department
Geetanjali Institute of Technical Studies
Dabok, Udaipur-313022

SULTAN CHAND & SONS®
Educational Publishers
New Delhi

SULTAN CHAND & SONS®
Educational Publishers
23, Daryaganj, New Delhi-110 002
Phones : 011-23281876, 23266105, 41625022 *(Showroom & Shop)*
011-23247051, 40234454 *(Office)*
E-Mail : sultanchand74@yahoo.com; info@sultanchandandsons.com
Fax : 011-23266357; Website : www.sultanchandandsons.com

ISBN : 978-93-49290-62-4 (TC-1319)
Price : ₹ **295.00**
First Edition: 2025

EVERY GENUINE COPY OF THIS BOOK HAS A HOLOGRAM

In our endeavour to protect you against counterfeit/fake books, we have pasted a copper hologram over the cover of this book. The hologram displays the full visual image, unique 3D multi-level, multi-colour effects of our logo from different angles when tilted or properly illuminated under a single light source, such as 3D depth effect, kinetic effect, pearl effect, gradient effect, trailing effect, emboss effect, glitter effect, randomly sparking tiny dots, micro text, laser numbering, etc.

A fake hologram does not display all these effects.

Always ask the bookseller to put his stamp on the first page of this book.

Author's Acknowledgement: The writing of a Textbook always involves creation of a huge debt towards innumerable authors and publications. We owe our gratitude to all of them. We acknowledge our indebtedness in extensive footnotes throughout the book. If, for any reason, any acknowledgement has been left out we beg to be excused. We assure to carry out correction in the subsequent edition, as and when it is known.

Printed at: Taj Press, Sector-63, Noida

Preface

> "When the student is ready, the teacher will appear. When the student is truly ready, the teacher will disappear." — *Lao Tzu*

This book offers a comprehensive introduction to Natural Language Processing (NLP) and its diverse range of applications, from machine translation to sentiment analysis. This book provides both theoretical foundations and practical insights into the methods and technologies behind NLP systems. It aims to give readers a well-rounded understanding of how machines process natural language, the associated challenges, and the ways NLP techniques can solve real-world problems.

NLP has emerged as one of the most transformative subfields within artificial intelligence, bridging the gap between human communication and machine understanding. As NLP continues to advance, it is making profound impacts across various sectors, including healthcare, business, and beyond. The potential of this field seems boundless, particularly as it evolves alongside developments in machine learning and large language models.

Human language is the most fundamental means of communication, and enabling machines to understand and process it is a research-worthy endeavor. From healthcare to business, NLP allows for automation, enhances human-computer interaction, and facilitates the extraction of valuable insight. Research in NLP is pivotal for building the next generation of intelligent systems that can truly understand human language in all its complexity.

The scope of this book encompasses the fundamental concepts, methodologies, and advanced techniques in the field of NLP. It covers a wide range of topics, from the basics of language processing, such as tokenization and morphological analysis, to more advanced areas like machine translation, information retrieval, and deep learning-based NLP models. The book explores classical methods (*e.g.,* rule-based systems, probabilistic models) and modern approaches (*e.g.,* neural networks, transformer-based models like BERT and GPT). With hands-on implementations and practical examples throughout, it serves as a comprehensive guide for readers.

This book is intended for students, researchers, and industry professionals seeking a deeper understanding of Natural Language Processing. Whether you are new to the field or looking to enhance your expertise, this book offers a structured approach to learning both the theory and practical techniques in NLP. It is ideal for university courses on NLP, AI, and machine learning, as well as professionals and enthusiasts looking to explore the applications of NLP in real-world scenarios.

By the end of the book, readers will have a solid foundation in NLP and be equipped to tackle a variety of tasks, from basic text preprocessing to advanced applications like machine translation, question answering, and sentiment analysis.

We hope this book not only deepens your understanding of NLP but also inspires you to explore its many possibilities.

Goonjan Jain
Kanika Garg

Acknowledgement

For my daughters – Arshi & Ameyaa

– Goonjan Jain

I would like to express my heartfelt gratitude to my family and friends and my husband who have supported and encouraged me throughout the process.

– Kanika Garg

Snapshot of the Book

S. No.	*Chapter Name*	*Pages*	*Figures*	*Tables*	*Code Snippets*	*Exercise Questions*	*Practical Questions*
1.	Introduction to Language Processing	1-19	13	1	–	10	6
2.	Language Modeling	21-44	3	11	5	9	10
3.	Lexical Analysis	45-91	9	3	24	23	11
4.	Syntactic Analysis	93-129	9	10	2	21	6
5.	Semantic Analysis	131-162	10	5	7	13	7
6.	Discourse Processing	163-190	10	5	3	30	4
7.	Natural Language Generation	191-213	14	4	4	20	7
8.	Tasks in NLP	215-268	10	7	27	38	10
9.	Advanced Deep Learning and Large Language Models	269-309	26	8	4	30	10
	Previous Year Questions papers	311-316	–	–	–	–	–
	Bibliography	317-324	–	–	–	–	–
	Total	324	104	54	76	194	71

Contents

About the Book

This book is a comprehensive guide to Natural Language Processing (NLP), designed for both beginners and advanced learners. Whether you're just starting or looking to refine your skills, this book takes you through every aspect of NLP—from the basics of text processing to cutting-edge machine learning techniques used in NLP today. It combines theoretical foundations with practical examples using Python, making complex NLP concepts accessible and actionable.

The book is rich with practical exercises, hands-on Python code snippets, and visual aids, ensuring that readers not only understand the concepts but also see how they apply in real-world scenarios. By the end of the book, readers will be proficient in using NLP libraries and will have a clear understanding of how to implement NLP solutions in Python.

Salient Features

- **Comprehensive Coverage:** From tokenization, stemming, and lemmatization to more advanced topics like named entity recognition (NER), sentiment analysis, and deep learning for NLP.
- **Python-Powered:** Every concept is paired with Python code examples, ensuring readers can practice and apply their knowledge immediately.
- **Beginner-Friendly to Advanced:** The book is structured to cater to all levels, beginning with foundational topics before advancing to complex NLP models.
- **Visual Aids:** Includes a variety of images, diagrams, and visual representations to simplify complex ideas.
- **Real-World Examples:** Contains numerous real-world NLP projects and use-cases to demonstrate practical applications.
- **Hands-On Exercises:** End-of-chapter exercises encourage readers to put their skills to the test and deepen their understanding.
- **Focus on Popular NLP Libraries:** Introduces popular Python NLP libraries like NLTK, spaCy, and Hugging Face, providing a well-rounded toolkit for solving NLP tasks.

About the Authors

Dr. Goonjan Jain has over a decade of experience in teaching, research, and software development. She is currently an Assistant Professor at Delhi Technological University, Delhi (formerly DCE). Prior to joining academia, she worked with a multinational technology company. Dr. Jain holds a Ph.D. with a specialization in Natural Language Processing (NLP) and an M.Tech. in Computer Science and Technology, both from Jawaharlal Nehru University (JNU), Delhi. Dr. Jain has made significant contributions to the field of Natural Language Processing, particularly in the areas of language understanding and computational linguistics. She has published many research papers in prestigious international journals and conferences, including *Nature Portfolio's Scientific Reports* and COLING.

Dr. Kanika Garg is a PhD holder in Natural Language Processing (NLP) and currently serves as an Associate Professor at Geetanjali Institute of Technical Studies, Udaipur. With deep expertise in NLP, she has successfully bridged the gap between academia and industry, bringing real-world experience into the classroom. In addition to teaching, Dr. Kanika has authored several publications in SCI-indexed and Scopus-indexed journals. Her research focuses on advancing NLP techniques and applications in diverse domains. Her work reflects a deep commitment to furthering the understanding of language technologies, both in theory and in practical application. With years of experience in both academic research and industry, Dr. Kanika is passionate about making NLP accessible to a wider audience. This passion is evident in her writing, where she skilfully blends academic rigor with practical insights, providing readers with the tools they need to succeed in the rapidly growing field of NLP.

Dedicated to the Service of Education • SULTAN • 1950

Sultan Chand & Sons
Publishers of Standard Educational Textbooks
23 Daryaganj, New Delhi-110002
Phones (S) : 011-23281876, 23266105, 41625022
(O): 011-23247051, 40234454
Email : sultanchand74@yahoo.com
info@sultanchandandsons.com

Scan to Visit Us

TC 1319

ISBN 978-93-91820-60-2