Krishna Bhavsar, Naresh Kumar,
Pratap Dangeti

Natural Language Processing with Python Cookbook

Over 60 recipes to implement text analytics solutions using deep learning principles

Packt>

Natural Language Processing with Python Cookbook

Natural language processing (NLP) is a field of computer science, artificial intelligence, and computational linguistics concerned with the interaction between computers and human (natural) languages; in particular, it's about programming computers to process large natural language corpora.

This book includes unique recipes that will teach you various aspects of performing NLP with NLTK—the leading Python platform for the task. You will come across various recipes in the book, covering (among other topics) natural language understanding, NLP, and syntactic analysis. You will learn how to understand language, plan sentences, and work around various ambiguities. You will learn how to efficiently use NLTK and implement text classification, identify parts of speech, tag words, and more. You will also learn how to analyze sentence structures and master lexical analysis, syntactic and semantic analysis, pragmatic analysis, and the application of deep learning techniques.

By the end of this book, you will have all the knowledge you need to implement NLP with Python.

Things you will learn:

- Explore various corpora that is available with NLTK and understand how to use WordNet corpus
- Understand how to manage and process raw text like HTML, RSS, PDF, word documents, and so on
- Learn how to pre-process raw text using techniques like tokenization, stemming, spell checker, and so on and also implement them on your own using regular expressions
- Understand the basics of pattern matching in textual analytics with regular expressions
- Learn to use and write your own POS taggers and grammars
- Learn how to perform named entity (NE) extraction and also learn parsers like RD, shift reduce, chart parsers
- Generate text from Shakespeare's writing using LSTM
- Utilize the BABI dataset and LSTM to model episodes
- Develop chat bot in generative way with deep learning

ISBN 978-1-78728-932-1
90000
9 781787 289321

Packt>
www.packtpub.com

Natural Language Processing with Python Cookbook

Krishna Bhavsar, Naresh Kumar, Pratap Dangeti

Natural Language Processing with Python Cookbook

Over 60 recipes to implement text analytics solutions using deep learning principles

Krishna Bhavsar
Naresh Kumar
Pratap Dangeti

Packt

BIRMINGHAM - MUMBAI

Natural Language Processing with Python Cookbook

First published: November 2017

Production reference: 1221117

Published by Packt Publishing Ltd.
Livery Place
35 Livery Street
Birmingham
B3 2PB, UK.
ISBN 978-1-78728-932-1

www.packtpub.com

Credits

Authors
Krishna Bhavsar
Naresh Kumar
Pratap Dangeti

Reviewer
Juan Tomas Oliva Ramos

Commissioning Editor
Veena Pagare

Acquisition Editor
Aman Singh

Content Development Editor
Aishwarya Pandere

Technical Editors
Dinesh Pawar
Suwarna Rajput

Copy Editor
Vikrant Phadkay

Project Coordinator
Nidhi Joshi

Proofreader
Safis Editing

Indexer
Tejal Daruwale Soni

Graphics
Tania Dutta

Production Coordinator
Shraddha Falebhai

About the Authors

Krishna Bhavsar has spent around 10 years working on natural language processing, social media analytics, and text mining in various industry domains such as hospitality, banking, healthcare, and more. He has worked on many different NLP libraries such as Stanford CoreNLP, IBM's SystemText and BigInsights, GATE, and NLTK to solve industry problems related to textual analysis. He has also worked on analyzing social media responses for popular television shows and popular retail brands and products. He has also published a paper on sentiment analysis augmentation techniques in 2010 NAACL. he recently created an NLP pipeline/toolset and open sourced it for public use. Apart from academics and technology, Krishna has a passion for motorcycles and football. In his free time, he likes to travel and explore. He has gone on pan-India road trips on his motorcycle and backpacking trips across most of the countries in South East Asia and Europe.

First and foremost, I would like to thank my mother for being the biggest motivating force and a rock-solid support system behind all my endeavors in life. I would like to thank the management team at Synerzip and all my friends for being supportive of me on this journey. Last but not least, special thanks to Ram and Dorothy for keeping me on track during this professionally difficult year.

Naresh Kumar has more than a decade of professional experience in designing, implementing, and running very-large-scale Internet applications in Fortune Top 500 companies. He is a full-stack architect with hands-on experience in domains such as e-commerce, web hosting, healthcare, big data and analytics, data streaming, advertising, and databases. He believes in open source and contributes to it actively. Naresh keeps himself up-to-date with emerging technologies, from Linux systems internals to frontend technologies. He studied in BITS-Pilani, Rajasthan with dual degree in computer science and economics.

Pratap Dangeti develops machine learning and deep learning solutions for structured, image, and text data at TCS, in its research and innovation lab in Bangalore. He has acquired a lot of experience in both analytics and data science. He received his master's degree from IIT Bombay in its industrial engineering and operations research program. Pratap is an artificial intelligence enthusiast. When not working, he likes to read about next-gen technologies and innovative methodologies. He is also the author of the book *Statistics for Machine Learning* by Packt.

I would like to thank my mom, Lakshmi, for her support throughout my career and in writing this book. I dedicate this book to her. I also thank my family and friends for their encouragement, without which it would not have been possible to write this book.

About the Reviewer

Juan Tomas Oliva Ramos is an environmental engineer from the University of Guanajuato, Mexico, with a master's degree in administrative engineering and quality. He has more than 5 years of experience in the management and development of patents, technological innovation projects, and the development of technological solutions through the statistical control of processes.

He has been a teacher of statistics, entrepreneurship, and the technological development of projects since 2011. He became an entrepreneur mentor and started a new department of technology management and entrepreneurship at Instituto Tecnologico Superior de Purisima del Rincon Guanajuato, Mexico.

Juan is an Alfaomega reviewer and has worked on the book *Wearable Designs for Smart Watches, Smart TVs and Android Mobile Devices*.

Juan has also developed prototypes through programming and automation technologies for the improvement of operations, which have been registered for patents.

I want to thank God for giving me wisdom and humility to review this book.
I thank Packt for giving me the opportunity to review this amazing book and to collaborate with a group of committed people
I want to thank my beautiful wife, Brenda, our two magic princesses (Maria Regina and Maria Renata) and our next member (Angel Tadeo), all of you, give me the strength, happiness, and joy to start a new day. Thanks for being my family.

www.PacktPub.com

For support files and downloads related to your book, please visit www.PacktPub.com. Did you know that Packt offers eBook versions of every book published, with PDF and ePub files available? You can upgrade to the eBook version at www.PacktPub.comand as a print book customer, you are entitled to a discount on the eBook copy. Get in touch with us at service@packtpub.com for more details. At www.PacktPub.com, you can also read a collection of free technical articles, sign up for a range of free newsletters and receive exclusive discounts and offers on Packt books and eBooks.

https://www.packtpub.com/mapt

Get the most in-demand software skills with Mapt. Mapt gives you full access to all Packt books and video courses, as well as industry-leading tools to help you plan your personal development and advance your career.

Why subscribe?

- Fully searchable across every book published by Packt
- Copy and paste, print, and bookmark content
- On demand and accessible via a web browser

Customer Feedback

Thanks for purchasing this Packt book. At Packt, quality is at the heart of our editorial process. To help us improve, please leave us an honest review on this book's Amazon page at `https://www.amazon.com/dp/178728932X`. If you'd like to join our team of regular reviewers, you can email us at `customerreviews@packtpub.com`. We award our regular reviewers with free eBooks and videos in exchange for their valuable feedback. Help us be relentless in improving our products!

Table of Contents