

MU

Sem  
7

Computer Engineering

New Syllabus  
2022-23

- ☆ With Solved Latest UNIVERSITY QUESTION PAPERS.
- ☆ Multiple Choice Questions.
- ☆ Lab Included.

Strictly as per the New Syllabus (REV-2019 'C' Scheme) of  
Mumbai University w.e.f. academic year 2022-2023

Department Optional Course - 3 (Code: CSDC7019)

# NATURAL LANGUAGE PROCESSING

**Prof. R. M. Baphana**

(Adjunct Faculty at C.O.E.P. (Pune)  
Teaching Ph. D. & M. Tech Students  
(Artificial Intelligence & Robotics)

**Dr. Tatwadarshi P. Nagarhall**

(Vidyavardhini's COE & Technology, Vasai Rd)

ISBN : 978-93-5583-091-3



9 789355 830913

**TECH-NEO**  
PUBLICATIONS

*Uthoos Athoos Inspiro Innooation*

A Sachin Shah Venture

www.techneobooks.in  
info@techneobooks.in

M7-83A



Price ₹ 395/-

  
HEAD

Dept of Computer Engg.,  
Vidyavardhini's College of  
Engineering and Technology,  
Vasai Road 401 202



29



## Natural Language Processing

(Course Code : CSDC7013)

(MU - Semester VII - Computer Engineering

- For New Syll. 2022-2023)

Authors : Prof. R. M. Baphana

Dr. Tatwadarshi P. Nagarhalli

First Edition for New Syllabus : July 2022

Tech-Neo ID : M7-83

ISBN : 978-93-5583-067-8

### Copyright © by Authors

All rights reserved.

No part of this publication may be reproduced, copied, or stored in a retrieval system, distributed or transmitted in any form or by any means, including photocopy, recording, or other electronic or mechanical methods, without the prior written permission of the Publisher.

This book is sold subject to the condition that it shall not, by the way of trade or otherwise, be lent, resold, hired out, or otherwise circulated without the publisher's prior written consent in any form of binding or cover other than which it is published and without a similar condition including this condition being imposed on the subsequent purchaser and without limiting the rights under copyright reserved above.

### Published by

#### ▶ Mr. Sachin S. Shah

Managing Director, B. E (Industrial Electronics)  
An Alumunus of IIM Ahmedabad

#### ▶ Mrs. Nayana S. Shah, & Mr. Rahul S. Shah

### Permanent Address

Tech-Neo Publications LLP

Sr. No. 38/1, Behind Pari Company, Khedekar  
Industrial Estate, Narhe, Maharashtra,  
Pune-411041.

Email : info@technobooks.in

Website : www.technobooks.in

### Printed at : Image Offset (Mr. Rahul Shah)

Dugane Ind. Area, Survey No. 28/25, Dhayari, Near  
Pari Company, Pune - 411041. Maharashtra State,  
India.

E-mail : rahulshahimage@gmail.com

## About Managing Director...

- Mr. Sachin Shah

### ▶ Over 25 years of experience in Academic Publishing...

With over two and a half decades of experience in bringing out more than 1200 titles in Engineering, Polytechnic, Pharmacy, Computer Sciences and Information Technology.

### ▶ A driven Educationalist...

1. B.E. (Industrial Electronics) (1992 Batch) from Bharati Vidyapeeth's College of Engineering, affiliated to University of Pune.
2. An Alumnus of IIM Ahmedabad.
3. A Co-Author of bestselling book on "Basic Electrical Engineering" Basic Electronics Engineering" for Degree Course in Engineering
4. For over a decade, been working as a Consultant for Higher Education in USA and several other countries.

### ▶ With path-breaking career...

- A publishing career that started with handwritten cyclostyled notes back in 1992.
- Has to his credit, setting up and expansion of one of the leading companies in higher education publishing.

### ▶ An experienced professional and an expert...

- An energetic, creative & resourceful professional with extensive experience of closely working with the best & the most eminent authors of Publishing Industry, ensures high standards of quality in contents.
- Helping students to attain better understanding and in-depth knowledge of the subject.
- Simplifying the methods of learning and bridging the gap between the best authors in the publishing industry and the student community for decades.

First edition published 2023  
by CRC Press  
6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742

and by CRC Press  
4 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

*CRC Press is an imprint of Taylor & Francis Group, LLC*

© 2023 selection and editorial matter, [Narendra M. Shekoker, Hari Vasudevan, Surya S. Durbha, Antonis Michalas, Tatwadarshi P. Nagarhalli, Ramchandra Sharad Mangrulkar, Monika Mangla]; individual chapters, the contributors

Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, access [www.copyright.com](http://www.copyright.com) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. For works that are not available on CCC please contact [mpkbookspermissions@tandf.co.uk](mailto:mpkbookspermissions@tandf.co.uk)

*Trademark notice:* Product or corporate names may be trademarks or registered trademarks and are used only for identification and explanation without intent to infringe.

ISBN: 978-1-032-11128-5 (hbk)

ISBN: 978-1-032-32979-6 (pbk)

ISBN: 978-1-003-21855-5 (ebk)

DOI: 10.1201/9781003218555

Typeset in Minion  
by Apex CoVantage, LLC