

Chapman & Hall/CRC
Machine Learning & Pattern Recognition Series

HANDBOOK OF
**NATURAL
LANGUAGE
PROCESSING**
SECOND EDITION

Edited by
NITIN INDURKHYA
FRED J. DAMERAU



CRC Press

Taylor & Francis Group

Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

Contents

List of Figures	ix
List of Tables	xiii
Editors.....	xv
Board of Reviewers	xvii
Contributors.....	xix
Preface	xxi

PART I Classical Approaches

1 Classical Approaches to Natural Language Processing <i>Robert Dale</i>	3
2 Text Preprocessing <i>David D. Palmer</i>	9
3 Lexical Analysis <i>Andrew Hippiusley</i>	31
4 Syntactic Parsing <i>Peter Ljunglöf and Mats Wirén</i>	59
5 Semantic Analysis <i>Cliff Goddard and Andrea C. Schalley</i>	93
6 Natural Language Generation <i>David D. McDonald</i>	121

PART II Empirical and Statistical Approaches

7 Corpus Creation <i>Richard Xiao</i>	147
8 Treebank Annotation <i>Eva Hajičová, Anne Abeillé, Jan Hajič, Jiří Mírovský, and Zdeňka Urešová</i>	167
9 Fundamental Statistical Techniques <i>Tong Zhang</i>	189
10 Part-of-Speech Tagging <i>Tunga Güngör</i>	205
11 Statistical Parsing <i>Joakim Nivre</i>	237
12 Multiword Expressions <i>Timothy Baldwin and Su Nam Kim</i>	267

13	Normalized Web Distance and Word Similarity	<i>Paul M.B. Vitányi and Rudi L. Cilibrasi</i>	293
14	Word Sense Disambiguation	<i>David Yarowsky</i>	315
15	An Overview of Modern Speech Recognition	<i>Xuedong Huang and Li Deng</i>	339
16	Alignment	<i>Dekai Wu</i>	367
17	Statistical Machine Translation	<i>Abraham Ittycheriah</i>	409

PART III Applications

18	Chinese Machine Translation	<i>Pascale Fung</i>	425
19	Information Retrieval	<i>Jacques Savoy and Eric Gaussier</i>	455
20	Question Answering	<i>Diego Mollá-Aliod and José-Luis Vicedo</i>	485
21	Information Extraction	<i>Jerry R. Hobbs and Ellen Riloff</i>	511
22	Report Generation	<i>Leo Wanner</i>	533
23	Emerging Applications of Natural Language Generation in Information Visualization, Education, and Health Care	<i>Barbara Di Eugenio and Nancy L. Green</i>	557
24	Ontology Construction	<i>Philipp Cimiano, Johanna Völker, and Paul Buitelaar</i>	577
25	BioNLP: Biomedical Text Mining	<i>K. Bretonnel Cohen</i>	605
26	Sentiment Analysis and Subjectivity	<i>Bing Liu</i>	627
	Index		667