

CONTENTS IN DETAIL

INTRODUCTION	xv
Using Python for Natural Language Processing	xvi
The spaCy Library	xvi
Who Should Read This Book?	xvii
What's in the Book?	xvii

1	
HOW NATURAL LANGUAGE PROCESSING WORKS	1
How Can Computers Understand Language?	2
Mapping Words and Numbers with Word Embedding	2
Using Machine Learning for Natural Language Processing	3
Why Use Machine Learning for Natural Language Processing?	5
What Is a Statistical Model in NLP?	8
Neural Network Models	9
Convolutional Neural Networks for NLP	11
What Is Still on You	12
Keywords	12
Context	13
Meaning Transition	13
Summary	14

2	
THE TEXT-PROCESSING PIPELINE	15
Setting Up Your Working Environment	16
Installing Statistical Models for spaCy	16
Basic NLP Operations with spaCy	17
Tokenization	18
Lemmatization	18
Applying Lemmatization for Meaning Recognition	19
Part-of-Speech Tagging	21
Using Part-of-Speech Tags to Find Relevant Verbs	23
Context Is Important	24
Syntactic Relations	24
Try This	28
Named Entity Recognition	29
Summary	29

3 WORKING WITH CONTAINER OBJECTS AND CUSTOMIZING SPACY 31

spaCy's Container Objects	31
Getting the Index of a Token in a Doc Object	32
Iterating over a Token's Syntactic Children	33
The doc.sents Container	33
The doc.noun_chunks Container	35
Try This	35
The Span Object	36
Try This	37
Customizing the Text-Processing Pipeline	37
Disabling Pipeline Components	38
Loading a Model Step by Step	38
Customizing the Pipeline Components	40
Using spaCy's C-Level Data Structures	42
How It Works	43
Preparing Your Working Environment and Getting Text Files	43
Your Cython Script	44
Building a Cython Module	45
Testing the Module	45
Summary	46

4 EXTRACTING AND USING LINGUISTIC FEATURES 47

Extracting and Generating Text with Part-of-Speech Tags	48
Numeric, Symbolic, and Punctuation Tags	48
Extracting Descriptions of Money	49
Try This	50
Turning Statements into Questions	51
Try This	54
Using Syntactic Dependency Labels in Text Processing	55
Distinguishing Subjects from Objects	55
Deciding What Question a Chatbot Should Ask	56
Try This	60
Summary	61

5 WORKING WITH WORD VECTORS 63

Understanding Word Vectors	64
Defining Meaning with Coordinates	64
Using Dimensions to Represent Meaning	65
The Similarity Method	66
Choosing Keywords for Semantic Similarity Calculations	67
Installing Word Vectors	68
Taking Advantage of Word Vectors That Come with spaCy Models	68
Using Third-Party Word Vectors	68
Comparing spaCy Objects	69
Using Semantic Similarity for Categorization Tasks	70
Extracting Nouns as a Preprocessing Step	71

Try This	72
Extracting and Comparing Named Entities	72
Summary	74

6 FINDING PATTERNS AND WALKING DEPENDENCY TREES 75

Word Sequence Patterns	76
Finding Patterns Based on Linguistic Features	76
Try This	77
Checking an Utterance for a Pattern	77
Using spaCy’s Matcher to Find Word Sequence Patterns	79
Applying Several Patterns	80
Creating Patterns Based on Customized Features	81
Choosing Which Patterns to Apply	83
Using Word Sequence Patterns in Chatbots to Generate Statements	83
Try This	86
Extracting Keywords from Syntactic Dependency Trees	86
Walking a Dependency Tree for Information Extraction	87
Iterating over the Heads of Tokens	87
Condensing a Text Using Dependency Trees	89
Try This	91
Using Context to Improve the Ticket-Booking Chatbot	91
Making a Smarter Chatbot by Finding Proper Modifiers	94
Summary	95

7 VISUALIZATIONS 97

Getting Started with spaCy’s Built-In Visualizers	98
displaCy Dependency Visualizer	98
displaCy Named Entity Visualizer	99
Visualizing from Within spaCy	100
Visualizing Dependency Parsing	100
Try This	102
Sentence-by-Sentence Visualizations	102
Customizing Your Visualizations with the Options Argument	103
Using Dependency Visualizer Options	103
Try This	104
Using Named Entity Visualizer Options	104
Exporting a Visualization to a File	106
Using displaCy to Manually Render Data	107
Formatting the Data	108
Try This	108
Summary	109

8 INTENT RECOGNITION 111

Extracting the Transitive Verb and Direct Object for Intent Recognition	112
Obtaining the Transitive Verb/Direct Object Pair	113
Extracting Multiple Intents with token.conjuncts	113

Try This	115
Using Word Lists to Extract the Intent	115
Finding the Meanings of Words Using Synonyms and Semantic Similarity	117
Recognizing Synonyms Using Predefined Lists	117
Try This	119
Recognizing Implied Intents Using Semantic Similarity	120
Try This	121
Extracting Intent from a Sequence of Sentences	122
Walking the Dependency Structures of a Discourse	122
Replacing Proforms with Their Antecedents	122
Try This	125
Summary	125

9

STORING USER INPUT IN A DATABASE **127**

Converting Unstructured Data into Structured Data	128
Extracting Data into Interchange Formats	129
Moving Application Logic to the Database	130
Building a Database-Powered Chatbot	132
Gathering the Data and Building a JSON Object	132
Converting Number Words to Numbers	133
Preparing Your Database Environment	135
Sending Data to the Underlying Database	137
When a User’s Request Doesn’t Contain Enough Information	138
Try This	139
Summary	139

10

TRAINING MODELS **141**

Training a Model’s Pipeline Component	142
Training the Entity Recognizer	143
Deciding Whether You Need to Train the Entity Recognizer	143
Creating Training Examples	144
Automating the Example Creation Process	144
Disabling the Other Pipeline Components	146
The Training Process	146
Evaluating the Updated Recognizer	148
Creating a New Dependency Parser	149
Custom Syntactic Parsing to Understand User Input	149
Deciding on Types of Semantic Relations to Use	150
Creating Training Examples	150
Training the Parser	152
Testing Your Custom Parser	152
Try This	153
Summary	153

11		
DEPLOYING YOUR OWN CHATBOT		155
How Implementing and Deploying a Chatbot Works		156
Using Telegram as a Platform for Your Bot		157
Creating a Telegram Account and Authorizing Your Bot		157
Getting Started with the python-telegram-bot Library		158
Using the telegram.ext Objects		159
Creating a Telegram Chatbot That Uses spaCy		160
Expanding the Chatbot		161
Holding the State of the Current Chat		162
Putting All the Pieces Together		163
Try This		167
Summary		167
12		
IMPLEMENTING WEB DATA AND PROCESSING IMAGES		169
How It Works		170
Making Your Bot Find Answers to Questions from Wikipedia		170
Determining What the Question Is About		171
Try This		174
Using Wikipedia to Answer User Questions		174
Try This		175
Reacting to Images Sent in a Chat		175
Generating Descriptive Tags for Images Using Clarifai		176
Using Tags to Generate Text Responses to Images		178
Putting All the Pieces Together in a Telegram Bot		178
Importing the Libraries		178
Writing the Helper Functions		178
Writing the Callback and main() Functions		180
Testing the Bot		181
Try This		182
Summary		183
LINGUISTIC PRIMER		185
Dependency Grammars vs. Phrase Structure Grammars		185
Common Grammar Concepts		187
Transitive Verbs and Direct Objects		187
Prepositional Objects		187
Modal Auxiliary Verbs		188
Personal Pronouns		188
INDEX		189