

Multilingual Natural Language Processing Applications

The background of the book cover is a jigsaw puzzle. The puzzle depicts a cityscape with a large, ornate building in the center, possibly a cathedral or a government building, surrounded by other structures and a body of water in the foreground. Several puzzle pieces are missing, particularly in the upper left and middle sections, revealing a plain white background behind the puzzle pieces.

From Theory to Practice

Edited by Daniel M. Bikel and Imed Zitouni

Register Your Book

at ibmpressbooks.com/ibmregister

Upon registration, we will send you electronic sample chapters from two of our popular IBM Press books. In addition, you will be automatically entered into a monthly drawing for a free IBM Press book.

Registration also entitles you to:

- Notices and reminders about author appearances, conferences, and online chats with special guests
- Access to supplemental material that may be available
- Advance notice of forthcoming editions
- Related book recommendations
- Information about special contests and promotions throughout the year
- Chapter excerpts and supplements of forthcoming books

Contact us

If you are interested in writing a book or reviewing manuscripts prior to publication, please write to us at:

Editorial Director, IBM Press
c/o Pearson Education
800 East 96th Street
Indianapolis, IN 46240

e-mail: IBMPress@pearsoned.com

Visit us on the Web: ibmpressbooks.com

Multilingual Natural Language Processing Applications: From Theory to Practice

Table of Contents

Contents

Preface

Acknowledgments

About the Authors

Part I: In Theory

Chapter 1 Finding the Structure of Words

1.1 Words and Their Components

1.2 Issues and Challenges

1.3 Morphological Models

1.4 Summary

Chapter 2 Finding the Structure of Documents

2.1 Introduction

2.2 Methods

2.3 Complexity of the Approaches

2.4 Performances of the Approaches

2.5 Features

2.6 Processing Stages

2.7 Discussion

2.8 Summary

Chapter 3 Syntax

3.1 Parsing Natural Language

3.2 Treebanks: A Data-Driven Approach to Syntax

Table of Contents

- 3.3 Representation of Syntactic Structure
- 3.4 Parsing Algorithms
- 3.5 Models for Ambiguity Resolution in Parsing
- 3.6 Multilingual Issues: What Is a Token?
- 3.7 Summary

Chapter 4 Semantic Parsing

- 4.1 Introduction
- 4.2 Semantic Interpretation
- 4.3 System Paradigms
- 4.4 Word Sense
- 4.5 Predicate-Argument Structure
- 4.6 Meaning Representation
- 4.7 Summary

Chapter 5 Language Modeling

- 5.1 Introduction
- 5.2 n-Gram Models
- 5.3 Language Model Evaluation
- 5.4 Parameter Estimation
- 5.5 Language Model Adaptation
- 5.6 Types of Language Models
- 5.7 Language-Specific Modeling Problems
- 5.8 Multilingual and Crosslingual Language Modeling
- 5.9 Summary

Chapter 6 Recognizing Textual Entailment

- 6.1 Introduction
- 6.2 The Recognizing Textual Entailment Task
- 6.3 A Framework for Recognizing Textual Entailment
- 6.4 Case Studies
- 6.5 Taking RTE Further
- 6.6 Useful Resources

Table of Contents

6.7 Summary

Chapter 7 Multilingual Sentiment and Subjectivity Analysis

7.1 Introduction

7.2 Definitions

7.3 Sentiment and Subjectivity Analysis on English

7.4 Word- and Phrase-Level Annotations

7.5 Sentence-Level Annotations

7.6 Document-Level Annotations

7.7 What Works, What Doesn't

7.8 Summary

Part II: In Practice

Chapter 8 Entity Detection and Tracking

8.1 Introduction

8.2 Mention Detection

8.3 Coreference Resolution

8.4 Summary

Chapter 9 Relations and Events

9.1 Introduction

9.2 Relations and Events

9.3 Types of Relations

9.4 Relation Extraction as Classification

9.5 Other Approaches to Relation Extraction

9.6 Events

9.7 Event Extraction Approaches

9.8 Moving Beyond the Sentence

9.9 Event Matching

9.10 Future Directions for Event Extraction

9.11 Summary

Chapter 10 Machine Translation

10.1 Machine Translation Today

Table of Contents

10.2 Machine Translation Evaluation

10.3 Word Alignment

10.4 Phrase-Based Models

10.5 Tree-Based Models

10.6 Linguistic Challenges

10.7 Tools and Data Resources

10.8 Future Directions

10.9 Summary

Chapter 11 Multilingual Information Retrieval

11.1 Introduction

11.2 Document Preprocessing

11.3 Monolingual Information Retrieval

11.4 CLIR

11.5 MLIR

11.6 Evaluation in Information Retrieval

11.7 Tools, Software, and Resources

11.8 Summary

Chapter 12 Multilingual Automatic Summarization

12.1 Introduction

12.2 Approaches to Summarization

12.3 Evaluation

12.4 How to Build a Summarizer

12.5 Competitions and Datasets

12.6 Summary

Chapter 13 Question Answering

13.1 Introduction and History

13.2 Architectures

13.3 Source Acquisition and Preprocessing

13.4 Question Analysis

13.5 Search and Candidate Extraction

Table of Contents

- 13.6 Answer Scoring
- 13.7 Crosslingual Question Answering
- 13.8 A Case Study
- 13.9 Evaluation
- 13.10 Current and Future Challenges
- 13.11 Summary and Further Reading

Chapter 14 Distillation

- 14.1 Introduction
- 14.2 An Example
- 14.3 Relevance and Redundancy
- 14.4 The Rosetta Consortium Distillation System
- 14.5 Other Distillation Approaches
- 14.6 Evaluation and Metrics
- 14.7 Summary

Chapter 15 Spoken Dialog Systems

- 15.1 Introduction
- 15.2 Spoken Dialog Systems
- 15.3 Forms of Dialog
- 15.4 Natural Language Call Routing
- 15.5 Three Generations of Dialog Applications
- 15.6 Continuous Improvement Cycle
- 15.7 Transcription and Annotation of Utterances
- 15.8 Localization of Spoken Dialog Systems
- 15.9 Summary

Chapter 16 Combining Natural Language Processing Engines

- 16.1 Introduction
- 16.2 Desired Attributes of Architectures for Aggregating Speech and NLP Engines
- 16.3 Architectures for Aggregation
- 16.4 Case Studies

Table of Contents

16.5 Lessons Learned

16.6 Summary

16.7 Sample UIMA Code

Index