

5. Categorizing and Tagging Words

5.1 Using a Tagger

5.2 Tagged Corpora

- 2.1 Representing Tagged Tokens
- 2.2 Reading Tagged Corpora
- 2.3 A Universal Part-of-Speech Tagset
- 2.4 Nouns
- 2.5 Verbs
- 2.6 Adjectives and Adverbs
- 2.7 Unsimplified Tags
- 2.8 Exploring Tagged Corpora

5.3 Mapping Words to Properties Using Python Dictionaries

- 3.1 Indexing Lists vs Dictionaries
- 3.2 Dictionaries in Python
- 3.3 Defining Dictionaries
- 3.4 Default Dictionaries
- 3.5 Incrementally Updating a Dictionary
- 3.6 Complex Keys and Values
- 3.7 Inverting a Dictionary

5.4 Automatic Tagging

- 4.1 The Default Tagger
- 4.2 The Regular Expression Tagger
- 4.3 The Lookup Tagger
- 4.4 Evaluation

5.5 N-Gram Tagging

- 5.1 Unigram Tagging
- 5.2 Separating the Training and Testing Data
- 5.3 General N-Gram Tagging
- 5.4 Combining Taggers
- 5.5 Tagging Unknown Words
- 5.6 Storing Taggers
- 5.7 Performance Limitations

5.6 Transformation-Based Tagging

5.7 How to Determine the Category of a Word

- 7.1 Morphological Clues
- 7.2 Syntactic Clues
- 7.3 Semantic Clues
- 7.4 New Words
- 7.5 Morphology in Part of Speech Tagsets

5.8 Summary

5.9 Further Reading

5.10 Exercises