

目 次

はじめに

第1章 言語研究とプログラミング	1
1.1. 本章の概要	1
1.2. なぜプログラミングが必要か	1
1.2.1. 大規模データによる言語研究	1
1.2.2. 機能を作る	3
1.2.3. 留意点	5
1.3. プログラミングの前に	6
1.4. 本書の活用	7
1.4.1. 構成	7
1.4.2. 専用ウェブサイトの活用	7
1.5. まとめ	9

第 I 部 プログラミングによるテキスト処理

第2章 テキストデータに親しもう	12
2.1. 本章の概要	12
2.2. テキストファイルの利点	12
2.3. テキストファイルを使う	14
2.3.1. テキストエディタとは	15
2.3.2. テキストエディタのインストール	16
2.3.3. テキストエディタによる検索例 1	18
2.3.4. テキストエディタによる検索例 2	19
2.3.5. テキストエディタによる置換	21
2.4. 文字コードと改行コード	23
2.4.1. 文字コードとは	23
2.4.2. 文字コードを調べる	24

2.4.3. 文字コードを変更する	25
2.4.4. 改行コードとは	27
2.5. まとめ	28
第3章 正規表現	29
3.1. 本章の概要	29
3.2. 正規表現とは	29
3.3. 正規表現の書き方	30
3.3.1. 「?」について	31
3.3.2. 「.」について	31
3.3.3. 「+」について	33
3.3.4. 「*」(アスタリスク)について	33
3.3.5. 「[]」(ブラケット)について	34
3.3.6. 「 」(パイプ)について	36
3.3.7. 「^」(キャレット)と「\$」(ドル)について	37
3.3.8. 後方参照	38
3.3.9. 組み合わせ	39
3.4. 正規表現を用いた置換	41
3.5. まとめ	44
3.6. 章末問題	44

第 II 部 Python の基本

第4章 Python に触れてみよう	48
4.1. 本章の概要	48
4.2. なぜ Python か	48
4.3. Python のインストール	49
4.4. Python の起動	51
4.5. Python で計算する	52
4.5.1. Python を電卓として使ってみる	52
4.5.2. 変数	53
4.6. Python で文字列を扱う	55
4.6.1. 文字列の 1 文字目が何か調べる：インデクシング	56
4.6.2. 文字数を調べる：len() 関数	58

4.6.3. 数値と文字列の区別	59
4.6.4. データを文字列に変換する：str() 関数	60
4.6.5. データを数値に変換する：int() 関数	60
4.7. Python を終了する	61
4.8. まとめ	62
4.9. 章末問題	62
第5章 Python でファイルの内容を表示してみよう	64
5.1. 本章の概要	64
5.2. Python のプログラムを保存・実行する	64
5.2.1. エディタでプログラムを書く	65
5.2.2. プログラムを実行するための準備	66
5.2.3. プログラムの実行	68
5.2.4. エラーが出てしまったら	69
5.3. プログラムにコメントを書く	72
5.3.1. Python スクリプトの中に日本語で書くには	72
5.4. 実行結果をファイルに保存する：リダイレクト	74
5.5. 各行の解説	75
5.6. まとめ	77
5.7. 章末問題	77
第6章 Python で検索しよう：条件分岐	78
6.1. 本章の概要	78
6.2. プログラムを条件分岐させる	78
6.3. 文字列に関する条件式	81
6.3.1. 文字列が含まれるかどうかを調べる：in 演算子	82
6.3.2. startswith() と endswith() 関数	83
6.4. and, or, not	83
6.4.1. 否定：not	84
6.4.2. かつ：and	84
6.4.3. または：or	85
6.5. else と elif	86
6.6. if の実例	88
6.6.1. ある条件を満たす行だけ表示する	88
6.6.2. 大文字小文字を区別せず検索する	89

6.6.3. 何も書かれていない行を削除する	91
6.7. まとめ	92
6.8. 章末問題	92
第7章 繰り返し処理を覚えよう：ループ	94
7.1. 本章の概要	94
7.2. ループの基本	94
7.3. ループの制御	97
7.3.1. 必要のない行をスキップする：continue	97
7.3.2. ループを中断する：break	99
7.4. ループの応用例	101
7.4.1. ファイルに行番号を振る：カウンタ	101
7.4.2. ファイルの最初の10行だけ表示する：カウンタ	103
7.4.3. キーワードがあったことを覚えておく：フラグ	104
7.4.4. どの行にも「ない」ことを確かめる	106
7.5. まとめ	108
7.6. 章末問題	109
第8章 単語の一覧表を作ろう：リスト	111
8.1. 本章の概要	111
8.2. リストの基本	111
8.2.1. リストのインデクシングとスライシング	113
8.2.2. リストへの追加	115
8.2.3. リストの並べ替え	117
8.3. リストとループ	119
8.4. ファイルの行を並べ替える	121
8.5. 出現した単語のリストを作る	123
8.5.1. 文字列とリストの変換：split と join	123
8.5.2. 出現した単語のリストを作る	126
8.5.3. 句読点などの処理	129
8.6. 表形式のデータを処理する	130
8.6.1. 発音辞書を使って特定の発音をもつ単語のみ表示する	130
8.7. まとめ	134
8.8. 章末問題	135

第9章 頻度表を作ろう：ディクショナリ	136
9.1. 本章の概要	136
9.2. ディクショナリの基本	136
9.2.1. ディクショナリの項目を1つ1つ処理する	139
9.3. ディクショナリで頻度表を作ろう	141
9.3.1. 頻度表をアルファベット順に並べ替える	144
9.3.2. 頻度表を頻度順に並べ替える	144
9.4. 発音辞書をディクショナリとして読み込んで利用する	145
9.5. まとめ	149
9.6. 章末問題	149
第10章 ファイル操作	151
10.1. 本章の概要	151
10.2. ファイル入出力の仕方のいろいろ	151
10.2.1. 入力ファイル名を実行時に指定する	152
10.2.2. ファイル名を指定して書き込む	155
10.2.3. 検索語をコマンドラインで指定できるようにする	156
10.3. 複数あるファイルを一気に処理する	158
10.3.1. Python でファイル一覧を表示する	158
10.3.2. フォルダ内の全ファイルの内容を表示する	160
10.3.3. ファイル名付きで全ファイルを表示する	161
10.3.4. フォルダ内の全ファイルから検索する	162
10.3.5. フォルダ内の全ファイルを書き換える	163
10.4. まとめ	166
10.5. 章末問題	167
第11章 Python で正規表現を使ってみよう	168
11.1. 本章の概要	168
11.2. Python での正規表現検索の基本	168
11.2.1. マッチしたかどうかで条件分岐する	169
11.2.2. マッチした行を一覧表示する	170
11.2.3. マッチした単語を集計する	171
11.2.4. マッチした部分をリストにする	174
11.3. 置換	175
11.3.1. 置換機能を用いて検索結果を強調する	176

11.4.	まとめ	176
11.5.	章末問題	177

第 III 部 Python の応用：日本語の処理を中心に

第 12 章	形態素解析プログラムを活用しよう	180
12.1.	本章の概要	180
12.2.	形態素解析プログラムとは	180
12.3.	形態素解析プログラムを導入する	181
12.3.1.	ソフトウェアのダウンロード	182
12.3.2.	ソフトウェアのインストール	184
12.4.	形態素解析プログラムを使う	186
12.4.1.	第一工程：入力データを指定する	186
12.4.2.	第二工程：解析オプションを指定する	187
12.4.3.	第三工程：出力方法を指定する	188
12.5.	形態素解析の留意点	191
12.5.1.	処理の難しいデータ	191
12.5.2.	形態素解析の単位	192
12.6.	まとめ	194
12.7.	章末問題	195
第 13 章	Python で日本語を扱おう	196
13.1.	本章の概要	196
13.2.	文字コードを指定してファイルを読み書きする	196
13.3.	日本語の頻度表を作る	198
13.3.1.	動詞のみ集計する	202
13.4.	まとめ	204
13.5.	章末問題	204
第 14 章	Python で KWIC 検索を作ろう	206
14.1.	本章の概要	206
14.2.	KWIC とは	206
14.3.	KWIC 検索を作る	207

14.3.1.	データを読み込む	208
14.3.2.	range() を使ったループによる検索	213
14.3.3.	文脈を表示する	216
14.3.4.	KWIC 検索プログラムの完成	219
14.4.	まとめ	222
14.5.	章末問題	222
第 15 章	コロケーション検索を作ろう	223
15.1.	本章の概要	223
15.2.	コロケーション検索プログラムの作成	223
15.2.1.	「条件 1」について	224
15.2.2.	「条件 2」について	225
15.2.3.	「条件 3」について	225
15.3.	発展	228
15.3.1.	複数のテキストファイルに適用する	228
15.3.2.	汎用性のあるプログラムのために	232
15.3.3.	コロケーション KWIC プログラムの完成	237
15.4.	プログラムの設定変更例	239
15.4.1.	出現文脈を広げる	239
15.4.2.	他のコロケーション検索 1	240
15.4.3.	他のコロケーション検索 2	240
15.4.4.	他のコロケーション検索 3	240
15.5.	まとめ	241
15.6.	章末問題	241
補遺	Mac をお使いの方のための情報	243
1.	データのダウンロード	243
2.	テキストエディタについて	244
3.	Python のインストール	248
4.	Python プログラムの実行	251
5.	形態素解析	254
文献紹介		255

機能一覽 259

索 引 275