

『Python ではじめるテキストアナリティクス入門』

第1刷用正誤表

(2022年6月24日現在)

このたびは、標記書籍をお買い求めいただき誠にありがとうございました。

標記書籍に誤りがございました。訂正し、深くお詫び申し上げます。

頁数	位置	誤	正
49	下から11行目	Pyhon	Python
53	1つめの「Windows PowerShell (管理者)」	wsl —install	wsl --install
113	プログラムリスト6.4 analyze_cooccurrency_miyajima.py 9-10行目	include_pos = ('NOUN', 'VERB', 'ADJ') + ('PROPN', 'ADV') stopwords = ('する', 'ある', 'ない', 'いう', 'もの', 'こと', 'よう', 'なる', 'ほう') + ('いる', 'くる', '行く', '思う')	include_pos = ('NOUN', 'PROPN', 'VERB', 'ADJ') stopwords = ('する', 'ある', 'ない', 'いう', 'もの', 'こと', 'よう', 'なる', 'ほう', 'いる', 'くる', 'お', 'つ', 'おる', 'とき', 'しまう', 'いく', 'みる', 'ため', 'ところ', '際', '他', '時', '中', '方', '目', '回', '年', '点', '前', '後', '思う', '行く')
	プログラムリスト 6.4 analyze_cooccurrency_miyajima.py 27 行目	count_model = CountVectorizer(token_pattern=f'\b\\w{token_length}\b')	token_pattern=f'\\b\\w{token_length}\\b' count_model = CountVectorizer(token_pattern=token_pattern)
114	プログラムリスト 6.4 analyze_cooccurrency_miyajima.py 94 行目	weight_cutoff = 0.05 G = create_network(words, word_counts, Xc, weight_cutoff)	G = create_network(words, word_counts, Xc, 0.015)
119	プログラムリスト 6.8 do_ner3.py		11 行目を 24 行目の前へ移動 (23 行目となる)
120-	プログラムリスト6.9		13 行目を 26 行目の前へ移動 (25 行目となる)

121	count_food1.py		
121	プログラムリスト6.9 count_food1.py 34行目	if ent.label_ == 'Dish' or ent.label_ == 'Food_other':	if ent.label_ == 'Dish':
122	プログラムリスト6.10 count_food2.py		13行目を26行目の前へ移動(25行目となる)
123	プログラムリスト6.10 count_food2.py 34行目	if ent.label_ == 'Dish' or ent.label_ == 'Food_other':	if ent.label_ == 'Dish':
123	プログラムリスト6.10 count_food2.py 43行目	揚げもみじ饅頭	揚げ紅葉饅頭
130	プログラムリスト6.13 analyze_cooccurrence_themepark.py 10行目	stopwords = ('する', 'ある', 'ない', 'いう', 'もの', 'こと', 'よう', 'なる', 'ほう') + ('いる', 'くる', '行く', '思う')	stopwords = ('する', 'ある', 'ない', 'いう', 'もの', 'こと', 'よう', 'なる', 'ほう', 'いる', 'くる', 'お', 'つ', 'おる', 'とき', 'しまう', 'いく', 'みる', 'やる', 'ため', 'ところ', '際', '他', '時', '中', '方', '回', '目', '年', '月', '日', '分', '事', '等', '内', '間', '半', '頃', '前', '後', '今回', '入る', '思う', '行う', '行く', '行ける', '行なう', '来る', '思う', '見る', '見える', '見れる', '観る', '訪れる')
	プログラムリスト6.13 analyze_cooccurrence_themepark.py 16-19行目	# 旅行の時期を指定 ここでは2014年度から2018年度とする from_date = pd.to_datetime("2014-04") to_date = pd.to_datetime("2019-03") df = df[(df['旅行の時期'] >= from_date) & (df['旅行の時期'] <= to_date)]	# 旅行の時期が欠損している口コミを削除 df = df.dropna(subset=['旅行の時期']) # 旅行の時期を年度に直し, 「年度」という項目を追加 df['年度'] = df['旅行の時期'].apply(lambda x: x.year if x.month >= 4 else x.year-1) # 旅行の時期を指定 df = df[(df['年度'] >= 2014) & (df['年度'] <= 2018)]

131	プログラムリスト 6.13 analyze_cooccurrence_themepark.py 26-27行目	weight_cutoff = 0.02 G = create_network(words, word_counts, Xc, weight_cutoff)	G = create_network(words, word_counts, Xc, 0.020)
177	プログラムリスト 8.8 2ch_daily_emotions.py 37行目	ツイート本文	スレッドタイトル
213	10行目	特徴量とするよりは	特徴量とするよりは