

5.1

課題 5_1text1 から text2 までの語彙数

text1: 17231

text2: 6403

text3: 2628

text4: 9201

text5: 5441

text6: 1855

text7: 11387

text8: 882

text9: 6349

5.2

課題 5_2 アルファベットの出現率

text1	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z		
0.082	0.018	0.024	0.04	0.123	0.022	0.022	0.066	0.069	0.001	0.008	0.045	0.024	0.069	0.073	0.018	0.002	0.055	0.067	0.092	0.028	0.009	0.023	0.001	0.018	0.001			
text2	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z		
0.077	0.015	0.024	0.043	0.127	0.023	0.018	0.061	0.07	0.002	0.005	0.039	0.028	0.073	0.08	0.015	0.001	0.063	0.062	0.086	0.028	0.011	0.024	0.002	0.022	0.0			
0.999	text3	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	
0.102	0.018	0.017	0.061	0.127	0.024	0.015	0.088	0.055	0.003	0.006	0.036	0.026	0.074	0.068	0.012	0.0	0.051	0.057	0.09	0.024	0.009	0.02	0.0	0.018	0.001			
text4	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z		
0.072	0.014	0.032	0.037	0.13	0.027	0.017	0.049	0.075	0.001	0.003	0.038	0.024	0.075	0.081	0.022	0.001	0.063	0.062	0.098	0.03	0.012	0.018	0.002	0.016	0.001			
text5	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z		
0.081	0.014	0.023	0.027	0.097	0.014	0.021	0.057	0.075	0.01	0.015	0.05	0.031	0.066	0.095	0.023	0.0	0.048	0.053	0.078	0.054	0.008	0.025	0.002	0.033	0.002			
text6	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z		
0.087	0.018	0.027	0.035	0.109	0.015	0.031	0.069	0.063	0.001	0.013	0.047	0.023	0.064	0.078	0.013	0.002	0.069	0.054	0.083	0.043	0.009	0.022	0.001	0.023	0.001			
text7	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z		
0.083	0.016	0.036	0.038	0.118	0.021	0.02	0.041	0.075	0.002	0.008	0.041	0.027	0.072	0.073	0.023	0.001	0.069	0.071	0.092	0.029	0.01	0.014	0.003	0.017	0.001			
text8	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z		
0.073	0.012	0.021	0.04	0.113	0.025	0.032	0.029	0.085	0.002	0.018	0.056	0.034	0.077	0.076	0.021	0.001	0.059	0.08	0.064	0.03	0.012	0.013	0.001	0.026	0.0			
text9	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z		
0.084	0.015	0.025	0.043	0.123	0.022	0.02	0.063	0.069	0.001	0.009	0.044	0.026	0.065	0.074	0.017	0.001	0.056	0.066	0.09	0.029	0.008	0.023	0.001	0.023	0.0			

5.3

課題 5_3 text1 から text9 までの出現率上位 5 番目までの 4 文字以上の単語

text 1	that		3085
	with		1722
	this		1394
	whale		1226
	from		1088
text 2	that		1385
	with		992
	have		819
	elinor		685
	which		593
text 3	unto		598
	that		520
	said		477
	with		293
	thou		284
text 4	that		1793
	have		1011
	which		1006

	with		958
	will		914
text 5			
	part		1022
	join		1021
	action		347
	that		284
	what		201
text 6			
	arthur		261
	that		106
	launcelot		101
	what		92
	knight		84
text 7			
	that		848
	t-1		806
	said		628
	with		398
	from		391
text 8			
	lady		88
	seeks		72
	with		46
	male		42
	looking		34
text 9			
	that		871
	with		561
	syme		519
	said		507
	they		318

5.4

課題 5_4 text1 から text9 までのストップワードの比率

text 1	0.4137045230600531
text 2	0.4714570266146805
text 3	0.4503618979537128
text 4	0.47647149141838624
text 5	0.29022439457898247
text 6	0.28042671067366065
text 7	0.302812984226628
text 8	0.17444010684199712
text 9	0.4391660526201725

5.5

課題 5_5 wagahaiha_nekodearu.txt に含まれる名詞の比率
形態素解析を行いすべての品詞をすべて足したものを分母とし、分子は名詞の数とした。

```
(mecab-python3      0.996.2 )
分母:212662
分子:60262
割合:0.283369854510914
```

```
Mecab(0.996)
分母:212662
分子:60263
割合:0.283374556808
```

5.6

課題 5_6 wagahaiha_nekodearu.txt に含まれる出現頻度上位 30 位までの名詞

1	位	の	1612	回	
2	位	事	1207	回	
3	位	もの	981	回	
4	位	君	973	回	
5	位	主人	933	回	
6	位	ん	707	回	
7	位	よう	697	回	
8	位	人	604	回	
9	位	一	561	回	
10	位	何	540	回	
11	位	吾輩	481	回	
12	位	これ	414	回	
13	位	それ	395	回	
14	位	時	344	回	
15	位	迷亭	343	回	
16	位	傍点	318	回	
17	位	ところ	315	回	
18	位	方	314	回	
19	位	三	311	回	—— 311 回
20	位	二	303	回	
21	位	上	297	回	
22	位	寒月	286	回	
23	位	そう	284	回	
24	位	顔	283	回	
25	位	先生	274	回	
26	位	人間	272	回	
27	位	僕	268	回	
28	位	さん	261	回	
29	位	気	251	回	

```
Mecab(0.996)
上と同様
```

5.7

課題 5_7 wagahaiha_nekodearu.txt に top30_jnoun が含まれる頻度

```
(mecab-python3      0.996.2 )
単語の総数: 60262
トップ30の名詞の数: 15388
```

top 30 の出現率の名詞が全体の単語に占める割合: 0.25535163121038135

Mecab(0.996)

単語の総数: 60263

トップ30の名詞の数: 15388

top 30 の出現率の名詞が全体の単語に占める割合: 0.25534739392330286