

```
In [1]: #影男のテキストデータからワードクラウドを生成. なお, ストップワードは以下の語と品詞と
#stopwords = ('する', 'ある', 'ない', 'いう', 'もの', 'こと', 'よう', 'なる', 'ほう')
#include_pos = ('NOUN', 'VERB', 'ADJ')
# ファイルを開く
with open("C:\Users\syuhe\Downloads\58486_ruby_68458kageotoko.txt", 'r', encoding='utf-8') as f:
    # ファイルの内容を読み込む
    text = f.read()

import re
import spacy
# 《》で囲まれたルビを除去
text = re.sub(r'《[^》]+》', '', text)

# [#]で囲まれた入力者注を除去
text = re.sub(r'#[^#]+', '', text)

# 冒頭18行と末尾14行を除去
lines = text.split('\n')
lines = lines[18:-14]
text = '\n'.join(lines)

nlp = spacy.load("ja_ginza")

chunk_size = 1000 # 適切なサイズに調整
chunks = [text[i:i+chunk_size] for i in range(0, len(text), chunk_size)]
# 各chunkに含まれる改行文字と全角スペースを削除
chunks = [chunk.replace('\n', '').replace('　', '') for chunk in chunks]

# ストップワードと品詞の指定
stopwords = {'する', 'ある', 'ない', 'いう', 'もの', 'こと', 'よう', 'なる', 'ほう'}
include_pos = {'NOUN', 'VERB', 'ADJ'}

filtered_words = []
```

```
In [2]: for chunk in chunks:
        doc = nlp(chunk)
        filtered_words.extend([token.text for token in doc if token.text not in stopwords and token.pos in include_pos])
```

```
In [3]: from wordcloud import WordCloud
import matplotlib.pyplot as plt

# ワードクラウドの生成
filtered_text = '\n'.join(filtered_words)

font_path = 'C:/Windows/Fonts/msgothic.ttc'

wordcloud = WordCloud(width=800, height=400, background_color='white', font_path=font_path)
```

```
In [4]: plt.figure(figsize=(10, 5))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off') # Turn off axis numbers and ticks
plt.show()
```

