附录：绘制一个中文的词云

2023 年 9 月 18 日

绘制一个中文的词云

1. 准备一个中文文本

如果要显示中文的词图，首先文本必须是中文的，例如路径'datasets/news_sample01.txt' 的中文新闻稿:

[1]:
```
f = open('datasets/news_sample01.txt', 'r', encoding='utf-8')
text = f.read()
text = text.replace('\n', ' ')  # 删除换行符
text[:30]
```

[1]:
```
' 人民币国际化稳步向前 今年一季度人民币跨境结算规模稳步扩大 '
```

2. 准备分词工具

这里使用 jieba 库来进行中文分词，如何使用 jieba，参加本章节下课件： .ipynb

[2]:
```
import jieba
words_list = jieba.cut(text)
words_list = list(words_list)
text = ' '.join(words_list)
```

```
Building prefix dict from the default dictionary …
Loading model from cache
/var/folders/7m/rgg1hhpj3yb1j0cf0c82yq1w0000gn/T/jieba.cache
Loading model cost 0.336 seconds.
Prefix dict has been built successfully.
```

3. 准备一个白底的图形

在网上下载图片，例如在搜索引擎上搜索“人民币 + 图标 + 白底”:

使用修改的图片的软件修改图片大小，在保证宽 * 高比不变的情况下，宽改成 1500pixels，避免图片过大或者过小。

读取准备好的图形，作为掩码的形状。然后使用 imageio 库来读取图片。

[3]:
```
import imageio as imageio
mask_img = imageio.imread('image/rmb.png')
print('图片大小: ', mask_img.shape)
```

```
/var/folders/7m/rgg1hhpj3yb1j0cf0c82yq1w0000gn/T/ipykernel_29103/3419627869.py:2
: DeprecationWarning: Starting with ImageIO v3 the behavior of this function
will switch to that of iio.v3.imread. To keep the current behavior (and make
this warning dissapear) use `import imageio.v2 as imageio` or call
`imageio.v2.imread` directly.
  mask_img = imageio.imread('image/rmb.png')
```

```
图片大小:  (1500, 1500, 4)
```

4. 准备颜色

以 CSS 颜色为例，

选择一个色系，例如’purple’,

[4]:
```
from wordcloud import get_single_color_func
color_func = get_single_color_func('purple')
```

5. 指定中文字体路径

接着，需要指定 WordCloud 类的参数 font_path 为一个中文字体路径。如何找到自己电脑中的中文字体路径呢？参照本章节 .ipynb

[5]:
```
from wordcloud import WordCloud
wc = WordCloud(
                font_path='/System/Library/Fonts/Supplemental/Songti.ttc',
                mask=mask_img,
```

```
                max_words=200,
                max_font_size=100,
                min_font_size=10,
                width=1500,
                height=1500,
                background_color="white",
                 color_func=color_func,
                repeat=True)
wc.generate(text)
```

[5]: <wordcloud.wordcloud.WordCloud at 0x10a784130>

[6]:
```
import matplotlib.pyplot as plt
plt.imshow(wc)
plt.axis("off")
plt.savefig("image/ciyun_cn.jpg", dpi=500) # 或者保存成.png .svg .pdf .eps 等
plt.show()
```