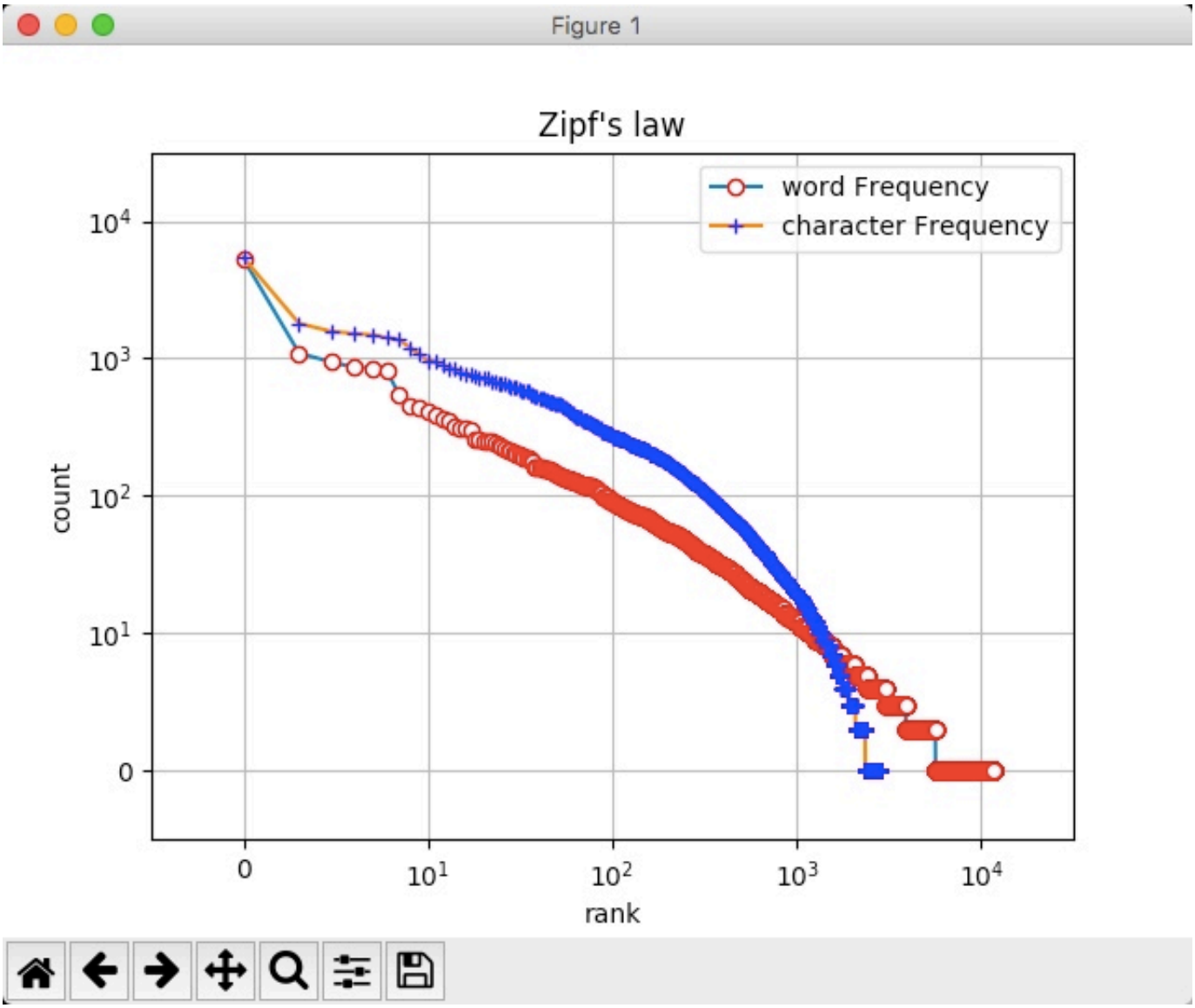
验证Zipf's law与正则匹配

需要库：jieba/matplotlib/openpyxl/regex/math/collections

- 选择小说--《皮肤的秘密》中文译本
- 首先对文本进行词频统计与字频统计，词频统计用到了jieba库分词
- 按照词(字)频顺序列出所有的词(字)及其出现次数
- 绘制排序-数量曲线，验证 Zipf's law ，这里用 Matplotlib.pyplot库画图
- 利用正则表达式查找文件中皮肤后相关的字词（2-4字）
- 把字频、词频、正则搜索数据写入excel

结果如下：



word	count	rank		word	count	rank		word	count	rank
的	5290	1		的	5395	1		医生	9	1
皮肤	1095	2		皮	1798	2		褶皱	7	2
会	940	3		是	1573	3		问题	7	3
在	862	4		有	1520	4		状况	6	4
是	851	5		一	1486	5		炎症	5	5
与	805	6		会	1417	6		来说	5	6
也	542	7		肤	1366	7		深处	5	7
都	452	8		不	1198	8		干燥	5	8
和	436	9		在	1066	9		的秘密	4	9
了	412	10		人	964	10		角质化	4	10
有	382	11		性	944	11		开裂	4	11
中	368	12		生	897	12		感染	4	12
我们	356	13		为	850	13		—大脑	3	13
人	320	14		与	837	14		在偷听	3	14
为	309	15		于	784	15		与光	3	15
对	308	16		时	776	16		的作用	3	16
上	301	17		体	768	17		海洛因	3	17
时	260	18		大	732	18		补救法	3	18
或	257	19		上	728	19		瘙痒	3	19
但	251	20		部	724	20		部位	3	20
就	249	21		们	724	21				
可以	249	22		来	677	22				
还	240	23		素	671	23				
您	232	24		过	658	24				
部位	226	25		能	651	25				
到	221	26		可	648	26				
能	214	27		发	642	27				