



南京工业大学  
NANJING TECH  
UNIVERSITY

# 用户数据采集与关联分析

## (结课作业)

吴志祥

18205185639

1030624832@qq.com



## 第二讲 词频统计

1. 基于CNKI数据库统计分析2014-2024年（近10年），“信息资源管理”或“网络营销”或其他你感兴趣的主题变化趋势。

基于 CNKI 数据库 2014-2024 年文献统计，信息资源管理与网络营销领域均呈现技术驱动、需求导向的演化特征。信息资源管理从传统图书馆、档案数字化管理起步，经大数据时代的数据治理与智慧化转型，近年聚焦 **AGI** 应用、数据要素市场化等主题，政策导向与技术革新共同推动其从“资源有序化”迈向“价值赋能”。网络营销则从早期的渠道拓展与流量争夺，转向私域运营、数据驱动的精细化运营，如今依托 **AIGC**、跨平台联动实现技术赋能，且愈发注重与实体经济融合，凸显“长期价值创造”的核心逻辑。两大领域均呈现跨学科融合趋势，前者侧重公共服务与理论构建，后者贴近市场实践与商业价值，技术迭代始终是贯穿十年的核心演进动力。



## 第二讲 词频统计

2. 完成ppt中的程序运行，包括全文词频统计，词频统计；

**本文档程序基于jleba分词的全文本词频统计**

**全文本词频统计的步骤**

- 打乱数据
- 分词
- 使用倒排索引 (像谷歌的索引)
- 计算词频, 统计词频的逆频率
- 排序
- 输出 tf-idf 矩阵
- 输出词频矩阵和 tf-idf 矩阵

```

10 [1]: from random import
11
12 [2]: #defining how iterations because normal iter packages is not available
13 #normal already available: jleba is a more robust algorithm using gettfidf package (0.4.0.0)
14
15 [3]: import jleba
16
17 [4]: data = [" ".join([word for word in word.split() if word in set(w for w in stopwords.words('english'))]) for word in data]
18 # 去除停用词, 只保留有效词, 返回清洗后的数据
19
20 [5]: [jleba.tokenize_with_weights(word) for word in data] # 返回分词和权重矩阵
21
22 Building graph first from the default dictionary
23 Loading words from jleba (C:\Users\user\AppData\Local\Temp\jleba_index
24 Building words list in 128 seconds
25 jleba.index has been built (done:0.000151)
26
27 [6]: words = list(jleba.tokenize_with_weights(data)) # 返回分词和权重矩阵
28
29 [7]: words
30
31 words
32
33 [8]: words
34
35 words
36
37 [9]: words
38
39 words
40
41 [10]: words
42
43 words
44
45 [11]: words
46
47 words
48
49 [12]: words
50
51 words
52
53 [13]: words
54
55 words
56
57 [14]: words
58
59 words
60
61 [15]: words
62
63 words
64
65 [16]: words
66
67 words
68
69 [17]: words
70
71 words
72
73 [18]: words
74
75 words
76
77 [19]: words
78
79 words
80
81 [20]: words
82
83 words
84
85 [21]: words
86
87 words
88
89 [22]: words
90
91 words
92
93 [23]: words
94
95 words
96
97 [24]: words
98
99 words
100
101 [25]: words
102
103 words
104
105 [26]: words
106
107 words
108
109 [27]: words
110
111 words
112
113 [28]: words
114
115 words
116
117 [29]: words
118
119 words
120
121 [30]: words
122
123 words
124
125 [31]: words
126
127 words
128
129 [32]: words
130
131 words
132
133 [33]: words
134
135 words
136
137 [34]: words
138
139 words
140
141 [35]: words
142
143 words
144
145 [36]: words
146
147 words
148
149 [37]: words
150
151 words
152
153 [38]: words
154
155 words
156
157 [39]: words
158
159 words
160
161 [40]: words
162
163 words
164
165 [41]: words
166
167 words
168
169 [42]: words
170
171 words
172
173 [43]: words
174
175 words
176
177 [44]: words
178
179 words
180
181 [45]: words
182
183 words
184
185 [46]: words
186
187 words
188
189 [47]: words
190
191 words
192
193 [48]: words
194
195 words
196
197 [49]: words
198
199 words
200
201 [50]: words
202
203 words
204
205 [51]: words
206
207 words
208
209 [52]: words
210
211 words
212
213 [53]: words
214
215 words
216
217 [54]: words
218
219 words
220
221 [55]: words
222
223 words
224
225 [56]: words
226
227 words
228
229 [57]: words
230
231 words
232
233 [58]: words
234
235 words
236
237 [59]: words
238
239 words
240
241 [60]: words
242
243 words
244
245 [61]: words
246
247 words
248
249 [62]: words
250
251 words
252
253 [63]: words
254
255 words
256
257 [64]: words
258
259 words
260
261 [65]: words
262
263 words
264
265 [66]: words
266
267 words
268
269 [67]: words
270
271 words
272
273 [68]: words
274
275 words
276
277 [69]: words
278
279 words
280
281 [70]: words
282
283 words
284
285 [71]: words
286
287 words
288
289 [72]: words
290
291 words
292
293 [73]: words
294
295 words
296
297 [74]: words
298
299 words
300
301 [75]: words
302
303 words
304
305 [76]: words
306
307 words
308
309 [77]: words
310
311 words
312
313 [78]: words
314
315 words
316
317 [79]: words
318
319 words
320
321 [80]: words
322
323 words
324
325 [81]: words
326
327 words
328
329 [82]: words
330
331 words
332
333 [83]: words
334
335 words
336
337 [84]: words
338
339 words
340
341 [85]: words
342
343 words
344
345 [86]: words
346
347 words
348
349 [87]: words
350
351 words
352
353 [88]: words
354
355 words
356
357 [89]: words
358
359 words
360
361 [90]: words
362
363 words
364
365 [91]: words
366
367 words
368
369 [92]: words
370
371 words
372
373 [93]: words
374
375 words
376
377 [94]: words
378
379 words
380
381 [95]: words
382
383 words
384
385 [96]: words
386
387 words
388
389 [97]: words
390
391 words
392
393 [98]: words
394
395 words
396
397 [99]: words
400
401 words
402
403 [100]: words
404
405 words
406
407 [101]: words
408
409 words
410
411 [102]: words
412
413 words
414
415 [103]: words
416
417 words
418
419 [104]: words
420
421 words
422
423 [105]: words
424
425 words
426
427 [106]: words
428
429 words
430
431 [107]: words
432
433 words
434
435 [108]: words
436
437 words
438
439 [109]: words
440
441 words
442
443 [110]: words
444
445 words
446
447 [111]: words
448
449 words
450
451 [112]: words
452
453 words
454
455 [113]: words
456
457 words
458
459 [114]: words
460
461 words
462
463 [115]: words
464
465 words
466
467 [116]: words
468
469 words
470
471 [117]: words
472
473 words
474
475 [118]: words
476
477 words
478
479 [119]: words
480
481 words
482
483 [120]: words
484
485 words
486
487 [121]: words
488
489 words
490
491 [122]: words
492
493 words
494
495 [123]: words
496
497 words
498
499 [124]: words
500
501 words
502
503 [125]: words
504
505 words
506
507 [126]: words
508
509 words
510
511 [127]: words
512
513 words
514
515 [128]: words
516
517 words
518
519 [129]: words
520
521 words
522
523 [130]: words
524
525 words
526
527 [131]: words
528
529 words
530
531 [132]: words
532
533 words
534
535 [133]: words
536
537 words
538
539 [134]: words
540
541 words
542
543 [135]: words
544
545 words
546
547 [136]: words
548
549 words
550
551 [137]: words
552
553 words
554
555 [138]: words
556
557 words
558
559 [139]: words
560
561 words
562
563 [140]: words
564
565 words
566
567 [141]: words
568
569 words
570
571 [142]: words
572
573 words
574
575 [143]: words
576
577 words
578
579 [144]: words
580
581 words
582
583 [145]: words
584
585 words
586
587 [146]: words
588
589 words
590
591 [147]: words
592
593 words
594
595 [148]: words
596
597 words
598
599 [149]: words
600
601 words
602
603 [150]: words
604
605 words
606
607 [151]: words
608
609 words
610
611 [152]: words
612
613 words
614
615 [153]: words
616
617 words
618
619 [154]: words
620
621 words
622
623 [155]: words
624
625 words
626
627 [156]: words
628
629 words
630
631 [157]: words
632
633 words
634
635 [158]: words
636
637 words
638
639 [159]: words
640
641 words
642
643 [160]: words
644
645 words
646
647 [161]: words
648
649 words
650
651 [162]: words
652
653 words
654
655 [163]: words
656
657 words
658
659 [164]: words
660
661 words
662
663 [165]: words
664
665 words
666
667 [166]: words
668
669 words
670
671 [167]: words
672
673 words
674
675 [168]: words
676
677 words
678
679 [169]: words
680
681 words
682
683 [170]: words
684
685 words
686
687 [171]: words
688
689 words
690
691 [172]: words
692
693 words
69
```

## 第二讲 词频统计

2. 完成ppt中的程序运行，包括全文词频统计，词频统计；

**本文档程序基于jleba分词的全文本词频统计**

**全文本词频统计的步骤**

- 打乱数据
- 分词
- 使用倒排索引 (像谷歌的索引)
- 计算词频, 统计词频的逆频率
- 排序
- 输出 tf-idf 矩阵
- 输出词频矩阵和 tf-idf 矩阵

```

10 [1]: from random import
11
12 [2]: #defining how iterations because normal iter packages is not available
13 #normal already available: jleba is a more robust algorithm using gettfidf package (0.4.0.0)
14
15 [3]: import jleba
16
17 [4]: data = [" ".join([word for word in word.split() if word in set(w for w in stopwords.words('english'))]) for word in data]
18 # 去除停用词, 只保留有效词, 返回清洗后的数据
19
20 [5]: [jleba.tokenize_with_weights(word) for word in data] # 返回分词和权重矩阵
21
22 Building graph first from the default dictionary
23 Loading words from jleba (C:\Users\user\AppData\Local\Temp\jleba_index
24 Building words list in 128 seconds
25 jleba.index has been built (done:0.000151)
26
27 [6]: words = list(jleba.tokenize_with_weights(data)) # 返回分词和权重矩阵
28
29 [7]: words
30
31 words
32 1. '词频'
33 2. '词频'
34 3. '词频'
35 4. '词频'
36 5. '词频'
37 6. '词频'
38 7. '词频'
39 8. '词频'
40 9. '词频'
41 10. '词频'
42 11. '词频'
43 12. '词频'
44 13. '词频'
45 14. '词频'
46 15. '词频'
47 16. '词频'
48 17. '词频'
49 18. '词频'
50 19. '词频'
51 20. '词频'
52 21. '词频'
53 22. '词频'
54 23. '词频'
55 24. '词频'
56 25. '词频'
57 26. '词频'
58 27. '词频'
59 28. '词频'
60 29. '词频'
61 30. '词频'
62 31. '词频'
63 32. '词频'
64 33. '词频'
65 34. '词频'
66 35. '词频'
67 36. '词频'
68 37. '词频'
69 38. '词频'
70 39. '词频'
71 40. '词频'
72 41. '词频'
73 42. '词频'
74 43. '词频'
75 44. '词频'
76 45. '词频'
77 46. '词频'
78 47. '词频'
79 48. '词频'
80 49. '词频'
81 50. '词频'
82 51. '词频'
83 52. '词频'
84 53. '词频'
85 54. '词频'
86 55. '词频'
87 56. '词频'
88 57. '词频'
89 58. '词频'
90 59. '词频'
91 60. '词频'
92 61. '词频'
93 62. '词频'
94 63. '词频'
95 64. '词频'
96 65. '词频'
97 66. '词频'
98 67. '词频'
99 68. '词频'
100 69. '词频'
101 70. '词频'
102 71. '词频'
103 72. '词频'
104 73. '词频'
105 74. '词频'
106 75. '词频'
107 76. '词频'
108 77. '词频'
109 78. '词频'
110 79. '词频'
111 80. '词频'
112 81. '词频'
113 82. '词频'
114 83. '词频'
115 84. '词频'
116 85. '词频'
117 86. '词频'
118 87. '词频'
119 88. '词频'
120 89. '词频'
121 90. '词频'
122 91. '词频'
123 92. '词频'
124 93. '词频'
125 94. '词频'
126 95. '词频'
127 96. '词频'
128 97. '词频'
129 98. '词频'
130 99. '词频'
131 100. '词频'
132 101. '词频'
133 102. '词频'
134 103. '词频'
135 104. '词频'
136 105. '词频'
137 106. '词频'
138 107. '词频'
139 108. '词频'
140 109. '词频'
141 110. '词频'
142 111. '词频'
143 112. '词频'
144 113. '词频'
145 114. '词频'
146 115. '词频'
147 116. '词频'
148 117. '词频'
149 118. '词频'
150 119. '词频'
151 120. '词频'
152 121. '词频'
153 122. '词频'
154 123. '词频'
155 124. '词频'
156 125. '词频'
157 126. '词频'
158 127. '词频'
159 128. '词频'
160 129. '词频'
161 130. '词频'
162 131. '词频'
163 132. '词频'
164 133. '词频'
165 134. '词频'
166 135. '词频'
167 136. '词频'
168 137. '词频'
169 138. '词频'
170 139. '词频'
171 140. '词频'
172 141. '词频'
173 142. '词频'
174 143. '词频'
175 144. '词频'
176 145. '词频'
177 146. '词频'
178 147. '词频'
179 148. '词频'
180 149. '词频'
181 150. '词频'
182 151. '词频'
183 152. '词频'
184 153. '词频'
185 154. '词频'
186 155. '词频'
187 156. '词频'
188 157. '词频'
189 158. '词频'
190 159. '词频'
191 160. '词频'
192 161. '词频'
193 162. '词频'
194 163. '词频'
195 164. '词频'
196 165. '词频'
197 166. '词频'
198 167. '词频'
199 168. '词频'
200 169. '词频'
201 170. '词频'
202 171. '词频'
203 172. '词频'
204 173. '词频'
205 174. '词频'
206 175. '词频'
207 176. '词频'
208 177. '词频'
209 178. '词频'
210 179. '词频'
211 180. '词频'
212 181. '词频'
213 182. '词频'
214 183. '词频'
215 184. '词频'
216 185. '词频'
217 186. '词频'
218 187. '词频'
219 188. '词频'
220 189. '词频'
221 190. '词频'
222 191. '词频'
223 192. '词频'
224 193. '词频'
225 194. '词频'
226 195. '词频'
227 196. '词频'
228 197. '词频'
229 198. '词频'
230 199. '词频'
231 200. '词频'
232 201. '词频'
233 202. '词频'
234 203. '词频'
235 204. '词频'
236 205. '词频'
237 206. '词频'
238 207. '词频'
239 208. '词频'
240 209. '词频'
241 210. '词频'
242 211. '词频'
243 212. '词频'
244 213. '词频'
245 214. '词频'
246 215. '词频'
247 216. '词频'
248 217. '词频'
249 218. '词频'
250 219. '词频'
251 220. '词频'
252 221. '词频'
253 222. '词频'
254 223. '词频'
255 224. '词频'
256 225. '词频'
257 226. '词频'
258 227. '词频'
259 228. '词频'
260 229. '词频'
261 230. '词频'
262 231. '词频'
263 232. '词频'
264 233. '词频'
265 234. '词频'
266 235. '词频'
267 236. '词频'
268 237. '词频'
269 238. '词频'
270 239. '词频'
271 240. '词频'
272 241. '词频'
273 242. '词频'
274 243. '词频'
275 244. '词频'
276 245. '词频'
277 246. '词频'
278 247. '词频'
279 248. '词频'
280 249. '词频'
281 250. '词频'
282 251. '词频'
283 252. '词频'
284 253. '词频'
285 254. '词频'
286 255. '词频'
287 256. '词频'
288 257. '词频'
289 258. '词频'
290 259. '词频'
291 260. '词频'
292 261. '词频'
293 262. '词频'
294 263. '词频'
295 264. '词频'
296 265. '词频'
297 266. '词频'
298 267. '词频'
299 268. '词频'
300 269. '词频'
301 270. '词频'
302 271. '词频'
303 272. '词频'
304 273. '词频'
305 274. '词频'
306 275. '词频'
307 276. '词频'
308 277. '词频'
309 278. '词频'
310 279. '词频'
311 280. '词频'
312 281. '词频'
313 282. '词频'
314 283. '词频'
315 284. '词频'
316 285. '词频'
317 286. '词频'
318 287. '词频'
319 288. '词频'
320 289. '词频'
321 290. '词频'
322 291. '词频'
323 292. '词频'
324 293. '词频'
32
```

这部分是指定类型的频次统计

我们采用功勋科学家黄旭华院士的传记序言文本

### 1 词频统计

统计文本中，指定词汇的次数，步骤：

- 直接定义需要统计的词汇（term）即直接定义列表
- 打开文本
- 采用文本跟词汇匹配的方式，进行词汇统计（注：没有采用采用分词的方式）
- 用字典存储：词汇-频次
- 画图展示

2.

```
In [1]: # 文件的打开与关闭
In [2]: #f_name = open('name.txt',encoding = 'GB18030') #使用mac的小伙伴，需要耐心调试下编码GB18030
In [3]: #f_name = open('name.txt')
In [4]: #data_name = f_name.read()
In [5]: #data_name[:70]
In [6]: #print(data_name[:50])
In [7]: #f_name.close()
In [8]: # 将文本转化为列表
In [9]: #names = data_name.split(',') # split一下names就是列表
In [10]: # 以上是课程演示里面的方法，特定的需要统计的词汇放置在txt文本中
# 我们需要统计的词汇很少，就直接定义需要统计的词汇
In [11]: terms = ['黄旭华','核潜艇','国立交通大学']
In [12]: print(terms)
['黄旭华', '核潜艇', '国立交通大学']
In [13]: terms
Out[13]: ['黄旭华', '核潜艇', '国立交通大学']
In [14]: # 学习一种新的数据结构，字典
In [15]: terms_dict = {}
In [16]: f_txt = open('科学家博物馆-黄旭华传记序言.txt',encoding = 'utf-8')
In [17]: data_txt = f_txt.read()
In [18]: f_txt.close()
In [19]: print(data_txt[:1000])
```

在核潜艇领域，我国已形成一套完整的研究、设计、试验、制造、测试的核潜艇产业体系，而且装备了一支具有极高战略威慑力的、成梯次配备的、已近实现装备运营的核潜艇部队。回顾我国核潜艇的发展历程，人们自然会想起以黄旭华为代表的五位两院院士及无数第一代核潜艇研制人员的皓首穷经、筚路蓝缕、无私奉献，正是他们所铸就的国之重器使我国彻底摆脱了超级大国的核讹诈，更使我们在民族复兴的道路上迈出了坚实的一步。

而今，由黄旭华院士等人所开创的核潜艇工程以令世人震撼的力量，继续承载着捍卫“中国梦”的伟大重任。

黄旭华是我国著名船舶专家、核潜艇研究设计专家、中国工程院首批院士、中国第一代核动力潜艇研制创始人之一。1924年2月24日，黄旭华出生于广东省汕尾市海丰县田陇镇，原籍广东省揭阳县。1949年，他毕业于国立交通大学造船系船舶制造专业，先后从事过民用船舶和军用舰艇的研究设计工作。1958年，黄旭华开始参与并领导我国第一代核潜艇的研究设计工作，先后出任第一代核潜艇副总设计师、第二任总设计师，历任中国船舶工业总公司及中船重工集团公司第七一九所副总工程师、副所长、所长、党委书记。黄旭华先后于1978年获全国科学大会奖、1982年获国防科工委二等奖，1986年被授予船舶工业总公司劳动模范，1989年被授予全国先进工作者，他参与完成的我国第一代核潜艇研制获1985年国家科学技术进步奖特等奖、导弹核潜艇研制获1996年国家科学技术进步奖特等奖。

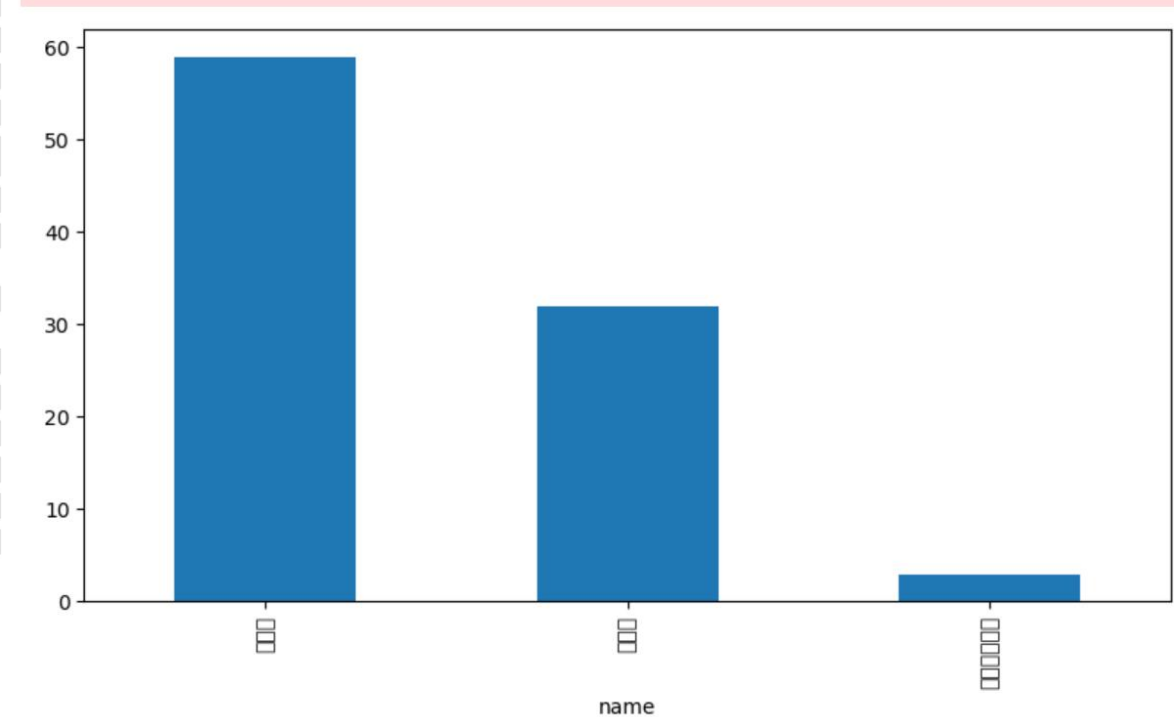
黄旭华出生于以医为主、兼理农商之家，正直、勇敢、仁厚、坚毅的父母自小给予了他良好的道德与文化的熏陶。在历经了树基小学、作叽小学、丰怀中学、广益中学、桂林中学、教育部特设大学先修班的坎坷求学历程之后，他以优异的成绩进入了当时著名的国立交通大学，系统学习造船专业理论与技术，以期实现“科学强国”的报国理想。同期在地下党的培养下，历经风雨的洗礼成长为一名坚强的共产党员。

新中国成立后，经过党校系统培训学习，黄旭华在政治思想上逐步成熟。经过苏联军事舰船的转让仿制的锤炼，黄旭华在专业技术上也崭露头角。1958年，黄旭华因为政治素质过硬、专业技术精湛，成为开启“09工程”的最初29位专业技术人员之一，从此将自己的一生献给了祖国的核潜艇事业。在核潜艇的研制过程中，黄旭华秉持“自力更生、艰苦奋斗、大力协同、无私奉献”的核潜艇精神，倡导以常规技术系统集成的科学理念，克服重重困

```
In [20]: # 用count函数统计文本中的词汇
```

你没有发现下面的代码很神奇吗？data\_txt点count一下，就可以统计词！实际上是字符串匹配的过

全文词频统计，指定类型词频统计；



## 第二讲 词频统

## 2. 完成ppt中的程

## 这部分是指定类型的频次统计

### 1 词频统计

统计文本中，指定人名的次数，步骤：

- 打开人名的文本
- 对人名进行列表化处理
- 打开文本
- 采用文本跟人名匹配的方式，进行人名统计（注：没有采用采用分词的方式）
- 用字典存储：人名-频次
- 画图展示

```
In [1]: # 文件的打开与关闭

In [2]: f_name = open('name.txt', encoding = 'GB18030') #使用mac的小伙伴，需要耐心调试下编码GB18030

In [3]: #f_name = open('name.txt')

In [3]: data_name = f_name.read()

In [4]: data_name[:70]

Out[4]: '諸葛亮|關羽|劉備|曹操|孫權|關羽|張飛|呂布|周瑜|趙雲|龐統|司馬懿|黃忠|馬超'

In [5]: print(data_name[:50])

諸葛亮|關羽|劉備|曹操|孫權|關羽|張飛|呂布|周瑜|趙雲|龐統|司馬懿|黃忠|馬超

In [6]: f_name.close()

In [7]: # 将文本转化为列表

In [8]: names = data_name.split('|') # split一下names就是列表

In [10]: print(names)

['諸葛亮', '關羽', '劉備', '曹操', '孫權', '關羽', '張飛', '呂布', '周瑜', '趙雲', '龐統', '司馬懿', '黃忠', '馬超']

In [11]: names

Out[11]: ['諸葛亮',
          '關羽',
          '劉備',
          '曹操',
          '孫權',
          '關羽',
          '張飛',
          '呂布',
          '周瑜',
          '趙雲',
          '龐統',
          '司馬懿',
          '黃忠',
          '馬超']

In [12]: # 学习一种新的数据结构，字典

In [13]: name_dict = {}

In [14]: f_txt = open('sanguo.txt', encoding = 'GB18030')

In [15]: data_txt = f_txt.read()

In [16]: f_txt.close()

In [17]: print(data_txt[:100])

《三国演义》（全）

（明）羅貫中著

第一回
宴桃園豪傑三結義斬黃巾英雄首立功
話說天下大勢，分久必合，合久必分：周末七國分爭，並
入于秦；及秦滅之後，楚、漢分爭，又並入於漢；漢朝自高祖
斬白

In [18]: # 用count函数统计文本中的词汇
```

[illegible]



## 第二讲 词频统计

3. 链接功勋科学家：把ppt中的文本换成功勋科学家黄旭华院士的传记序言文本（文件夹中，科学家博物馆-黄旭华传记序言.txt）， 1）统计全文词频； 2）统计指定词频，如“黄旭华”；

黄旭华： 47 次

核潜艇： 23 次

采集： 22 次

小组： 14 次

资料： 13 次

学术： 12 次

成长： 10 次

研制： 9 次

工作： 8 次

工程： 7 次

“黄旭华” 共出现了 47 次。



## 第二讲 词频统计

### 4. 阅读论文“2018-Wang 等 - Long live the scientists Tracking the scientific”，并做阅读总结（1页PPT即可）。

这篇论文以谷歌图书和学术文献为数据基础，聚焦物理学家的科学声誉追踪，其研究视角与方法给人诸多启发。论文最令人印象深刻的是将“科学声誉”这一抽象概念转化为可量化的研究对象，通过分析科学家姓名在全球数字化书籍和学术文献中的出现频次，打破了传统依赖奖项、引用量等单一指标评估学术影响的局限，为科学声誉研究提供了新颖的量化路径。研究中关于“群体内偏好”的发现极具洞察力——牛顿在英式英语语境中始终保有更高声誉，而爱因斯坦在美式英语、德语等语境中更受关注，这一现象深刻揭示了科学声誉并非绝对客观，而是受语言、地域等文化因素影响，让我们看到科学传播中文化语境的重要作用。同时，论文通过共现分析明确了牛顿的万有引力定律、运动定律与爱因斯坦的相对论、量子理论是其声誉的核心支撑，却也发现部分提及源于科学家的哲学观点、生平轶事等，这说明科学声誉的构成是多元的，不仅限于学术成就本身。

更值得深思的是，论文证实了伟大科学家的声誉具有跨世纪的持久性，牛顿、爱因斯坦等先贤的名字在当代仍被广泛提及，其科学精神通过书籍传播得以延续。这一发现让我们意识到，科学研究的价值不仅在于推动学术进步，更在于形成跨越时空的文化影响力。此外，论文提出谷歌图书可作为替代计量工具，拓展了学术影响评估的边界，为衡量科学家在学术界之外的社会价值提供了新的思路，也提醒我们在大数据时代，可通过更丰富的数据源挖掘学术研究的深层意义。

整体而言，这篇论文以严谨的数据分析、创新的研究视角，让我们对科学声誉的形成、演变与影响因素有了更全面的认知，也展现了数字人文方法在科学史研究中的强大潜力，为相关领域的后续研究提供了宝贵的借鉴。