Python编程新思维及实战

嵩天



实例3:金庸武侠写作风格分析

嵩天





Python编程新思维及实战

"金庸武侠写作风 格分析"需求分析

避免断更,请加微信501863613

需求分析

飞雪连天射白鹿 笑书神侠倚碧鸳





需求分析

金庸武侠写作风格是什么?

- 从词语频度角度简要分析作者的写作风格:字频+词频
- · "文笔优美" NO "辞藻华丽" NO, "用词风格" YES

需求分析

金庸武侠写作风格是什么?

• 字频: 金庸最喜欢用哪些字?

• 词频: 金庸最喜欢用哪些词?

《天龙八部》 最常用50字频 总字数: 1,253,020

他: 9763 不: 20032 子: 6385 心: 4882 无: 3597 **-:** 18686 大: 9054 有: 6232 见: 4562 好: 3580 那: 6224 的: 18443 来: 8800 时: 3575 么: 4504 之: 8008 到: 5577 是: 16485 只: 4489 要: 3519 说: 7127 道: 15430 段: 5494 个: 4483 誉: 3490 也: 5486 中: 7059 了: 15289 声: 4343 又: 3482 人: 12554 在: 6638 手: 5327 自: 4222 头: 3372 我: 10971 得: 6595 去: 5058 身: 4193 着: 3370 下: 6443 她: 3297 你: 10039 便: 4957 然: 3980 出: 4921 老: 3244 这: 9906 上: 6416 过: 3740

《笑傲江湖》 最常用50字频 总字数: 1,265,974

你: 7759 不: 17464 上: 5547 也: 4230 派: 3544 有: 5515 么: 3499 令: 7204 便: 3937 **-:** 15550 道: 14253 大: 7095 中: 5366 出: 3879 岳: 3384 的: 13825 来: 7022 子: 5351 山: 3864 然: 3336 剑: 5312 之: 6963 是: 13415 只: 3847 好: 3215 了: 13215 狐: 6754 师: 5115 声: 3776 自: 3180 人: 10572 冲: 6636 下: 4970 去: 3671 又: 3077 说: 5847 我: 8703 那: 4612 们: 3005 心: 3621 到: 4254 见: 3561 这: 8472 在: 5819 身: 2984 手: 4233 他: 8336 得: 5784 个: 3553 过: 2939

《天龙八部》 v.s. 《笑傲江湖》

不: 20032 他: 9763

一: 18686 大: 9054

的: 18443 来: 8800

是: 16485 之: 8008

了: 15289 中: 7059

人: 12554 在: 6638

我: 10971 得: 6595

你: 10039 下: 6443

这: 9906 上: 6416

不: 17464 你: 7759

一: 15550 令: 7204

道: 14253 大: 7095

的: 13825 来: 7022

是: 13415 之: 6963

了: 13215 狐: 6754

人: 10572 冲: 6636

我: 8703 说: 5847

这: 8472 在: 5819

他: 8336 得: 5784

金庸全14部作品 最常用10个字符

《雪山飞狐》 一了的是不人道我他这 《侠客行》 不一的是了道人他石我 一的不道是了人中这子 《鸳鸯刀》 一了不的是他道这我来 《连城诀》 《倚天屠龙记》 一不的是道了人他这我 的一了是人不道他这在 《白马啸西风》 《碧血剑》 一不了道的是人他来袁 《天龙八部》 不一的是道了人我你这 一不的道是了他人这来 《射雕英雄传》 《鹿鼎记》 不道了一的是小人韦宝 《神雕侠侣》 一不是的道了过他人这 《书剑恩仇录》 一不道了的是人来他上 《笑傲江湖》 不一道的是了人我这他 《飞狐外传》 一的不是了人道他这胡

金庸全14部作品 最常用20个字符

《天龙八部》 不一的是道了人我你这他大来之说中在得下上 的一了是人不道他这在来我你苏中上李秀到大 《白马啸西风》 一不了道的是人他来袁这承我志大你在手上青 《碧血剑》 《飞狐外传》 一的不是了人道他这胡大来手在中上斐你得下 《射雕英雄传》 一不的道是了他人这来我在你上大靖郭下中黄 一了的是不人道我他这大来上子那在说中下手 《雪山飞狐》 一不道了的是人来他上这在大我你家去见手下 《书剑恩仇录》 不一道的是了人我这他你令大来之狐冲说在得 《笑傲江湖》 不一的是了道人他石我这你来天中那在得说上 《侠客行》 《连城诀》 一了不的是他道这我来人你中在狄得云上大到 一不是的道了过他人这我你杨来大在之中得上 《神雕侠侣》 一的不道是了人中这子那刀来大上手得他你我 《鸳鸯刀》 一不的是道了人他这我张来无你中之大在说上 《倚天屠龙记》 不道了一的是小人韦宝这大我你他来说子上得 《鹿鼎记》

金庸全14部作品 最常用10个字符的共性字符

不是道一了的

金庸全14部作品 最常用20个字符的共性字符

不道一了是人他这的来



金庸全14部作品 最常用20个字符的共性字符

一人来,道:这不是他的了

避免断風, 遺加微信501863613

Python编程新思维及实战

"金庸武侠写作风 格分析"实例编写

代码纵览:字频

```
#飞雪连天射白鹿 笑书神侠倚碧鸳
fnames = {"飞狐外传", "雪山飞狐", "连城诀", "天龙八部", "射雕英雄传", "白马啸西风", "鹿鼎记", \
         "笑傲江湖", "书剑恩仇录", "神雕侠侣", "侠客行", "倚天屠龙记", "碧血剑", "鸳鸯刀"}
def PrintJYChars(fname)
   txt = open(fname, "r").read()
   d = {} ; cnt = 0 ; rst = ""
   for w in txt:
      cnt += 1
      d[w] = d. get(w, 0) + 1
   try:
          de / d[w]
       except:
          pass
   ls = list(d.items())
   Is. sort (key = lambda x: x[1], reverse = True)
   for i in range (20):
       word, count = Is[i]
       rst += word
   print(rst)
   return rst
txt = PrintJYChars("天龙八部" + ".txt")
A = set(txt. split("\n") [-1])
for fname in fnames:
   txt = PrintJYChars(fname + ".txt")
   A &= set(txt.split("\n")[-1])
print(A)
```

#飞雪连天射白鹿 笑书神侠倚碧鸳 fnames = {"飞狐外传", "雪山飞狐", "连城诀", "天龙八部", "射雕英雄传", "白马啸西风", "鹿鼎记", \ "笑傲江湖", "书剑恩仇录", "神雕侠侣", "侠客行", "倚天屠龙记", "碧血剑", "鸳鸯刀"} def PrintJYChars(fname) txt = open(fname, "r").read() d = {} ; cnt = 0 ; rst = "" for w in txt: cnt += 1 d[w] = d. get(w, 0) + 1for w in ", . "": ? \n []:": try: *de* / d[w] except: pass ls = list(d.items()) Is. sort(key=lambda x:x[1], reverse=True) *for* i *in* range (20): word, count = Is[i] rst += word print(rst) return rst txt = PrintJYChars("天龙八部" + ".txt") $A = set(txt. split("\n")[-1])$ for fname in fnames: txt = PrintJYChars(fname + ".txt") A &= set(txt.split("\n")[-1]) print(A)

多文件遍历





```
#飞雪连天射白鹿 笑书神侠倚碧鸳
fnames = {"飞狐外传", "雪山飞狐", "连城诀", "天龙八部", "射雕英雄传", "白马啸西风", "鹿鼎记", \
         "笑傲江湖", "书剑恩仇录", "神雕侠侣", "侠客行", "倚天屠龙记", "碧血剑", "鸳鸯刀"}
def PrintJYChars(fname)
   txt = open(fname, "r").read()
   d = {} ; cnt = 0 ; rst = ""
   for w in txt:
       cnt += 1
       d[w] = d. get(w, 0) + 1
   try:
          de / d[w]
       except:
          pass
   Is = list(d.items())
   Is. sort (key= lambda x: x[1], reverse= True)
   for i in range (20):
       word, count = Is[i]
       rst += word
   print(rst)
   return rst
txt = PrintJYChars("天龙八部" + ".txt")
A = set(txt. split("\n")[-1])
for fname in fnames:
   txt = PrintJYChars(fname + ".txt")
   A &= set(txt.split("\n")[-1])
print(A)
```

输出前20最 多字符函数



```
#飞雪连天射白鹿 笑书神侠倚碧鸳
fnames = {"飞狐外传", "雪山飞狐", "连城诀", "天龙八部", "射雕英雄传", "白马啸西风", "鹿鼎记", \
         "笑傲江湖", "书剑恩仇录", "神雕侠侣", "侠客行", "倚天屠龙记", "碧血剑", "鸳鸯刀"}
def PrintJYChars(fname)
   txt = open(fname, "r").read()
   d = {} ; cnt = 0 ; rst = ""
   for w in txt:
       cnt += 1
      d[w] = d. get(w, 0) + 1
   try:
          de / d[w]
       except:
          pass
   Is = list(d.items())
   Is. sort (key = lambda x: x[1], reverse = True)
   for i in range (20):
       word, count = Is[i]
       rst += word
   print(rst)
   return rst
txt = PrintJYChars("天龙八部" + ".txt")
A = set(txt. split("\n")[-1])
for fname in fnames:
   txt = PrintJYChars(fname + ".txt")
   A &= set(txt.split("\n")[-1])
print(A)
```

读入文件遍 历每个字符 采用字典统 计字频





```
#飞雪连天射白鹿 笑书神侠倚碧鸳
fnames = {"飞狐外传", "雪山飞狐", "连城诀", "天龙八部", "射雕英雄传", "白马啸西风", "鹿鼎记", \
         "笑傲江湖", "书剑恩仇录", "神雕侠侣", "侠客行", "倚天屠龙记", "碧血剑", "鸳鸯刀"}
def PrintJYChars(fname)
   txt = open(fname, "r").read()
   d = {} ; cnt = 0 ; rst = ""
   for w in txt:
       cnt += 1
       d[w] = d. get(w, 0) + 1
   try:
          de / d[w]
       except:
          pass
   Is = list(d.items())
   Is. sort (key = lambda x: x[1], reverse = True)
   for i in range (20)
       word, count = Is[i]
       rst += word
   print(rst)
   return rst
txt = PrintJYChars("天龙八部" + ".txt")
A = set(txt. split("\n") [-1])
for fname in fnames:
   txt = PrintJYChars(fname + ".txt")
   A &= set(txt.split("\n")[-1])
print(A)
```

```
#飞雪连天射白鹿 笑书神侠倚碧鸳
fnames = {"飞狐外传", "雪山飞狐", "连城诀", "天龙八部", "射雕英雄传", "白马啸西风", "鹿鼎记", \
         "笑傲江湖", "书剑恩仇录", "神雕侠侣", "侠客行", "倚天屠龙记", "碧血剑", "鸳鸯刀"}
def PrintJYChars(fname)
   txt = open(fname, "r").read()
   d = {} ; cnt = 0 ; rst = ""
   for w in txt:
       cnt += 1
       d[w] = d. get(w, 0) + 1
   try:
          de / d[w]
       except:
          pass
   Is = list(d.items())
   Is. sort (key = lambda x: x[1], reverse = True)
   for i in range (20):
       word, count = Is[i]
       rst += word
   print(rst)
   return rst
txt = PrintJYChars("天龙八部" + ".txt")
A = set(txt. split("\n")[-1])
for fname in fnames:
   txt = PrintJYChars(fname + ".txt")
   A &= set(txt.split("\n")[-1])
print(A)
```

将字典变成 二元元组的 列表,根据 第二项排序 列表



```
#飞雪连天射白鹿 笑书神侠倚碧鸳
fnames = {"飞狐外传", "雪山飞狐", "连城诀", "天龙八部", "射雕英雄传", "白马啸西风", "鹿鼎记", \
        "笑傲江湖", "书剑恩仇录", "神雕侠侣", "侠客行", "倚天屠龙记", "碧血剑", "鸳鸯刀"}
def PrintJYChars(fname)
   txt = open(fname, "r").read()
   d = {} ; cnt = 0 ; rst = ""
   for w in txt:
      cnt += 1
      d[w] = d. get(w, 0) + 1
   try:
         de / d[w]
      except:
         pass
                                                                                              将前20位最
   Is = list(d.items())
   Is. sort (key = lambda x: x[1], reverse = True)
                                                                                              多的字符组
   for i in range (20):
      word, count = Is[i]
                                                                                              成字符串输
      rst += word
   print(rst)
                                                                                              出并返回
   return rst
txt = PrintJYChars("天龙八部" + ".txt")
                                                避免断更,请加微信50186
A = set(txt. split("\n") [-1])
for fname in fnames:
   txt = PrintJYChars(fname + ".txt")
   A &= set(txt.split("\n")[-1])
```



print(A)



```
#飞雪连天射白鹿 笑书神侠倚碧鸳
fnames = {"飞狐外传", "雪山飞狐", "连城诀", "天龙八部", "射雕英雄传", "白马啸西风", "鹿鼎记", \
         "笑傲江湖", "书剑恩仇录", "神雕侠侣", "侠客行", "倚天屠龙记", "碧血剑", "鸳鸯刀"}
def PrintJYChars(fname)
   txt = open(fname, "r").read()
   d = {} ; cnt = 0 ; rst = ""
   for w in txt:
       cnt += 1
       d[w] = d. get(w, 0) + 1
    try:
          de / d[w]
       except:
          pass
   ls = list(d.items())
   Is. sort (key = lambda x: x[1], reverse = True)
   for i in range (20):
       word, count = Is[i]
       rst += word
   print(rst)
   return rst
txt = PrintJYChars("天龙八部" + ".txt")
A = set(txt. split("\n")[-1])
for fname in fnames
   txt = PrintJYChars(fname + ".txt")
   A &= set(txt.split("\n")[-1])
print(A)
```

随便找一部作品为起点前20个字符组成集合



```
#飞雪连天射白鹿 笑书神侠倚碧鸳
fnames = {"飞狐外传", "雪山飞狐", "连城诀", "天龙八部", "射雕英雄传", "白马啸西风", "鹿鼎记", \
         "笑傲江湖", "书剑恩仇录", "神雕侠侣", "侠客行", "倚天屠龙记", "碧血剑", "鸳鸯刀"}
def PrintJYChars(fname)
    txt = open(fname, "r").read()
    d = {} ; cnt = 0 ; rst = ""
    for w in txt:
       cnt += 1
       d[w] = d. get(w, 0) + 1
    for w in ", . "": ? \n []:":
       try:
           de / d[w]
       except:
           pass
    ls = list(d.items())
    Is. sort (key = lambda x: x[1], reverse = True)
    for i in range (20):
       word, count = Is[i]
       rst += word
    print(rst)
    return rst
txt = PrintJYChars("天龙八部" + ".txt")
A = set(txt. split("\n") [-1])
for fname in fnames:
   txt = PrintJYChars(fname + ".txt")
   A &= set(txt.split("\n")[-1])
print(A)
```

遍历所有作品,将最多字符计算交 集,并输出





程序结果

《天龙八部》 《白马啸西风》 《碧血剑》 《飞狐外传》 《射雕英雄传》 《雪山飞狐》 《书剑恩仇录》 《笑傲江湖》 《侠客行》 《连城诀》 《神雕侠侣》 《鸳鸯刀》 《倚天屠龙记》 《鹿鼎记》

不一的是道了人我你这他大来之说中在得下上 的一了是人不道他这在来我你苏中上李秀到大 一不了道的是人他来袁这承我志大你在手上青 一的不是了人道他这胡大来手在中上斐你得下 一不的道是了他人这来我在你上大靖郭下中黄 一了的是不人道我他这大来上子那在说中下手 一不道了的是人来他上这在大我你家去见手下 不一道的是了人我这他你令大来之狐冲说在得 不一的是了道人他石我这你来天中那在得说上 一了不的是他道这我来人你中在狄得云上大到 一不是的道了过他人这我你杨来大在之中得上 一的不道是了人中这子那刀来大上手得他你我 一不的是道了人他这我张来无你中之大在说上 不道了一的是小人韦宝这大我你他来说子上得

金庸全14部作品 最常用20个字符的共性字符

不道一了是人他这的来



金庸全14部作品 最常用20个字符的共性字符

一人来,道:这不是他的了

代码纵览: 词频

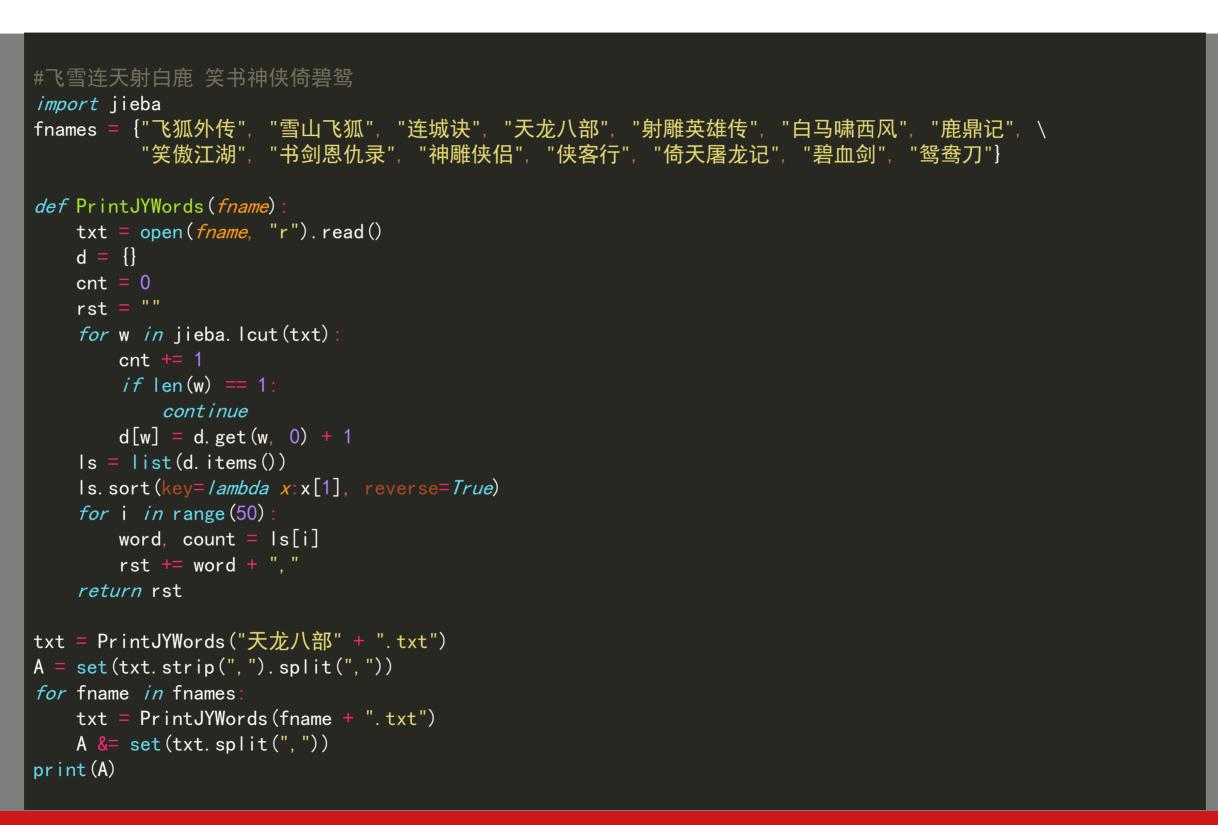
```
#飞雪连天射白鹿 笑书神侠倚碧鸳
import jieba
fnames = {"飞狐外传", "雪山飞狐", "连城诀", "天龙八部", "射雕英雄传", "白马啸西风", "鹿鼎记", \
         "笑傲江湖", "书剑恩仇录", "神雕侠侣", "侠客行", "倚天屠龙记", "碧血剑", "鸳鸯刀"}
def PrintJYWords(fname)
    txt = open(fname, "r").read()
   d = \{\}
    cnt = 0
    rst = ""
    for w in jieba. lcut(txt)
       cnt += 1
       if len(w) == 1:
           cont i nue
       d[w] = d. get(w, 0) + 1
    Is = list(d.items())
    Is. sort (key= lambda x:x[1], reverse= True)
    for i in range (50)
       word, count = Is[i]
       rst += word + ","
    return rst
txt = PrintJYWords("天龙八部" + ". txt")
A = set(txt.strip(",").split(","))
for fname in fnames:
    txt = PrintJYWords(fname + ".txt")
   A &= set(txt.split(","))
print(A)
```

2个字符以上 的词频统计



```
#飞雪连天射白鹿 笑书神侠倚碧鸳
import jieba
fnames = {"飞狐外传", "雪山飞狐", "连城诀", "天龙八部", "射雕英雄传", "白马啸西风", "鹿鼎记", \
         "笑傲江湖", "书剑恩仇录", "神雕侠侣", "侠客行", "倚天屠龙记", "碧血剑", "鸳鸯刀"}
def PrintJYWords(fname)
    txt = open(fname, "r").read()
    d = \{\}
    cnt = 0
    rst = ""
    for w in jieba. lcut(txt)
       cnt += 1
       if len(w) == 1:
           cont i nue
       d[w] = d. get(w, 0) + 1
    Is = list(d.items())
    Is. sort(key= lambda x:x[1], reverse=True)
    for i in range (50)
       word, count = Is[i]
       rst += word + ","
    return rst
| txt = PrintJYWords("天龙八部" + ".txt")
A = set(txt.strip(",").split(","))
for fname in fnames:
    txt = PrintJYWords(fname + ".txt")
   A &= set(txt.split(","))
print(A)
```

改名为 PrintJYWo rds()



引入jieba库 进行分词



```
#飞雪连天射白鹿 笑书神侠倚碧鸳
import jieba
fnames = {"飞狐外传", "雪山飞狐", "连城诀", "天龙八部", "射雕英雄传", "白马啸西风", "鹿鼎记", \
         "笑傲江湖", "书剑恩仇录", "神雕侠侣", "侠客行", "倚天屠龙记", "碧血剑", "鸳鸯刀"}
def PrintJYWords(fname)
   txt = open(fname, "r").read()
   d = \{\}
   cnt = 0
   rst = ""
   for w in jieba. lcut(txt)
       cnt += 1
       if len(w) == 1:
          cont i nue
       d[w] = d. get(w, 0) + 1
   Is = list(d.items())
   Is. sort (key= /ambda x:x[1], reverse= True)
   for i in range (50)
       word, count = Is[i]
       rst += word + ","
    return rst
txt = PrintJYWords("天龙八部" + ". txt")
                                            避免断更,请加微信50186
A = set(txt.strip(",").split(","))
for fname in fnames:
   txt = PrintJYWords(fname + ".txt")
   A &= set(txt.split(", "))
```

分词后一个 字符的不予 考虑



print(A)

```
#飞雪连天射白鹿 笑书神侠倚碧鸳
import jieba
fnames = {"飞狐外传", "雪山飞狐", "连城诀", "天龙八部", "射雕英雄传", "白马啸西风", "鹿鼎记", \
         "笑傲江湖", "书剑恩仇录", "神雕侠侣", "侠客行", "倚天屠龙记", "碧血剑", "鸳鸯刀"}
def PrintJYWords(fname)
    txt = open(fname, "r").read()
    d = \{\}
    cnt = 0
    rst = ""
    for w in jieba. lcut(txt)
       cnt += 1
       if len(w) == 1:
           cont i nue
       d[w] = d. get(w, 0) + 1
    Is = list(d.items())
    Is. sort(key= lambda x:x[1], reverse=True)
    for i in range (50)
       word, count = Is[i]
       rst += word + ","
    return rst
txt = PrintJYWords("天龙八部" + ".txt")
A = set(txt.strip(",").split(","))
for fname in fnames:
    txt = PrintJYWords(fname + ".txt")
   A &= set(txt. strip(", "). split(", "))
print(A)
```

每个作品输出排名前50的单词



程序结果

说道 心中 自己 出来 一声 一个 知道



似乎找到了金庸的写作风格…



Thank you