

# SongPoem

## R语言分析宋词300首

其实做法很简单，大概就是分这么几步：

- 把文本拆分成一个一个的单词
- 把单词按照出现的频率、次数进行排序
- 用可视化把结果展示出来

## 导入文本库

```
fileName <- "F:\\\\Code\\\\R\\\\Song\\\\songci300.txt"  
SC <- readChar(fileName, file.info(fileName)$size)
```

```
## Warning in readChar(fileName, file.info(fileName)$size): 在non-UTF-8 MBCS语  
## 言环境里只能读取字节
```

```
substr(SC, 1000, 1100)
```

```
## [1] "伤春，不如怜取眼前人。\\r\\n晏殊\\r\\n红笺小字，说尽平生意。鸿雁在云鱼在水，惆怅此情难  
寄。\\r\\n斜阳独倚西楼，遥山恰对帘钩。人面不知何处，绿波依旧东流。\\r\\n晏殊\\r\\n金风细细，叶叶  
梧桐坠。绿酒初尝人易醉，一枕窗"
```

## 用分词包『结巴R』里面的 worker() 公式完成分词

```
library(jiebaR)
```

```
## Warning: package 'jiebaR' was built under R version 3.3.3
```

```
## Loading required package: jiebaRD
```

```
## Warning: package 'jiebaRD' was built under R version 3.3.3
```

```
cc = worker()  
analysis <- as.data.frame(table(cc[SC]))  
names(analysis) <- c("word", "freq")  
analysis <- analysis[order(-analysis$freq),]  
analysis$word <- as.character(analysis$word)  
head(analysis)
```







```
wordcloud2(analysis[analysis$freq>1& analysis$freq < 300 & nchar(analysis$word) == 2,])
```





```
wordcloud2(analysis[analysis$freq>1 & analysis$freq < 300 & nchar(analysis$word) == 3,])
```



# 诗词创作 分析准备

```
cipai <- "画堂晨起，来报雪花坠。高卷帘栊 看 佳瑞，皓色远 迷 庭砌。盛气光引 炉烟，素草寒生  
玉佩。应是天仙狂醉，乱把白云揉碎。"  
tagger <- worker("tag")  
cipai_2 <- tagger <= cipai  
cipai_2
```

```
##          n          x          x          n          v          a          n          g          v          x  
## "画堂" "晨起" "来报" "雪花" "坠" "高" "卷帘" "栊" "看" "佳瑞"  
##          x          a          v          x          n          x          x          x          x          nr  
## "皓色" "远" "迷" "庭砌" "盛气" "光引" "炉烟" "素草" "寒生" "玉佩"  
##          x          n          x          d          p          nr          v  
## "应是" "天仙" "狂醉" "乱" "把" "白云" "揉碎"
```

从之前提炼的宋词词频库中，选取了至少出现过 两次 的一字或两字词语，作为诗词创作的素材库；然后提取词性文件，将素材根据词性分类：

```
example <- subset(analysis, freq >1 & nchar(word) <3 & freq < 300)  
cixing <- attributes(cipai_2)$names  
example_2 <- tagger <= example$word
```

# 开始创作

从范本词牌的 第一个词 开始，随机在素材库中选取 词性相同，字数相等 的单词，填入提前设置好的空白字符串中。

```
write_songci <- function(m){
  set.seed(m)
  empty <- ""
  for (i in 1:length(cipai_2)){
    temp_file <- example_2[attributes(example_2)$name == cixing[i]]
    temp_file <- temp_file[nchar(temp_file) == nchar(cipai_2[i])]
    empty <- paste0(empty, sample(temp_file,1))
  }

  result <- paste0(substr(empty, 1,4), ",", substr(empty,5,9),".",
    substr(empty, 10,16), ",", substr(empty, 17,22),".",
    substr(empty, 23,28), ",", substr(empty, 29,34),".",
    substr(empty, 35,40), ",", substr(empty, 41,46),".")
}
```

# 看看创作成果

```
lapply(1:5, write_songci)
```

```
## [[1]]
## [1] "乱鸦初静，帘外心眼断。慢银钩，榭吐灯前，轻寒暖抱春空。青草何似诗酒，竹槛春恨翩翩。犹在垂柳
平芜，欲凭佩环归来。"
##
## [[2]]
## [1] "楼头若到，帘外风味闻。嫩春色，旆舞看尽，困倚绿干泪满。华表心似争忍，声里芳思苏轼。轻鸥归路
湘天，却为言语衰草。"
##
## [[3]]
## [1] "风味午梦，春空踪迹认。碎游丝，枕觅难留，荒台阴赋赖有。万事香凝残醉，沙际遥山王孙。时见芳草
更苦，都凭清香更阑。"
##
## [[4]]
## [1] "草草双燕，赵令青苔聚。红琉璃，醺卧绿杨，天角难听云来。画图非烟长向，柳花怎得蓬壶。声住滋味
呼灯，只从秋水不负。"
##
## [[5]]
## [1] "小窗情伤，倚楼情味望。浓一梦，髻引残醉，已断细晴阑角。楼台轻别春山，燕股赖有双飞。燕股留人
人去，便在晓色闲愁。"
```