

```

library(httr)
library(jsonlite)
library(tidyverse)

options(stringsAsFactors = FALSE)
options(encoding = "UTF-8")

dcardurl <- "https://www.dcard.tw/_api/forums/"
board <- 'makeup'

#把url跟看板融合成一個網址，將抓的順序設定成熱門排序，所以用true
mainurl <- paste0(dcardurl,board,'/posts?popular=true')

# 抽出json，把他存入resdata這個data.frame裡面
resdata <- fromJSON(content(GET(mainurl), "text"))

#先查看前兩個Column
head(resdata[,c(1,2)])

#假設要抓200篇文章
n <- 200

# 因為不改limit值，所以他預設會每次抓20篇回來，我們把要抓的文章/20便是我們要抓的次數
# 還要再減一，因為我們一開始就先抓了前20筆
page <- (200/20)-1

#抓到的最後一篇文章id
end <- resdata$id[length(resdata$id)]

#寫一個loop，重複做page次
for(i in 1:page){
  # 從「目前抓到的最後一篇文章id」往前抓20篇
  url <- paste0(mainurl,"&before=",end)
  # 測試時可以把url印出來檢查有沒有抓對
  print(url)
  # 把抓到存入暫存的tmpres，這只是暫存
  tmpres <- fromJSON(content(GET(url), "text"))

  # 從tmpres裡更新「最後一篇文章的id」
  end <- tmpres$id[length(tmpres$id)]

  # 然後把我們新抓到的tmpres和之前已經有的resdata合併
  resdata <- bind_rows(resdata[,c(1:12)],tmpres[,c(1:12)])
}

#省記憶體
rm(tmpres)

#查看前幾筆
head(resdata)

cc<-worker()
count <-table(cc[resdata[,2]])

newd = data.frame(count)

```

```
head(newd[order(newd$Freq,decreasing = TRUE),],20)
newdd = newd[order(newd$Freq,decreasing = TRUE),]
wordcloud2(newdd)

word <- cc[resdata[,2]]
newd = data.frame(table(word))

newd %>%
  filter(!str_detect(word, "[a-zA-Z0-9]+")) %>% #去掉english and number
  filter(nchar(as.character(word)) > 1) %>% #一個字的去掉
  filter( Freq > 1) ->temp #可留下頻率>某數字

wordcloud2(temp)
```