

wordcloud-example

텍스트파일 읽어보기

```
adr <- file("56 노무현.txt", blocking=F, encoding = "UTF-8") #글씨 깨어진다면 encoding 바꾸  
어야 한다.
```

```
adr2 <- readLines(adr); head(adr2)
```

```
## [1] "국민 여러분, 안녕하십니까? 새해 복 많이 받으십시오."
```

```
## [2] ""
```

```
## [3] "지난해에도 어려움이 많으셨지요?"
```

```
## [4] "지난 3년 동안 정말 고생 많으셨습니다. 이 기간 전체가 제 임기 중이라 국민 여러분께 송구스럽기 짝이 없  
습니다."
```

```
## [5] ""
```

```
## [6] "그러나 국민 여러분, 반가운 소식도 있습니다."
```

```
close(adr)
```

한글 사전

```
useSejongDic() #세종 사전을 사용. 한글 단어 사전이다.
```

```
## Backup was just finished!
```

```
## 370957 words dictionary was built.
```

```
dics <- c('sejong', 'woorimalsam')
```

```
category <- c('news')
```

```
user_d <- data.frame(c('안철수', '박근혜', '노무현', '정규직'), c('nqpc'))
```

```
buildDictionary(ext_dic=dics, category_dic_nms = category, user_dic = user_d, replace  
_usr_dic=F)
```

```
## 716933 words dictionary was built.
```

텍스파일의 각 줄에서 명사 추출하기

```
nouns <- sapply(adr2, extractNoun, USE.NAMES=F) # 각 줄에서 명사만 추출해 낸다. KoNLP의 함  
수다
```

```
head(nouns)
```

```
## [[1]]
## [1] "국민"      "여러분" "안녕"      "하"        "새해"      "복"
##
## [[2]]
## [1] ""
##
## [[3]]
## [1] "지난해" "어려움"
##
## [[4]]
## [1] "3"      "년"      "동안"   "고생"    "기간"    "전체"    "저"
## [8] "임기"    "중"      "국민"    "여러분" "송구"    "짜"
##
## [[5]]
## [1] ""
##
## [[6]]
## [1] "국민"      "여러분" "소식"
```

두 글자 이상으로 필터링하기

```
c <- unlist(nouns) # 필터링을 위해 unlist 작업을 해서 저장합니다.
nouns <- Filter(function(x) {nchar(x) >= 2} ,c) # 두 글자 이상 되는 것만 필터링하기
head(nouns)
```

```
## [1] "국민"      "여러분" "안녕"      "새해"      "지난해" "어려움"
```

숫자 필터링하기

```
nouns <- gsub(" |\\d+", "", nouns) # 숫자 없애기
c <- unlist(nouns) # 필터링을 위해 unlist 작업을 해서 저장합니다.
nouns <- Filter(function(x) {nchar(x) >= 2} ,c) # 두 글자 이상 되는 것만 필터링하기
```

단어 카운트

```
wordcount <- table(unlist(nouns))
head(sort(wordcount, decreasing=T),30) # 그냥 확인해 보려고
```

```
##
##      우리      국민      문제      여러분      정부      일자리      하게      해결
##      34       31       29       22       22       19       19       19
##      사회      경제      정책      해서      책임      수준      중소기업      교육
##      18       17       17       17       14       10       10       9
##      그동안      주장      고용      비정규직      서비스      이상      지난해      하기
##      9        9        8        8        8        8        8        8
##      하지      대책      들이      때문      문화      분야
##      8        7        7        7        7        7
```

워드클라우드 시각화

```
pal <- brewer.pal(12,"Set3")
pal <- pal[-c(1:2)]
wordcloud(names(wordcount), freq=wordcount, scale=c(7, 0.2), min.freq=7,
random.order=F, random.color=F, rot.per=.2,
          colors=pal, family="AppleGothic")
```

```
## Warning in wordcloud(names(wordcount), freq = wordcount, scale = c(7,
## 0.2), : 해서 could not be fit on page. It will not be plotted.
```

```
## Warning in wordcloud(names(wordcount), freq = wordcount, scale = c(7,
## 0.2), : 지원 could not be fit on page. It will not be plotted.
```

```
## Warning in wordcloud(names(wordcount), freq = wordcount, scale = c(7,
## 0.2), : 추진 could not be fit on page. It will not be plotted.
```

```
## Warning in wordcloud(names(wordcount), freq = wordcount, scale = c(7,
## 0.2), : 필요 could not be fit on page. It will not be plotted.
```

