

Textmining with tm package

Jaeyoon Han

2016-02-01

```
library(tm)
```

```
## Loading required package: NLP
```

```
##
```

```
## Attaching package: 'NLP'
```

```
## The following object is masked from 'package:ggplot2':
```

```
##
```

```
##      annotate
```

```
library(magrittr)
```

```
library(SnowballC)
```

```
jobs <- readLines("./data/steve.txt")
```

```
# For Mac
```

```
jobs <- iconv(jobs, "WINDOWS-1252", "UTF-8")
```

```
head(jobs)
```

```
## [1] "'You've got to find what you love,' Jobs says"
```

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

```
## [3] "This is the text of the Commencement address by Steve Jobs, CEO of Apple Computer and of Pixar A
```

```
## [4] ""
```

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

```
## [6] "I am honored to be with you today at your commencement from one of the finest universities in t
```

```
steve_jobs <- jobs %>%
```

```
  tolower %>%
```

```
  removeNumbers %>%
```

```
  removeWords(stopwords()) %>%
```

```
  removePunctuation %>%
```

```
  stemDocument %>%
```

```
  stripWhitespace
```

```
head(steve_jobs)
```

```
## [1] " got find love jobs say"
```

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

```
## [3] " text commencement address steve jobs ceo apple computer pixar animation studios delivered june
```

```
## [4] ""
```

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

```
## [6] " honored today commencement one finest universities world never graduated college truth told cl
```

```

write(unlist(steve_jobs), "./data/jobs.txt")

jobs <- read.table("./data/jobs.txt", fill = TRUE)
jobs <- as.list(jobs)

jobs_words <- c()
for(i in 1 : length(jobs)){
  jobs_words <- c(jobs_words, as.character(jobs[[i]]))
}

head(jobs_words, 50)

```

```

## [1] "got"      "text"      "honored"   "first"     "dropped"
## [6] "started"  "really"    "high"      "years"     "spending"
## [11] "begin"    "romantic"  "temple"    "reed"      "learn"
## [16] "none"     "multiple"  "looking"   "connect"   "second"
## [21] "lucky"    "earlier"   "eventually" "really"    "slowly"
## [26] "see"      "next"      "now"       "pretty"    "work"
## [31] "great"    "third"     "read"      "answer"    "remembering"
## [36] "important" "year"      "go"        "family"    "lived"
## [41] "form"     "closest"   "one"       "old"       "time"
## [46] "truly"    "young"     "made"      "stewart"   "adventurous"

```

```

write(unlist(jobs_words), "./data/jobs_words.txt")
jobs_words <- read.table("./data/jobs_words.txt")

```