
인문학 텍스트 마이닝

Text Analysis “wordcloud”

Text Analysis “wordcloud”

- 경로 지정

```
> setwd("/Users/Seongmin_M/Desktop/Class")  
>
```

- 데이터 불러오기

```
> Harry_potter_1 <- file("/Users/Seongmin_M/Desktop/Class/Harry_potter_1.txt", blocking=F)
```

- 데이터 읽어들이기

```
> Harry_txtLines <- readLines(Harry_potter_1)
```

- 데이터 닫기

```
> close(Harry_potter_1)
```

- 데이터 다듬기

```
> Harry_txtLines <- gsub("()", "", Harry_txtLines)  
> Harry_txtLines <- gsub("<", "", Harry_txtLines)  
> Harry_txtLines <- gsub(">", "", Harry_txtLines)  
> Harry_txtLines <- gsub("[ \\t]{2,}", "", Harry_txtLines)
```

Text Analysis “wordcloud”

- tm 패키지 설치 및 라이브러리 불러오기

```
> install.packages("tm")
```

```
% Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
             %             %             Dload  Upload  Total  Spent    Left  Speed
  0      0     0     0     0     0     0     0  --:--:--  --:--:--  --:--:--    0100  644k  100
644k     0     0  1290k     0  --:--:--  --:--:--  --:--:--  1292k
```

The downloaded binary packages are in

```
/var/folders/28/g8cf_pvx46s5phqgwr6qq7jw0000gn/T//RtmpZs3wki/downloaded_packages
```

```
> library(tm)
```

```
필요한 패키지를 로딩중입니다: NLP
```

- 텍스트 데이터를 코퍼스데이터로 변환

```
> Harry_txtLines_corpus <- Corpus(VectorSource(Harry_txtLines))
```

- 마침표 제거하기

```
> Harry_txtLines_corpus <- tm_map(Harry_txtLines_corpus, function(x)removeWords(x,stopwords()
))
```

Text Analysis “wordcloud”

- 수치형 데이터로 형 변환

```
> Harry_Tdm <- TermDocumentMatrix(Harry_txtLines_corpus, control = list(wordLengths = c(2, Inf)))
```

- wordcloud 패키지 설치 및 라이브러리 불러오기

```
> install.packages("wordcloud")
% Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
           Dload  Upload   Total   Spent    Left   Speed
  0      0    0     0    0     0     0     0  ---:---:--  ---:---:--  ---:---:--    0  1  138k    1
2168    0     0  2067    0  0:01:08  0:00:01  0:01:07  2068100  138k  100  138k    0    0
126k     0  0:00:01  0:00:01  ---:---:--  126k
```

The downloaded binary packages are in

```
/var/folders/28/g8cf_pvx46s5phqgwr6qq7jw0000gn/T//RtmpZs3wki/downloaded_packages
```

```
>
```

```
> library(wordcloud)
```

```
필요한 패키지를 로딩중입니다: RColorBrewer
```

Text Analysis “wordcloud”

- 매트릭스 형태로 형 변환

```
> Harry_Tdm_M <- as.matrix(Harry_Tdm)
```

- 단어들의 출현빈도 카운팅

```
> Harry_wordFreq <- sort(rowSums(Harry_Tdm_M), decreasing = TRUE)
```

- 워드클라우드 색상 지정

```
> pal <- brewer.pal(8, "Dark2")
```

