

Lab04_KeyWordEx

- 01 텍스트 마이닝 패키지 불러오기
- 02 testText에 저장하기
- 03 텍스트에서 말뭉치(Corpus)로 변환
- 04. corpus 로 변환하여 저장 후, 전처리
- 05. 전처리가 후, 문서 내용 확인
- 06. 불용어 처리 수행
- 07. 단어문서행렬(Term Document Matrix, TDM)

01 텍스트 마이닝 패키지 불러오기

```
library(tm)
```

```
## Warning: package 'tm' was built under R version 3.4.4
```

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

```
## Warning: package 'NLP' was built under R version 3.4.1
```

02 testText에 저장하기

```
testText <- readLines("D:/dataset/textMining/anb-jarena-lee.txt")
```

```
## Warning in readLines("D:/dataset/textMining/anb-jarena-lee.txt"): 'D:/  
## dataset/textMining/anb-jarena-lee.txt'에서 불완전한 마지막 행이 발견되었습  
## 니다
```

```
print(testText)
```

```
## [1] "In 1804, after several months of profound spiritual anxiety, Jarena Lee"
## [2] "moved from New Jersey to Philadelphia. There she labored as a domestic"
## [3] "and worshiped among white congregations of Roman Catholics and mixed"
## [4] "congregations of Methodists. On hearing an inspired sermon by the"
## [5] "Reverend Richard Allen, founder of the Bethel African Methodist"
## [6] "Episcopal Church, Lee joined the Methodists. She was baptized in 1807."
## [7] "Prior to her baptism, she experienced the various physical and emotional"
## [8] "stages of conversion terrifying visions of demons and eternal"
## [9] "perdition; extreme feelings of ecstasy and depression; protracted"
## [10] "periods of meditation, fasting, and prayer; ennui and fever; energy and"
## [11] "vigor. In 1811 she married Joseph Lee, who pastored an African-American"
## [12] "church in Snow Hill, New Jersey. They had six children, four of whom"
## [13] "died in infancy."
```

```
str(testText)
```

```
## chr [1:13] "In 1804, after several months of profound spiritual anxiety, Jarena Lee" ...
```

03 텍스트에서 말뭉치(Corpus)로 변환

VectorSource : 벡터 형태의 소스를 만든다.

[Usage]

VectorSource(x)

x : A vector giving the texts

```
textText1 <- VectorSource(testText)
textText1
```

```
## $encoding
## [1] ""
##
## $length
## [1] 13
##
## $position
## [1] 0
##
## $reader
## function (elem, language, id)
## {
##     if (!is.null(elem$uri))
##         id <- basename(elem$uri)
##     PlainTextDocument(elem$content, id = id, language = language)
## }
## <environment: namespace:tm>
##
## $content
## [1] "In 1804, after several months of profound spiritual anxiety, Jarena Lee"
## [2] "moved from New Jersey to Philadelphia. There she labored as a domestic"
## [3] "and worshiped among white congregations of Roman Catholics and mixed"
## [4] "congregations of Methodists. On hearing an inspired sermon by the"
## [5] "Reverend Richard Allen, founder of the Bethel African Methodist"
## [6] "Episcopal Church, Lee joined the Methodists. She was baptized in 1807."
## [7] "Prior to her baptism, she experienced the various physical and emotional"
## [8] "stages of conversion terrifying visions of demons and eternal"
## [9] "perdition; extreme feelings of ecstasy and depression; protracted"
## [10] "periods of meditation, fasting, and prayer; ennui and fever; energy and"
## [11] "vigor. In 1811 she married Joseph Lee, who pastored an African-American"
## [12] "church in Snow Hill, New Jersey. They had six children, four of whom"
## [13] "died in infancy."
##
## attr(,"class")
## [1] "VectorSource" "SimpleSource" "Source"
```

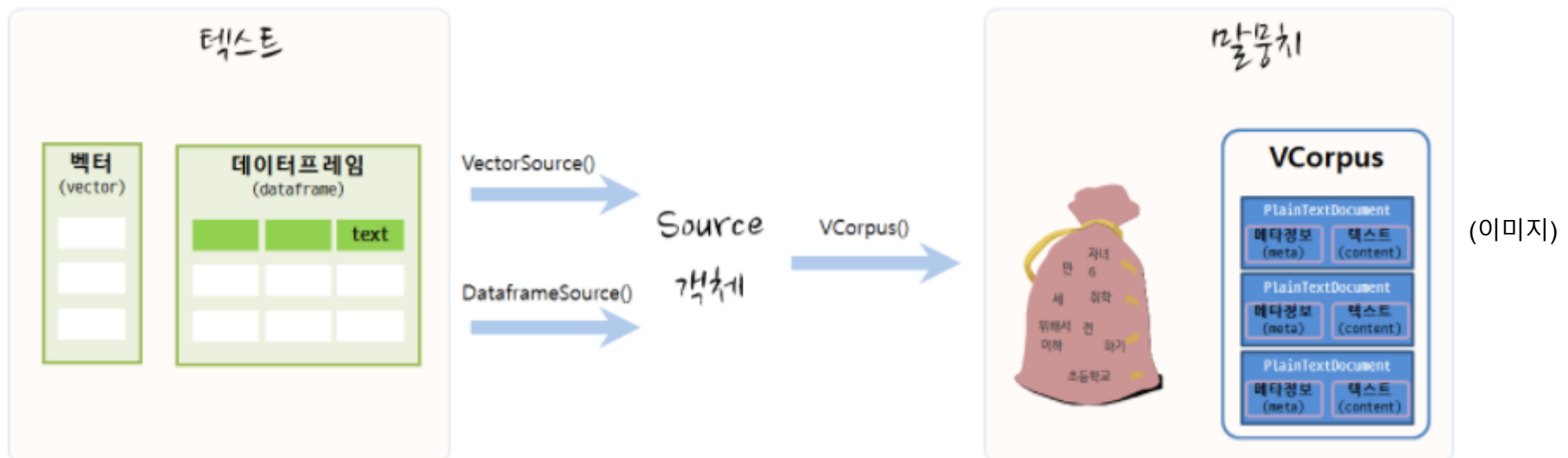
04. corpus 로 변환하여 저장 후, 전처리

Corpus : 텍스트 문서 집합을 구성한다.(Constructs a text document collection-corpus)

Usage : `Corpus(object, readerControl = list(reader = object@DefaultReader, language = "en_US", load = TRUE), dbControl = list(useDb = FALSE, dbName = "", dbType = "DB1"), ...)`

object : A Source Object (하나의 원본 객체)

Value : 리스트 형태의 객체



<https://statklee.github.io/ml/ml-wordcloud.html> (<https://statklee.github.io/ml/ml-wordcloud.html>) 참조

```
corpus <- Corpus(textText1)
corpus
```

```
## <<SimpleCorpus>>
## Metadata: corpus specific: 1, document level (indexed): 0
## Content: documents: 13
```

```
str(corpus)
```

```

## List of 13
## $ 1 :List of 2
## ..$ content: chr "In 1804, after several months of profound spiritual anxiety, Jarena Lee"
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2018-10-29 17:52:08"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "1"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 2 :List of 2
## ..$ content: chr "moved from New Jersey to Philadelphia. There she labored as a domestic"
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2018-10-29 17:52:08"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "2"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 3 :List of 2
## ..$ content: chr "and worshiped among white congregations of Roman Catholics and mixed"
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2018-10-29 17:52:08"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "3"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 4 :List of 2
## ..$ content: chr "congregations of Methodists. On hearing an inspired sermon by the"

```

```

## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2018-10-29 17:52:08"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "4"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 5 :List of 2
## ..$ content: chr "Reverend Richard Allen, founder of the Bethel African Methodist"
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2018-10-29 17:52:08"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "5"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 6 :List of 2
## ..$ content: chr "Episcopal Church, Lee joined the Methodists. She was baptized in 1807."
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2018-10-29 17:52:08"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "6"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 7 :List of 2
## ..$ content: chr "Prior to her baptism, she experienced the various physical and emotional"
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2018-10-29 17:52:08"

```

```

## .. ..$ description : chr(0)
## .. ..$ heading      : chr(0)
## .. ..$ id           : chr "7"
## .. ..$ language     : chr "en"
## .. ..$ origin       : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 8 :List of 2
## ..$ content: chr "stages of conversion terrifying visions of demons and eternal"
## ..$ meta :List of 7
## .. ..$ author      : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2018-10-29 17:52:08"
## .. ..$ description : chr(0)
## .. ..$ heading      : chr(0)
## .. ..$ id           : chr "8"
## .. ..$ language     : chr "en"
## .. ..$ origin       : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 9 :List of 2
## ..$ content: chr "perdition: extreme feelings of ecstasy and depression; protracted"
## ..$ meta :List of 7
## .. ..$ author      : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2018-10-29 17:52:08"
## .. ..$ description : chr(0)
## .. ..$ heading      : chr(0)
## .. ..$ id           : chr "9"
## .. ..$ language     : chr "en"
## .. ..$ origin       : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 10:List of 2
## ..$ content: chr "periods of meditation, fasting, and prayer; ennui and fever; energy and"
## ..$ meta :List of 7
## .. ..$ author      : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2018-10-29 17:52:08"
## .. ..$ description : chr(0)
## .. ..$ heading      : chr(0)
## .. ..$ id           : chr "10"

```

```

## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 11:List of 2
## ..$ content: chr "vigor. In 1811 she married Joseph Lee, who pastored an African-American"
## ..$ meta      :List of 7
## .. ..$ author      : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2018-10-29 17:52:08"
## .. ..$ description : chr(0)
## .. ..$ heading      : chr(0)
## .. ..$ id           : chr "11"
## .. ..$ language     : chr "en"
## .. ..$ origin       : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 12:List of 2
## ..$ content: chr "church in Snow Hill, New Jersey. They had six children, four of whom"
## ..$ meta      :List of 7
## .. ..$ author      : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2018-10-29 17:52:08"
## .. ..$ description : chr(0)
## .. ..$ heading      : chr(0)
## .. ..$ id           : chr "12"
## .. ..$ language     : chr "en"
## .. ..$ origin       : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 13:List of 2
## ..$ content: chr "died in infancy."
## ..$ meta      :List of 7
## .. ..$ author      : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2018-10-29 17:52:08"
## .. ..$ description : chr(0)
## .. ..$ heading      : chr(0)
## .. ..$ id           : chr "13"
## .. ..$ language     : chr "en"
## .. ..$ origin       : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"

```



```
##   .- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## - attr(*, "class")= chr [1:2] "SimpleCorpus" "Corpus"
```

```
strwrap(corpus[[1]])
```

```
## [1] "In 1804, after several months of profound spiritual anxiety,"
## [2] "Jarena Lee"
```

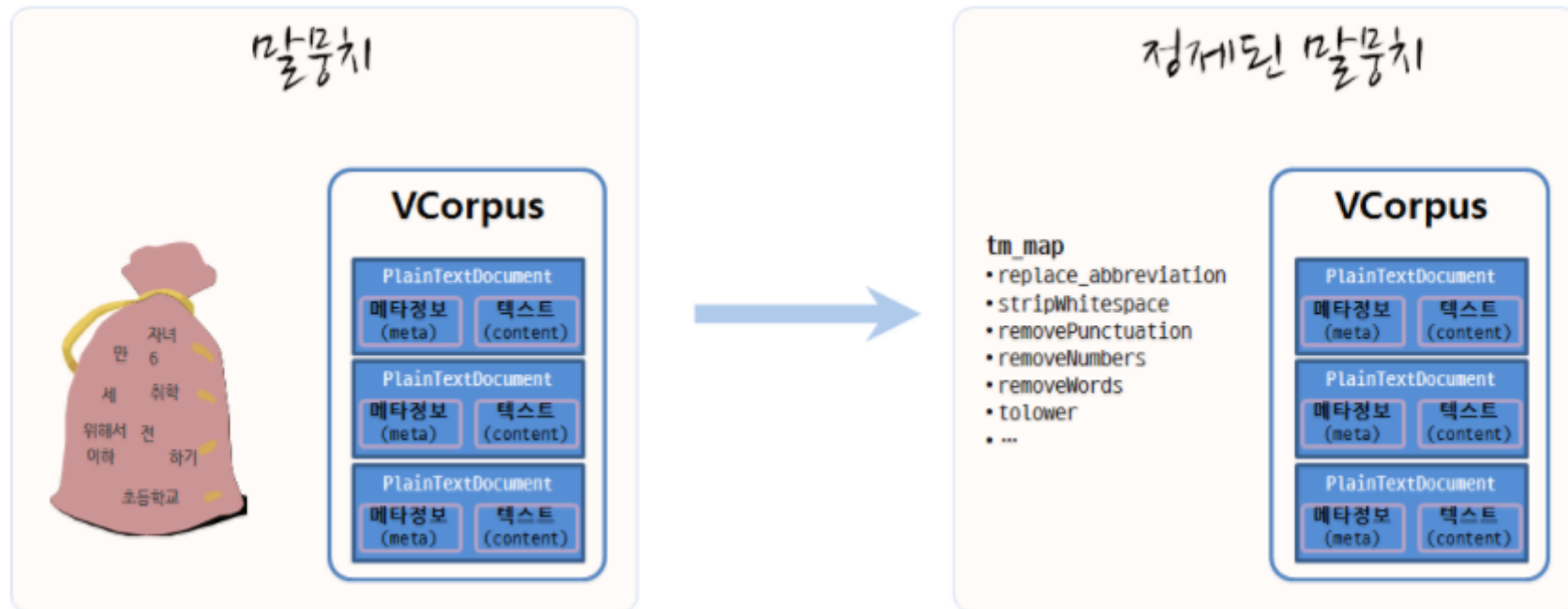
04-2. 말뭉치 정제

텍스트를 말뭉치 객체로 저장시켜 놓으면, 말뭉치 정제 과정을 거치게 된다.
기본 정제함수는 tm팩키지를 활용하고,
추가적으로 qdap 팩키지 특수 말뭉치 함수도 활용한다.

04-3. 문장 부호를 제거하기

tm_map : Transformations On Corpora
Corpora : Corpus의 복수형

Usage : tm_map(x, FUN, ...)
x : A corpus (하나의 Corpus)
FUN : text의 문서를 변경을 취하는 함수



(이미지 참조) <https://statklee.github.io/ml/ml-wordcloud.html> (<https://statklee.github.io/ml/ml-wordcloud.html>)

```
corpus <- tm_map(corpus, removePunctuation)
```

```
## Warning in tm_map.SimpleCorpus(corpus, removePunctuation): transformation
## drops documents
```

```
strwrap(corpus[[13]])
```

```
## [1] "died in infancy"
```

04-4. 특정 문자를 제거하기(여기서는 숫자 0~9를 제거)

```
### 04-5 Text 문서로부터 단어들을 삭제하는 함수  
removeWords(x, words)  
x : 문자 또는 문서 document  
words : 벡터로 주어진 단어를 삭제한다.
```

```
corpus <- tm_map(corpus, removeNumbers)
```

```
## Warning in tm_map.SimpleCorpus(corpus, removeNumbers): transformation drops  
## documents
```

```
#corpus <- tm_map(corpus, removeWords, c("1804", "0", "1", "2", "3", "4", "5", "6", "7", "8", "9")) 교재 내용 수정 필요  
strwrap(corpus[[1]])
```

```
## [1] "In after several months of profound spiritual anxiety Jarena Lee"
```

04-6. 입력 텍스트가 영문인 경우, 대문자를 소문자로 변환하기

```
corpus <- tm_map(corpus, tolower)
```

```
## Warning in tm_map.SimpleCorpus(corpus, tolower): transformation drops  
## documents
```

```
corpus[[1]]$content
```

```
## [1] "in after several months of profound spiritual anxiety jarena lee"
```

05. 전처리가 후, 문서 내용 확인

```
strwrap(corpus[[1]])
```

```
## [1] "in after several months of profound spiritual anxiety jarena lee"
```

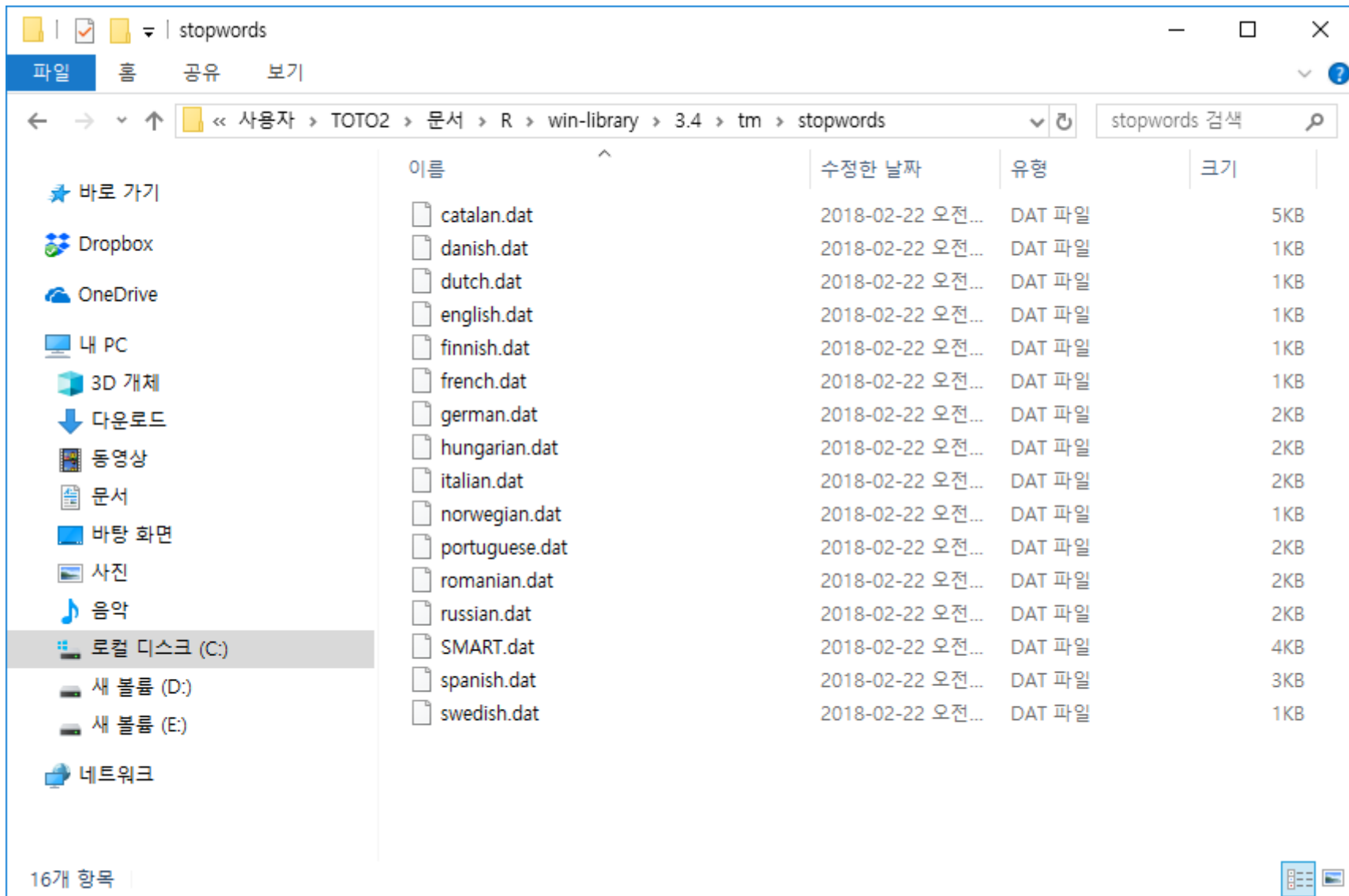
05-1. 전처리가 수행된 결과를 확인(2)

```
strwrap(corpus[[2]])
```

```
## [1] "moved from new jersey to philadelphia there she labored as a"  
## [2] "domestic"
```

06. 불용어 처리 수행

R 패키지 내장된 영문 불용어 사전활용.
가. 경로확인 .libPaths()
나. 사전 확인
stopwords/english.dat 추가 편집



```
.libPaths()
```

```
## [1] "C:/Users/WITHJS/Documents/R/win-library/3.4"
## [2] "C:/Program Files/R/R-3.4.0/library"
```

```
strwrap(corpus[[1]])
```

```
## [1] "in after several months of profound spiritual anxiety jarena lee"
```

06-1 영어 불용어 제거하기(조사[of등], 전치사(after, from 등))

```
corpus <- tm_map(corpus, removeWords, stopwords("english"))
```

```
## Warning in tm_map.SimpleCorpus(corpus, removeWords, stopwords("english")):  
## transformation drops documents
```

```
strwrap(corpus[[1]])
```

```
## [1] "several months profound spiritual anxiety jarena lee"
```

07. 단어문서행렬(Term Document Matrix, TDM)

단어문서행렬을 통한 방법이 일반적으로 많이 이용됨.
데이터 프레임 : DataframeSource,
벡터인 경우 : VectorSource

단어문서행렬 생성 : TermDocumentMatrix
불용어 처리가 된 corpus를 TermDocumentMatrix 변환 후, 각 단어의 출현 빈도를 측정.

07-1 TermDocumentMatrix 로 변환

```
tdm <- TermDocumentMatrix(corpus, control=list(tokenize="scan", wordLengths=c(2,7)))  
inspect(tdm)
```

```
## <<TermDocumentMatrix (terms: 46, documents: 13)>>
## Non-/sparse entries: 51/547
## Sparsity          : 91%
## Maximal term length: 7
## Weighting          : term frequency (tf)
## Sample            :
##               Docs
## Terms      1 10 11 12 2 3 5 6 7 8
## anxiety 1  0  0  0 0 0 0 0 0 0
## church   0  0  0  1 0 0 0 1 0 0
## jarena   1  0  0  0 0 0 0 0 0 0
## jersey   0  0  0  1 1 0 0 0 0 0
## labored  0  0  0  0 1 0 0 0 0 0
## lee      1  0  1  0 0 0 0 1 0 0
## months   1  0  0  0 0 0 0 0 0 0
## moved    0  0  0  0 1 0 0 0 0 0
## new      0  0  0  1 1 0 0 0 0 0
## several  1  0  0  0 0 0 0 0 0 0
```

```
nTerms(tdm)
```

```
## [1] 46
```

07-2 문서단어행렬 변환.

DocumentTermMatrix()는 TermDocumentMatrix() 함수로 통해 나온 행렬을 전치(transpose)시킨 것이다.

```
dtm <- DocumentTermMatrix(corpus)
dtm
```

```
## <<DocumentTermMatrix (documents: 13, terms: 72)>>  
## Non-/sparse entries: 79/857  
## Sparsity          : 92%  
## Maximal term length: 15  
## Weighting          : term frequency (tf)
```

```
dtm2 <- as.matrix(dtm)  
dtm2
```



```

##      Terms
## Docs anxiety jarena lee months profound several spiritual domestic jersey
## 1      1      1      1      1      1      1      1      0      0
## 2      0      0      0      0      0      0      0      1      1
## 3      0      0      0      0      0      0      0      0      0
## 4      0      0      0      0      0      0      0      0      0
## 5      0      0      0      0      0      0      0      0      0
## 6      0      0      1      0      0      0      0      0      0
## 7      0      0      0      0      0      0      0      0      0
## 8      0      0      0      0      0      0      0      0      0
## 9      0      0      0      0      0      0      0      0      0
## 10     0      0      0      0      0      0      0      0      0
## 11     0      0      1      0      0      0      0      0      0
## 12     0      0      0      0      0      0      0      0      1
## 13     0      0      0      0      0      0      0      0      0

##      Terms
## Docs labored moved new philadelphia among catholics congregations mixed
## 1      0      0      0      0      0      0      0      0      0
## 2      1      1      1      1      0      0      0      0      0
## 3      0      0      0      0      1      1      1      1      1
## 4      0      0      0      0      0      0      0      1      0
## 5      0      0      0      0      0      0      0      0      0
## 6      0      0      0      0      0      0      0      0      0
## 7      0      0      0      0      0      0      0      0      0
## 8      0      0      0      0      0      0      0      0      0
## 9      0      0      0      0      0      0      0      0      0
## 10     0      0      0      0      0      0      0      0      0
## 11     0      0      0      0      0      0      0      0      0
## 12     0      0      1      0      0      0      0      0      0
## 13     0      0      0      0      0      0      0      0      0

##      Terms
## Docs roman white worshiped hearing inspired methodists sermon african
## 1      0      0      0      0      0      0      0      0
## 2      0      0      0      0      0      0      0      0
## 3      1      1      1      0      0      0      0      0
## 4      0      0      0      1      1      1      1      0
## 5      0      0      0      0      0      0      0      1
## 6      0      0      0      0      0      1      0      0
## 7      0      0      0      0      0      0      0      0

```

```
## 8      0      0      0      0      0      0      0      0      0
## 9      0      0      0      0      0      0      0      0      0
## 10     0      0      0      0      0      0      0      0      0
## 11     0      0      0      0      0      0      0      0      0
## 12     0      0      0      0      0      0      0      0      0
## 13     0      0      0      0      0      0      0      0      0
```

```
##      Terms
```

```
## Docs allen bethel founder methodist reverend richard baptized church
```

```
## 1      0      0      0      0      0      0      0      0      0
## 2      0      0      0      0      0      0      0      0      0
## 3      0      0      0      0      0      0      0      0      0
## 4      0      0      0      0      0      0      0      0      0
## 5      1      1      1      1      1      1      1      0      0
## 6      0      0      0      0      0      0      0      1      1
## 7      0      0      0      0      0      0      0      0      0
## 8      0      0      0      0      0      0      0      0      0
## 9      0      0      0      0      0      0      0      0      0
## 10     0      0      0      0      0      0      0      0      0
## 11     0      0      0      0      0      0      0      0      0
## 12     0      0      0      0      0      0      0      0      1
## 13     0      0      0      0      0      0      0      0      0
```

```
##      Terms
```

```
## Docs episcopal joined baptism emotional experienced physical prior various
```

```
## 1      0      0      0      0      0      0      0      0      0
## 2      0      0      0      0      0      0      0      0      0
## 3      0      0      0      0      0      0      0      0      0
## 4      0      0      0      0      0      0      0      0      0
## 5      0      0      0      0      0      0      0      0      0
## 6      1      1      0      0      0      0      0      0      0
## 7      0      0      1      1      1      1      1      1      1
## 8      0      0      0      0      0      0      0      0      0
## 9      0      0      0      0      0      0      0      0      0
## 10     0      0      0      0      0      0      0      0      0
## 11     0      0      0      0      0      0      0      0      0
## 12     0      0      0      0      0      0      0      0      0
## 13     0      0      0      0      0      0      0      0      0
```

```
##      Terms
```

```
## Docs conversion demons eternal stages terrifying visions depression
```

```
## 1      0      0      0      0      0      0      0      0
```

```

## 2      0      0      0      0      0      0      0
## 3      0      0      0      0      0      0      0
## 4      0      0      0      0      0      0      0
## 5      0      0      0      0      0      0      0
## 6      0      0      0      0      0      0      0
## 7      0      0      0      0      0      0      0
## 8      1      1      1      1      1      1      0
## 9      0      0      0      0      0      0      1
## 10     0      0      0      0      0      0      0
## 11     0      0      0      0      0      0      0
## 12     0      0      0      0      0      0      0
## 13     0      0      0      0      0      0      0
##      Terms
## Docs ecstasy extreme feelings perdition protracted energy ennui fasting
## 1      0      0      0      0      0      0      0      0
## 2      0      0      0      0      0      0      0      0
## 3      0      0      0      0      0      0      0      0
## 4      0      0      0      0      0      0      0      0
## 5      0      0      0      0      0      0      0      0
## 6      0      0      0      0      0      0      0      0
## 7      0      0      0      0      0      0      0      0
## 8      0      0      0      0      0      0      0      0
## 9      1      1      1      1      1      0      0      0
## 10     0      0      0      0      0      1      1      1
## 11     0      0      0      0      0      0      0      0
## 12     0      0      0      0      0      0      0      0
## 13     0      0      0      0      0      0      0      0
##      Terms
## Docs fever meditation periods prayer africanamerican joseph married
## 1      0      0      0      0      0      0      0
## 2      0      0      0      0      0      0      0
## 3      0      0      0      0      0      0      0
## 4      0      0      0      0      0      0      0
## 5      0      0      0      0      0      0      0
## 6      0      0      0      0      0      0      0
## 7      0      0      0      0      0      0      0
## 8      0      0      0      0      0      0      0
## 9      0      0      0      0      0      0      0
## 10     1      1      1      1      0      0      0

```

```
##      11      0      0      0      0      1      1      1
##      12      0      0      0      0      0      0      0
##      13      0      0      0      0      0      0      0
##      Terms
## Docs pastored vigor children four hill six snow died infancy
##      1      0      0      0      0      0      0      0      0
##      2      0      0      0      0      0      0      0      0
##      3      0      0      0      0      0      0      0      0
##      4      0      0      0      0      0      0      0      0
##      5      0      0      0      0      0      0      0      0
##      6      0      0      0      0      0      0      0      0
##      7      0      0      0      0      0      0      0      0
##      8      0      0      0      0      0      0      0      0
##      9      0      0      0      0      0      0      0      0
##     10      0      0      0      0      0      0      0      0
##     11      1      1      0      0      0      0      0      0
##     12      0      0      1      1      1      1      0      0
##     13      0      0      0      0      0      0      1      1
```

07-3 발생빈도가 높은 상위 5개를 출력하기

```
print(frequency)
```

```
## function (x, ...)
## UseMethod("frequency")
## <bytecode: 0x000000001861fb68>
## <environment: namespace:stats>
```

```
frequency <- colSums(dtm2) # 단어 빈도수 구하기
frequency <- sort(frequency, decreasing=T) # 정렬하기
frequency[1:5]
```

```
##      lee      jersey      new congregations      methodists
##      3          2          2          2          2
```

07-4 발생빈도가 높은 상위 20개를 막대 그래프 형태로 출력하기

```
barplot(frequency[1:20], las=3)
```

