Lab 4 Exp 1

Date: 25.04.2022

AIM: WAP in R language to print the word count of a given data.

CODE:

```r
df <- read.csv('https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv')

count_str <- function(df){
  count = 0
  for (i in df){
    for (j in i){
      if (typeof(j) =="character"){
        if(nchar(j) > 0){
          count = count + 1
        }
    }

    }
  }
  return(count)
}

count_str(df)

count = 0
for (i in df$Cabin){

  if(nchar(i) > 0){
    count = count + 1
  }
}

count

```

9:27 count_str(df) R Script

OUTPUT:

```
> count
[1] 204
>
```

Lab 4 Exp 2

Date: 25.04.2022

AIM: Write a program in R Language to generate a word cloud

CODE:

```r
# Load
library("tm")
library("SnowballC")
library("wordcloud")
library("RColorBrewer")


# Read the text file from internet
filePath <- "http://www.sthda.com/sthda/RDoc/example-files/martin-luther-king-i-have-a-dream-speech.txt"
text <- readLines(filePath)


# Load the data as a corpus
docs <- Corpus(VectorSource(text))

toSpace <- content_transformer(function (x , pattern ) gsub(pattern, " ", x))
docs <- tm_map(docs, toSpace, "/")
docs <- tm_map(docs, toSpace, "@")
docs <- tm_map(docs, toSpace, "\\|")

# Convert the text to lower case
docs <- tm_map(docs, content_transformer(tolower))
# Remove numbers
docs <- tm_map(docs, removeNumbers)
# Remove english common stopwords
docs <- tm_map(docs, removeWords, stopwords("english"))
# Remove your own stop word
# specify your stopwords as a character vector
docs <- tm_map(docs, removeWords, c("blabla1", "blabla2"))
# Remove punctuations
docs <- tm_map(docs, removePunctuation)
# Eliminate extra white spaces
docs <- tm_map(docs, stripWhitespace)
# Text stemming
# docs <- tm_map(docs, stemDocument)

dtm <- TermDocumentMatrix(docs)
m <- as.matrix(dtm)
v <- sort(rowSums(m),decreasing=TRUE)
d <- data.frame(word = names(v),freq=v)
head(d, 10)

set.seed(1234)
wordcloud(words = d$word, freq = d$freq, min.freq = 1,
          max.words=200, random.order=FALSE, rot.per=0.35,
          colors=brewer.pal(8, "Dark2"))

```

19:37 (Top Level) R Script

OUTPUT:

```
> head(d, 10)
             word freq
will         will   17
freedom   freedom   13
ring         ring   12
dream       dream   11
day           day   11
let           let   11
every       every    9
one           one    8
able         able    8
together together    7
>
```