

# Word Cloud using Language R

Saturday, June 20, 2015

7:17 PM

The word cloud show the most frequently used in a given pages of text. For example we can check the frequently used words in Tweets, face book comments, articles in news paper.

Lets do the word cloud for an article in NewYork Times new paper.

1. Open the NYTimes.com and select an article and create a text file by name article.txt and save it in your harddisk.
2. Open R Studio
3. Install the required packages

```
install.packages(c("tm", "wordcloud",  
"RColorBrewer"))
```

4. Load the installed packages

```
library(tm)  
library(wordcloud)  
library(RColorBrewer)
```

5. Create [corpus](#) from text file

```
articleCorpus <-  
readLines("/Users/jbonu/git/BusinessIntelligenceDa  
taMining/WordCloud/NYTMArticle.txt")  
mycorpus = Corpus(VectorSource(articleCorpus))
```

6. Transformation

```
mycorpus1 = tm_map(mycorpus, removePunctuation)  
mycorpus2 = tm_map(mycorpus1, removeWords,  
stopwords("english"))  
mycorpus3 = tm_map(mycorpus2, tolower)  
mycorpus4 = tm_map(mycorpus3, stripWhitespace)  
mycorpus5 = tm_map(mycorpus4, PlainTextDocument)
```

## 7. Create TermDocumentMatrix

```
tdm <- TermDocumentMatrix(mycorpus5)
m1 <- as.matrix(tdm)
v1 <- sort(rowSums(m1), decreasing=TRUE)
d1 <- data.frame(word=names(v1), freq=v1)
```

## 8. Create word cloud

```
wordcloud(d1$word,d1
$freq,col=brewer.pal(8,"Set2"), min.freq="5")
```

## 9. WordCloud



