IST-687 Chapter Notes Template: After Completing Please Submit as a PDF.
Originality Assertion: By submitting this file you affirm that this writing is your own.

Name: Yicun Deng
Class: M003
Date: 11/3
Chapter Number: #14 #15
Title of Chapter: Word Perfect Happy Words

Module 11 Text Mining - Lectures and Roundtable

Module 11 Text Mining - Lectures and Roundtable

```
We are focusing on unstructured data this week
TM focus on the frequency, NLP focus on meaning
Packages: tm, wordcloud.
words.vec<-VectorSource(sba)
words.corpus<-Corpus(words.vec)
words.corpus
<<VCorpus>>
Metadata:corpus specific:0, document level(indexed):0
Content:documents:15
Words.corpus<-tm_map(words.corpus,content_transformer(tolower))
Words.corpus<-tm_map(words.corpus,removePunctuation)
Words.corpus<-tm_map(words.corpus,removeWords,stopwords(“english”))

Tdm<-TermDocumentMatrix(words.corpus)
Tdm
TermDocumentMatrix(terms:189,documents:15)
M<-as.matrix(tdm)
wordCounts<-rowSums(m)
wordCounts<-sort(wordCounts,decreasing=TRUE)
head(wordCounts)
cloudFrame<-data.frame(+word=names(sortedMatrix),freq=sortedMatrix)
wordcloud(cloudFrame$word,cloudFrame$freq)

wordcloud(names(wordCounts),wordCounts,min.freq=2,+max.words=50,rot.per=0.35,colors=br
ewer.pal(8,”Dark2”))
```

Module 11 Lectures Part 2

Module 11 Lectures Part 2

```
Pos<-“positive-words.txt”
Neg<-“negative-words.txt”
P<-scan(pos,character(0),sep=”\n”)
N<-scan(neg,character(0),sep=”\n”)
P<-p[-1:-34]
N←n[-1:-34]
```

```
totalWords<-sum(wordCounts)
words<-names(wordCounts)
matched<-match(words,p,nomatch=0)
head(matched,10)
mCounts<-wordCounts[which(matched!=0)]
length(mCounts)
mWords<-names(mCounts)
nPos<-sum(mCounts)
nPos
nNeg<-sum(nCounts)
nWords<-names(nCounts)
nNeg
length(nCoutns)
totalWords<-length(words)
ratioPos<-nPos/totalWords
ratioPos
ratioNeg<-nNeg/totalWords
ratioNeg
```

Module 11 R Coding

Module 11 R Coding

```
Tdm<-TermDocumentMatrix(words.corpus)
Tdm
TermDocumentMatrix(terms:189,documents:15)
M<-as.matrix(tdm)
wordCounts<-rowSums(m)
wordCounts<-sort(wordCounts,decreasing=TRUE)
head(wordCounts)
cloudFrame<-data.frame(+word=names(sortedMatrix),freq=sortedMatrix)
wordcloud(cloudFrame$word,cloudFrame$freq)
totalWords<-sum(wordCounts)
words<-names(wordCounts)
matched<-match(words,p,nomatch=0)
head(matched,10)
mCounts<-wordCounts[which(matched!=0)]
length(mCounts)
mWords<-names(mCounts)
nPos<-sum(mCounts)
nPos
nNeg<-sum(nCounts)
nWords<-names(nCounts)
nNeg
length(nCoutns)
totalWords<-length(words)
```

IST-687 Chapter Notes Template: After Completing Please Submit as a PDF.
Originality Assertion: By submitting this file you affirm that this writing is your own.

```
ratioPos<-nPos/totalWords
ratioPos
ratioNeg<-nNeg/totalWords
ratioNeg
```

Exercise Review

Download several speech script from online corpus and them processed it into dataframe, using the method from class to determine the positive rate and negative rate of the speech. In my opinion, it should be related to the topic of the speech

Question for Class

IMPORTANT: Hi Professor, I had shot an email to my groupmates but they didn't respond, should I come to the class hour to see if they are there or shot another email to them?