Sentimental Analysis

Importing Required Libraries

```r
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.0.4
```

```
## -- Attaching packages ------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.6     v dplyr   1.0.5
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## Warning: package 'ggplot2' was built under R version 4.0.4
```

```
## Warning: package 'tibble' was built under R version 4.0.3
```

```
## Warning: package 'tidyr' was built under R version 4.0.4
```

```
## Warning: package 'readr' was built under R version 4.0.4
```

```
## Warning: package 'purrr' was built under R version 4.0.4
```

```
## Warning: package 'dplyr' was built under R version 4.0.4
```

```
## Warning: package 'stringr' was built under R version 4.0.3
```

```
## Warning: package 'forcats' was built under R version 4.0.4
```

```
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(tidytext)
```

```
## Warning: package 'tidytext' was built under R version 4.0.5
```

```r
library(glue)
```

```
## Warning: package 'glue' was built under R version 4.0.3
```

```
##
## Attaching package: 'glue'
```

```
## The following object is masked from 'package:dplyr':
##
##     collapse
```

```
library(stringr)
```

Importing the folder containing all the text files

```
files = list.files('D:/My Files/Sem-2/R Programming-Akhil Sir/gutenberg')
print(files)
```

```
##  [1] "austen-emma.txt"        "austen-persuasion.txt"
##  [3] "austen-sense.txt"       "bible-kjv.txt"
##  [5] "blake-poems.txt"        "bryant-stories.txt"
##  [7] "burgess-busterbrown.txt" "carroll-alice.txt"
##  [9] "chesterton-ball.txt"    "chesterton-brown.txt"
## [11] "chesterton-thursday.txt" "edgeworth-parents.txt"
## [13] "melville-moby_dick.txt"  "milton-paradise.txt"
## [15] "README"                 "shakespeare-caesar.txt"
## [17] "shakespeare-hamlet.txt"  "shakespeare-macbeth.txt"
## [19] "whitman-leaves.txt"
```

Getting the Filename

```
filename = glue('D:/My Files/Sem-2/R Programming-Akhil Sir/gutenberg/', files[1])
head(filename)
```

```
## D:/My Files/Sem-2/R Programming-Akhil Sir/gutenberg/austen-emma.txt
```

Removing the trailing spaces

```
filename = trimws(filename)
head(filename)
```

```
## D:/My Files/Sem-2/R Programming-Akhil Sir/gutenberg/austen-emma.txt
```

Getting the text in the file

```
filetext = glue(read_file(filename))
```

Get rid of dollar symbol

```
filetext = gsub("\\$", "", filetext)
```

Tokenization- it can separate into words,charackers, or subwords

```
tokens = tibble(text = filetext) %>%
unnest_tokens(word, text)
print(head(tokens))
```

```
## # A tibble: 6 x 1
##   word
##   <chr>
## 1 emma
## 2 by
## 3 jane
## 4 austen
## 5 1816
## 6 volume
```

Getting the sentiment of the text

```
sent = tokens %>%
inner_join(get_sentiments("bing")) %>%
count(sentiment) %>%
spread(sentiment, n, fill = 0) %>%
mutate(sentiment = positive - negative)
```

```
## Joining, by = "word"
```

```
se = as.numeric(sent[1,'sentiment'])
se
```

```
## [1] 2347
```

Creating a function to extract the sentiment

```
GetSentiment = function(file){
  filename = glue('D:/My Files/Sem-2/R Programming-Akhil Sir/gutenberg/', file)
  filename = trimws(filename)

  filetext = glue(read_file(filename))
  filetext = gsub("\\$", "", filetext)

  # tokennize
  tokens = tibble(text = filetext) %>%
    unnest_tokens(word, text)
  sentiment = tokens %>%
  inner_join(get_sentiments("bing")) %>%
  count(sentiment) %>%
  spread(sentiment, n, fill = 0) %>%
  mutate(sentiment = positive - negative)

  fin_sent = as.numeric(sentiment['sentiment'])
  return (fin_sent)
}
```

Testing the function for one text file

```
GetSentiment(files[1])
```

```
## Joining, by = "word"
```

```
## [1] 2347
```

Applying the function to the entire folder and creating result in a dataframe

```
sentiments_frame = data.frame()
for (i in 1:18){
  sentiments_frame = rbind(sentiments_frame,c(files[i], GetSentiment(files[i])))
}
```

```
## Joining, by = "word"
## Joining, by = "word"
## Joining, by = "word"
## Joining, by = "word"
## Joining, by = "word"
## Joining, by = "word"
## Joining, by = "word"
## Joining, by = "word"
## Joining, by = "word"
## Joining, by = "word"
## Joining, by = "word"
## Joining, by = "word"
## Joining, by = "word"
## Joining, by = "word"
## Joining, by = "word"
## Joining, by = "word"
## Joining, by = "word"
## Joining, by = "word"
```

```
print(sentiments_frame)
```

```
##              X.austen.emma.txt. X.2347.
## 1              austen-emma.txt    2347
## 2         austen-persuasion.txt   1272
## 3             austen-sense.txt    1262
## 4               bible-kjv.txt    -1687
## 5              blake-poems.txt     60
## 6            bryant-stories.txt    458
## 7     burgess-busterbrown.txt     58
## 8             carroll-alice.txt    -54
## 9           chesterton-ball.txt   -603
## 10        chesterton-brown.txt   -486
## 11     chesterton-thursday.txt   -619
## 12        edgeworth-parents.txt   1481
## 13       melville-moby_dick.txt  -895
## 14          milton-paradise.txt    74
## 15                     README    -17
## 16        shakespeare-caesar.txt   119
## 17        shakespeare-hamlet.txt    -5
## 18      shakespeare-macbeth.txt    -32
```

Giving column names to the existing dataframe

```
colnames(sentiments_frame) = c('filename', 'sentiment')
sentiments_frame
```

```
##                        filename sentiment
## 1           austen-emma.txt      2347
## 2      austen-persuasion.txt      1272
## 3          austen-sense.txt      1262
## 4             bible-kjv.txt     -1687
## 5           blake-poems.txt        60
## 6         bryant-stories.txt       458
## 7   burgess-busterbrown.txt        58
## 8          carroll-alice.txt       -54
## 9        chesterton-ball.txt      -603
## 10      chesterton-brown.txt      -486
## 11  chesterton-thursday.txt      -619
## 12     edgeworth-parents.txt      1481
## 13    melville-moby_dick.txt      -895
## 14        milton-paradise.txt        74
## 15                   README       -17
## 16    shakespeare-caesar.txt       119
## 17    shakespeare-hamlet.txt        -5
## 18  shakespeare-macbeth.txt       -32
```

Adding a new feature of sentiment labels(positive and negative)

```
to_sentiment = function(x){
  if(x < 0){
    'Negative'
  }else{
    'Positive'
  }
}
```

```
sentiments_frame$sentmt_text = sapply(sentiments_frame$sentiment, to_sentiment)
sentiments_frame
```

```
##                        filename sentiment sentmt_text
## 1           austen-emma.txt      2347    Positive
## 2      austen-persuasion.txt      1272    Positive
## 3          austen-sense.txt      1262    Positive
## 4             bible-kjv.txt     -1687    Negative
## 5           blake-poems.txt        60    Positive
## 6         bryant-stories.txt       458    Positive
## 7   burgess-busterbrown.txt        58    Positive
## 8          carroll-alice.txt       -54    Negative
## 9        chesterton-ball.txt      -603    Negative
## 10      chesterton-brown.txt      -486    Negative
## 11  chesterton-thursday.txt      -619    Negative
## 12     edgeworth-parents.txt      1481    Positive
## 13    melville-moby_dick.txt      -895    Negative
## 14        milton-paradise.txt        74    Positive
## 15                   README       -17    Negative
## 16    shakespeare-caesar.txt       119    Positive
## 17    shakespeare-hamlet.txt        -5    Negative
## 18  shakespeare-macbeth.txt       -32    Negative
```