Sentiment Analysis of ftragedy (https://www.kaggle.com/sharkcpn/french-tragedies/data)

by NISHCHAY CHAWLA

==========

Introduction

What is sentiment analysis?

Sentiment analysis is the computational task of automatically determining what feelings a writer is expressing in text. Sentiment is often framed as a binary distinction (positive vs. negative), but it can also be a more fine-grained, like identifying the specific emotion an author is expressing (like fear, joy or anger).

How is it done?

1. Create or find a list of words associated with strongly positive or negative sentiment(lexicon). Here French Expanded Emotion Lexicon (FEEL (http://advanse.lirmm.fr/feel.php)) was used.
2. Count the number of positive and negative words in the text using lexicon.
3. Analyze the mix of positive to negative words. Many positive words and few negative words indicates positive sentiment (Polarity), while many negative words and few positive words indicates negative sentiment(Polarity).

Explanation

```
#read the file
test1 <- read_file(paste0('~/Desktop/Senti_analysis/ftragedy/',files[1]), locale = locale('f
r'))

#some text processing
test1 <- stripWhitespace(test1)

test_df <- data_frame(text = test1) %>% unnest_tokens(word, text)

test_clean_df <- test_df %>% filter(!word %in% stopwords('french'))

#Comparing with lexicon to get positive, negative and other emotions in text
test_feel_df <- inner_join(test_clean_df, FEEL_lex, by = 'word')

sum(test_feel_df['polarity']=='positive')
```

```
## [1] 1364
```

```
sum(test_feel_df['polarity']=='negative')
```

```
## [1] 1180
```

```
sum(test_feel_df$joy)
```

```
## [1] 111
```

```
# Function to automate the above task
get_polarity_and_sentiments <- function(file){
  file_name <- paste0('~/Desktop/Senti_analysis/ftragedy/',file)
  file_name <- trimws(file_name)
  file_text <- read_file(file_name, locale = locale('fr'))

  clean_text <- stripWhitespace(file_text)

  tokens <- data_frame(text = clean_text) %>% unnest_tokens(word,text)
  tokens <- tokens %>% filter(!word %in% stopwords('french'))


  token_feel_frame <- inner_join(tokens, FEEL_lex, by = 'word')

  sentiments_scorecard <- data.frame(
                      File = file,
                      Total_Positive = sum(token_feel_frame$polarity == 'positive'),
                      Total_Negative = sum(token_feel_frame$polarity == 'negative'),

                      Total_Joy_Words = sum(token_feel_frame$joy),
                      Total_Fear_Words = sum(token_feel_frame$fear),
                      Total_Sadness_Words = sum(token_feel_frame$sadness),
                      Total_Anger_Words = sum(token_feel_frame$anger),
                      Total_Surprise_Words = sum(token_feel_frame$surprise),
                      Total_Disgust_Words = sum(token_feel_frame$disgust))

  sentiments_scorecard <- sentiments_scorecard %>%
                          mutate(Polarity = Total_Positive-Total_Negative)
  return(sentiments_scorecard)
}
```

```
#results same as explanation above
get_polarity_and_sentiments(files[1])
```

File <fctr>	Total_Positive <int>	Total_Negative <int>	Total_Joy_Words <int>	Total
Crebillon_TR-V-1703-Idomenee.txt	1364	1180	111	

1 row | 1-5 of 10 columns

```
polarity_and_sentiments <- data_frame()


for(i in files){
  polarity_and_sentiments <- rbind(polarity_and_sentiments, get_polarity_and_sentiments(i))
}
```

```
head(polarity_and_sentiments,10)
```

File	Total_Positive	Total_Negative	Total_Joy_Words

<fctr>	<int>	<int>	<int>
Crebillon_TR-V-1703-Idomenee.txt	1364	1180	111
Crebillon_TR-V-1707-Atree.txt	1344	987	135
Crebillon_TR-V-1708-Electre.txt	1515	1110	152
Crebillon_TR-V-1711-Rhadamisthe.txt	1314	963	110
Crebillon_TR-V-1717-Semiramis.txt	1606	1023	153
Crebillon_TR-V-1726-Pyrrhus.txt	1669	1104	174
Crebillon_TR-V-1749-Catilina.txt	1625	1106	145
Crebillon_TR-V-1749-Xerces.txt	1797	1175	162
Crebillon_TR-V-1754-Triumvirat.txt	1571	1006	148
PCorneille_TR-V-1639-Medee.txt	1515	965	164

1-10 of 10 rows | 1-4 of 10 columns

```
#most positive
polarity_and_sentiments %>% filter(Total_Positive == max(Total_Positive))
```

File <fctr>	Total_Positive <int>	Total_Negative <int>	Total_Joy_Words <int>	To
PCorneille_TR-V-1661-Toisondor.txt	2140	1097	204	

1 row | 1-5 of 10 columns

```
#most negative
polarity_and_sentiments %>% filter(Total_Negative == max(Total_Negative))
```

File <fctr>	Total_Positive <int>	Total_Negative <int>	Total_Joy_Words <int>	Total
Crebillon_TR-V-1703-Idomenee.txt	1364	1180	111	

1 row | 1-5 of 10 columns