````
---
title: "Secondprocess"
author: "Michael Pearson"
date: "8/17/2020"
output:
  pdf_document: default
  word_document: default
  html_document: default
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
library(dplyr, quietly = TRUE)
library(readr, quietly = TRUE)
library(R.utils, quietly = TRUE)
library(SnowballC, quietly = TRUE)
library(tidyr, quietly = TRUE)
library(data.table, quietly = TRUE)
library(quanteda)
library(readtext)
```

## R Markdown

Do the combi thing for samples



``` {r trigrams except ns_ns, eval = TRUE}
tri_trigrams <- read.csv(file = "~/Documents/Coursera/Capstone Project/
20sample/tri_gram_s_s.csv" ,colClasses = c( NA, NA, NA, NA) )
tri_trigrams <- data.table(tri_trigrams)
combi_tri_s_s <- unite(tri_trigrams, bigrams, c("word1", "word2"), sep = "
")
rm(tri_trigrams)
write.csv(combi_tri_s_s,file = "~/Documents/Coursera/Capstone Project/
20sample/combi_tri_s_s.csv" )
rm(combi_tri_s_s)
tri_trigrams <- read.csv(file = "~/Documents/Coursera/Capstone Project/
20sample/tri_gram_ns_s.csv" ,colClasses = c(NA, NA, NA, NA) )
tri_trigrams <- data.table(tri_trigrams)
combi_tri_ns_s <- unite(tri_trigrams, bigrams, c("word1", "word2"), sep = "
")
rm(tri_trigrams)
write.csv(combi_tri_ns_s,file = "~/Documents/Coursera/Capstone Project/
20sample/combi_tri_ns_s.csv" )
rm(combi_tri_ns_s)
tri_trigrams <- read.csv(file = "~/Documents/Coursera/Capstone Project/
20sample/tri_gram_s_ns.csv" ,colClasses = c(NA, NA, NA, NA) )
tri_trigrams <- data.table(tri_trigrams)
combi_tri_s_ns <- unite(tri_trigrams, bigrams, c("word1", "word2"), sep = "
")
rm(tri_trigrams)
````
````

```
write.csv(combi_tri_s_ns,file = "~/Documents/Coursera/Capstone Project/
20sample/combi_tri_s_ns.csv" )
rm(combi_tri_s_ns)
```

now for no stemming and no stopwords

```{r for no stemming an no stopwords, eval = TRUE}
tri_trigrams <- read.csv(file = "~/Documents/Coursera/Capstone Project/
20sample/tri_gram_ns_ns.csv" ,colClasses = c(NA, NA, NA, NA) )
tri_trigrams <- data.table(tri_trigrams)
combi_tri_ns_ns <- unite(tri_trigrams, bigrams, c("word1", "word2"), sep =
" ")
rm(tri_trigrams)
write.csv(combi_tri_ns_ns,file = "~/Documents/Coursera/Capstone Project/
20sample/combi_tri_ns_ns.csv" )
rm(combi_tri_ns_ns)
```

Now for the test batch

```{r test tri grams, eval = TRUE}
tri_trigrams <- read.csv(file = "~/Documents/Coursera/Capstone Project/
20test/tri_gram_test.csv" ,colClasses = c(NA, NA, NA, NA) )
tri_trigrams <- data.table(tri_trigrams)
combi_tri_ns_ns <- unite(tri_trigrams, bigrams, c("word1", "word2"), sep =
" ")
rm(tri_trigrams)
write.csv(combi_tri_ns_ns,file = "~/Documents/Coursera/Capstone Project/
20test/combi_tri_test.csv" )
rm(combi_tri_ns_ns)
```

quadgrams

```{r quadgrams except ns_ns, eval = TRUE}
quad_quadgrams <- read.csv(file = "/Users/mutecypher/Documents/Coursera/
Capstone Project/20sample/quad_gram_s_s.csv",colClasses = c( NA, NA, NA,
NA,NA) )
quad_quadgrams <- data.table(quad_quadgrams)
combi_quad_s_s <- unite(quad_quadgrams, trigrams, c("word1", "word2",
"word3"), sep = " ")
rm(quad_quadgrams)
write.csv(combi_quad_s_s,file = "/Users/mutecypher/Documents/Coursera/
Capstone Project/20sample/combi_quad_s_s.csv" )
rm(combi_quad_s_s)
quad_quadgrams <- read.csv(file = "/Users/mutecypher/Documents/Coursera/
Capstone Project/20sample/quad_gram_ns_s.csv",colClasses = c( NA, NA, NA,
NA,NA) )
quad_quadgrams <- data.table(quad_quadgrams)
combi_quad_ns_s <- unite(quad_quadgrams, trigrams, c("word1", "word2",
"word3"), sep = " ")
rm(quad_quadgrams)
```

```
write.csv(combi_quad_ns_s,file = "/Users/mutecypher/Documents/Coursera/
Capstone Project/20sample/combi_quad_ns_s.csv" )
rm(combi_quad_ns_s)
quad_quadgrams <- read.csv(file = "/Users/mutecypher/Documents/Coursera/
Capstone Project/20sample/quad_gram_s_ns.csv",colClasses = c( NA, NA, NA,
NA,NA) )
quad_quadgrams <- data.table(quad_quadgrams)
combi_quad_s_ns <- unite(quad_quadgrams, trigrams, c("word1", "word2",
"word3"), sep = " ")
rm(quad_quadgrams)
write.csv(combi_quad_s_ns,file = "/Users/mutecypher/Documents/Coursera/
Capstone Project/20sample/combi_quad_s_ns.csv" )
rm(combi_quad_s_ns)
```

Now for ns_ns

```{r quad ns_ns, eval = TRUE}
quad_quadgrams <- read.csv(file = "/Users/mutecypher/Documents/Coursera/
Capstone Project/20sample/quad_gram_ns_ns.csv",colClasses = c( NA, NA, NA,
NA,NA) )
quad_quadgrams <- data.table(quad_quadgrams)
combi_quad_ns_ns <- unite(quad_quadgrams, trigrams, c("word1", "word2",
"word3"), sep = " ")
rm(quad_quadgrams)
write.csv(combi_quad_ns_ns,file = "/Users/mutecypher/Documents/Coursera/
Capstone Project/20sample/combi_quad_ns_ns.csv" )
rm(combi_quad_ns_ns)
```

Now for test

```{r quad grams for test, eval = TRUE}
quad_quadgrams <- read.csv(file = "/Users/mutecypher/Documents/Coursera/
Capstone Project/20test/quad_gram_test.csv",colClasses = c( NA, NA, NA,
NA,NA) )
quad_quadgrams <- data.table(quad_quadgrams)
combi_quad_ns_ns <- unite(quad_quadgrams, trigrams, c("word1", "word2",
"word3"), sep = " ")
rm(quad_quadgrams)
write.csv(combi_quad_ns_ns,file = "/Users/mutecypher/Documents/Coursera/
Capstone Project/20test/combi_quad_test.csv" )
rm(combi_quad_ns_ns)
```

quingrams

``` {r quingrams except ns_ns, eval = TRUE}
quin_quingrams <- read.csv(file = "/Users/mutecypher/Documents/Coursera/
Capstone Project/20sample/quin_gram_s_s.csv",colClasses = c( NA, NA, NA,
NA,NA) )
```

```
quin_quingrams <- data.table(quin_quingrams)
combi_quin_s_s <- unite(quin_quingrams, quadgrams, c("word1", "word2",
"word3", "word4"), sep = " ")
rm(quin_quingrams)
write.csv(combi_quin_s_s,file = "/Users/mutecypher/Documents/Coursera/
Capstone Project/20sample/combi_quin_s_s.csv" )
rm(combi_quin_s_s)
quin_quingrams <- read.csv(file = "/Users/mutecypher/Documents/Coursera/
Capstone Project/20sample/quin_gram_ns_s.csv",colClasses = c( NA, NA, NA,
NA,NA) )
quin_quingrams <- data.table(quin_quingrams)
combi_quin_ns_s <- unite(quin_quingrams, quadgrams, c("word1", "word2",
"word3", "word4"), sep = " ")
rm(quin_quingrams)
write.csv(combi_quin_ns_s,file = "/Users/mutecypher/Documents/Coursera/
Capstone Project/20sample/combi_quin_ns_s.csv" )
rm(combi_quin_ns_s)
quin_quingrams <- read.csv(file = "/Users/mutecypher/Documents/Coursera/
Capstone Project/20sample/quin_gram_s_ns.csv",colClasses = c( NA, NA, NA,
NA,NA) )
quin_quingrams <- data.table(quin_quingrams)
combi_quin_s_ns <- unite(quin_quingrams, quadgrams, c("word1", "word2",
"word3", "word4"), sep = " ")
rm(quin_quingrams)
write.csv(combi_quin_s_ns,file = "/Users/mutecypher/Documents/Coursera/
Capstone Project/20sample/combi_quin_s_ns.csv" )
rm(combi_quin_s_ns)
```

Now for ns_ns

```{r quin with ns_ns, eval = TRUE}
quin_quingrams <- read.csv(file = "/Users/mutecypher/Documents/Coursera/
Capstone Project/20sample/quin_gram_ns_ns.csv",colClasses = c( NA, NA, NA,
NA,NA) )
quin_quingrams <- data.table(quin_quingrams)
combi_quin_ns_ns <- unite(quin_quingrams, quadgrams, c("word1", "word2",
"word3", "word4"), sep = " ")
rm(quin_quingrams)
write.csv(combi_quin_ns_ns,file = "/Users/mutecypher/Documents/Coursera/
Capstone Project/20sample/combi_quin_ns_ns.csv")
rm(combi_quin_ns_ns)
```

Lastly, for test

```{r quin for test, eval = TRUE}
quin_quingrams <- read.csv(file = "/Users/mutecypher/Documents/Coursera/
Capstone Project/20test/quin_gram_test.csv",colClasses = c( NA, NA, NA,
NA,NA) )
quin_quingrams <- data.table(quin_quingrams)
```

```
combi_quin_ns_ns <- unite(quin_quingrams, quadgrams, c("word1", "word2",
"word3", "word4"), sep = " ")
rm(quin_quingrams)
write.csv(combi_quin_ns_ns,file = "/Users/mutecypher/Documents/Coursera/
Capstone Project/20test/combi_quin_test.csv")
rm(combi_quin_ns_ns)
```

Including Stuff at the end