Project 1

Will Graham; Richelle Lee; Robin Lin

2022-09-29

Ctrl + Shift + H: Set Working Directory

a <- scan("pg10.txt",what="character",skip=104) ## skip contents  
n <- length(a)  
a <- a[-((n-2886):n)] ## strip license  
a <- a[-grep("[0123456789]:[0123456789]",a)] ## strip out verse numbers  
  
# a <- a[1:100] # test

A Function Splitting Punctuation Marks

indexes\_split\_punct <- function(string){  
 ## Separating words with punctuation marks  
 indexes <- grep("[^a-z ^A-Z]",a) ## obtain the indexes of words containing non-letter characters   
 return(indexes)  
}  
  
split\_punct <- function(b, test){  
 punct\_indexes <- indexes\_split\_punct(b)  
 punct <- b[punct\_indexes]  
 new\_b <- c(1:(length(b) + length(punct)))  
 tracker <- 0  
 if (test == TRUE){  
 b <- b[1:100]  
 new\_b <- new\_b[1:115]  
 }  
   
   
 for (i in c(1:length(b))) {  
 if (b[i] %in% punct) {  
 new\_b[i+tracker] <- substr(b[i], 1, nchar(b[i]) - 1)  
 new\_b[i + tracker + 1] <- substr(b[i], nchar(b[i]), nchar(b[i]))  
 tracker <- tracker + 1  
 } else {new\_b[i + tracker] <- b[i]}  
 }  
 return(new\_b)  
}

#Lowering capitals  
new\_b <- split\_punct(a, TRUE)  
lower\_new\_b <- tolower(new\_b)  
unique\_words <- unique(lower\_new\_b)  
index\_match <- match(lower\_new\_b, unique\_words)  
tabulate\_index\_match <- tabulate(index\_match)  
sort\_results <- sort(tabulate\_index\_match, index.return = TRUE, decreasing = TRUE)  
top\_500 <- sort\_results$ix[1 : 10]  
result\_list <- unique\_words[top\_500]

common\_word\_match <- match(lower\_new\_b, result\_list)  
first\_column <- common\_word\_match[1 : (length(common\_word\_match) - 2)]  
second\_column <- first\_column[2 : (length(common\_word\_match) - 1)]  
third\_column <- first\_column[3 : length(common\_word\_match)]  
T\_array <- cbind(first\_column, second\_column, third\_column)  
  
T\_array <- T\_array[which(rowSums(T\_array) |> is.na() == FALSE),]  
print(result\_list[T\_array])

## [1] "." "and" "." "." "and" "there"   
## [7] "was" "light" "." "the" "the" "darkness"  
## [13] "." "," "and" "." "." "and"   
## [19] "darkness" "and" "and" "there" "was" "light"   
## [25] "." "and" "light" "darkness" "." "and"   
## [31] "and" "the" "and" "and" "the" "was"   
## [37] "the" "god" "was" "light" "." "and"   
## [43] "god" "," "." "and" "god" "the"   
## [49] "darkness" "the" "god"