**############################################################################################################ DIFFERENT WAYS OF OMITTING AND COMBINING WORDS IN A DOCUMENT-TERM MATRIX ############ #################################################################################################**

**################################################################################################################## OMITTING A SINGLE WORD FROM A DOCUMENT-TERM MATRIX ############################ #################################################################################################**

**corpOLD.dtm**

**dim(corpOLD.dtm)**

**labels(corpOLD.dtm)$Terms**

**ind=labels(corpOLD.dtm)$Terms=="purposes"**

**ind**

**corpNEW.dtm=corpOLD.dtm[,!ind]**

**dim(corpNEW.dtm)**

**labels(corpNEW.dtm)$Terms**

**as.matrix(corpNEW.dtm)**

**################################################################################################################## OMITTING SEVERAL WORDS FROM A DOCUMENT-TERM MATRIX ############################ #################################################################################################**

**corpOLD.dtm**

**dim(corpOLD.dtm)**

**labels(corpOLD.dtm)$Terms**

**delete=c("floor","proposition","purposes")**

**## write down terms to delete**

**length(delete)**

**corpTEMP.dtm=corpOLD.dtm**

**for (i in 1:length(delete)) {**

**ind=labels(corpTEMP.dtm)$Terms==delete[i]**

**corpTEMP.dtm=corpTEMP.dtm[,!ind]**

**}**

**corpNEW.dtm=corpTEMP.dtm**

**dim(corpNEW.dtm)**

**labels(corpNEW.dtm)$Terms**

**as.matrix(corpNEW.dtm)**

**################################################################################################################## COMBINING FREQUENCIES ACROSS SEVERAL TERMS #################################### ################# AND DELETING FREQUENCIES OF INDIVIDUAL TERMS ##################################**

**#################################################################################################**

**#### combining terms in dtm matrix: corpOLD.dtm**

**#### terms to be combined in: comb**

**#### new dtm matrix: corpNEW.dtm**

**library(slam)**

**corpOLD.dtm**

**dim(corpOLD.dtm)**

**comb=c("purpose","purposes") ## words to combine**

**groupcomb=toString(comb) ## label for combined terms**

**mm=length(comb)**

**corpNEW.dtm=corpOLD.dtm**

**h=labels(corpNEW.dtm)$Terms**

**ind1=c(1:length(h))**

**for (j in 1:mm) {**

**ind=h==comb[j]**

**index=ind1[ind==TRUE]**

**h[index]=groupcomb ## changing the labels of the terms to be combined to the group label**

**}**

**dimnames(corpNEW.dtm)$Terms=h**

**labels(corpNEW.dtm)$Terms**

**corpNEW.dtm=rollup(corpNEW.dtm,2L,labels(corpNEW.dtm)$Terms,FUN=sum)**

**dim(corpNEW.dtm)**

**labels(corpNEW.dtm)$Terms**

**as.matrix(corpNEW.dtm)**

**################################################################################################# ############# FUNCTION combine\_terms TO COMBINE TERMS (COLUMNS) IN A DT MATRIX ##################**

**##################### AND DELETING FREQUENCIES OF INDIVIDUAL TERMS ##############################**

**#################################################################################################**

**#### combining terms in dtm matrix: corpOLD.dtm**

**#### terms to be combined in: comb**

**#### new dtm matrix: corpNEW.dtm**

**combine\_terms = function(corpOLD.dtm,comb) {**

**groupcomb=toString(comb) ## label for combined terms**

**mm=length(comb)**

**corpNEW.dtm=corpOLD.dtm**

**h=labels(corpNEW.dtm)$Terms**

**ind1=c(1:length(h))**

**for (j in 1:mm) {**

**ind=h==comb[j]**

**index=ind1[ind==TRUE]**

**h[index]=groupcomb[1] ## changing the labels of the terms to be combined to the group label**

**}**

**dimnames(corpNEW.dtm)$Terms=h**

**labels(corpNEW.dtm)$Terms**

**corpNEW.dtm=rollup(corpNEW.dtm,2L,labels(corpNEW.dtm)$Terms,FUN=sum)**

**return(corpNEW.dtm)**

**}**

**library(slam)**

**corpOLD.dtm**

**dim(corpOLD.dtm)**

**labels(corpOLD.dtm)$Terms**

**as.matrix(corpOLD.dtm)**

**comb=c("purpose","purposes") ## words to combine**

**corpNEW.dtm=combine\_terms(corpOLD.dtm,comb)**

**dim(corpNEW.dtm)**

**labels(corpNEW.dtm)$Terms**

**as.matrix(corpNEW.dtm)**

**############################################################################################### ################# ANOTHER ALTERNATIVE WAY OF COMBINING TERMS IN A DT MATRIX ###################**

**###############################################################################################**

**#### combining terms in dtm matrix: corpOLD.dtm**

**#### terms to be combined in: comb**

**#### new dtm matrix: corpNEW.dtm**

**corpOLD.dtm**

**dim(corpOLD.dtm)**

**labels(corpOLD.dtm)$Terms**

**comb=c("purpose","purposes") ## words to combine**

**groupcomb=toString(comb)**

**groupcomb ## label for combined terms**

**mm=length(comb)**

**h=labels(corpOLD.dtm)$Terms**

**ind=h==comb[1]**

**for (i in 2:mm) {**

**ind1=h==comb[i]**

**ind=ind|ind1**

**}**

**corpR.dtm=corpOLD.dtm[,!ind]**

**dim(corpR.dtm)**

**labels(corpR.dtm)$Terms**

**as.matrix(corpR.dtm)**

**corpN.dtm=corpOLD.dtm[,ind]**

**dim(corpN.dtm)**

**if(dim(corpN.dtm)[2]>0) {**

**corpN.dtm[,1]=row\_sums(as.matrix(corpN.dtm))**

**corpN.dtm=corpN.dtm[,1]**

**dimnames(corpN.dtm)$Terms=groupcomb**

**}**

**corpNEW.dtm=as.DocumentTermMatrix(cbind(corpR.dtm,corpN.dtm),weighting=weightTf)**

**dim(corpNEW.dtm)**

**labels(corpNEW.dtm)$Terms**

**as.matrix(corpNEW.dtm)**