library(arules)

library(arulesViz)

library(devtools)

library(wordcloud2)

library(KoNLP)

library(wordcloud)

library(igraph)

library(RColorBrewer)

library(stringr)

library(dplyr)

useSejongDic()

#본인 파일 위치, 파일명 설정->data.frame으로 불러옴#

moms <- read.csv("C:/Rdata/액체세제/맘스홀릭\_(액체세제)\_599건.csv",as.is=T)

cook <- read.csv("C:/Rdata/액체세제/82cook액체세제2019-8-17 17시 6분.csv",as.is=T)

blog <- read.csv("C:/Rdata/액체세제/블로그 액체세제 173건.csv",as.is=T)

inst <- read.csv("C:/Rdata/액체세제/인스타\_액체세제\_500.csv",as.is=T)

#오류: 열의 개수가 열의 이름들보다 많습니다 -> 이거 뜨면 파일을 열어서 CSV (쉼표로 분리)로 저장

#각 파일의 "Words" 추출하여 하나의 변수 안에 넣음#

moms.df <- paste(moms$Contents, moms$Comments)

cook.df <- paste(cook$Contents, cook$Comments)

blog.df <- paste(blog$Contents, blog$Comments)

inst.df <- paste(inst$Contents, inst$Comments)

head(moms.df)

mstr(moms.df) # chr [1:599]

str(cook.df) # chr [1:23]

str(blog.df) # chr [1:174]

str(inst.df) # chr [1:501]

tail(moms.df)

names(moms.df)

#합치기

com <- c(moms.df, cook.df, blog.df, inst.df)

length(com)

str(com2)

class(com2) #matrix

#불필요한 문자 제거#

ed.dete<- gsub("[^ 가-힣]","", com)

head(ed.dete)

#세종 사전 단어 추가#->분리되서는 안되는 명사들

mergeUserDic(data.frame(c('섬유유연제','퍼실','미세먼지','냄새제거'), c('ncn')))

#명사 추출

noun.dete <- extractNoun(ed.dete)

head(noun.dete)

#단어 정제# ->삭제와 확인 반복하면서 데이터 정제

word.unlist<-unlist(noun.dete)

wordcount<-table(word.unlist)

head(sort(wordcount, decreasing=T),100)

word.unlist <- gsub('^세제$','',word.unlist)

word.unlist <- gsub('^사용$','',word.unlist)

word.unlist <- gsub('^한$','',word.unlist)

word.unlist <- gsub('^액체$','',word.unlist)

word.unlist <- gsub('^것$','',word.unlist)

word.unlist <- gsub('^수$','',word.unlist)

word.unlist <- gsub('^세탁$','',word.unlist)

word.unlist <- gsub('^빨래$','',word.unlist)

word.unlist <- gsub('^제품$','',word.unlist)

word.unlist <- gsub('^저$','',word.unlist)

word.unlist <- gsub('^하게$','',word.unlist)

word.unlist <- gsub('^아기$','',word.unlist)

word.unlist <- gsub('^세탁세제$','',word.unlist)

word.unlist <- gsub('^세탁기$','',word.unlist)

word.unlist <- gsub('^들$','',word.unlist)

word.unlist <- gsub('^번$','',word.unlist)

word.unlist <- gsub('^전$','',word.unlist)

word.unlist <- gsub('^물$','',word.unlist)

word.unlist <- gsub('^나$','',word.unlist)

word.unlist <- gsub('^후$','',word.unlist)

word.unlist <- gsub('^적$','',word.unlist)

word.unlist <- gsub('^일반$','',word.unlist)

word.unlist <- gsub('^생각$','',word.unlist)

word.unlist <- gsub('^제$','',word.unlist)

word.unlist <- gsub('^때문$','',word.unlist)

word.unlist <- gsub('^유아$','',word.unlist)

word.unlist <- gsub('^하기$','',word.unlist)

word.unlist <- gsub('^우리$','',word.unlist)

word.unlist <- gsub('^해$','',word.unlist)

word.unlist <- gsub('^해서$','',word.unlist)

word.unlist <- gsub('^적$','',word.unlist)

word.unlist <- gsub('^나$','',word.unlist)

word.unlist <- gsub('^분$','',word.unlist)

word.unlist <- gsub('^우리$','',word.unlist)

word.unlist <- gsub('^하기$','',word.unlist)

word.unlist <- gsub('^해$','',word.unlist)

word.unlist <- gsub('^때문$','',word.unlist)

word.unlist <- gsub('^아이$','',word.unlist)

word.unlist <- gsub('^등$','',word.unlist)

word.unlist <- gsub('^형$','',word.unlist)

word.unlist <- gsub('^시트$','',word.unlist)

word.unlist <- gsub('^카페$','',word.unlist)

word.unlist <- gsub('^게시$','',word.unlist)

word.unlist <- gsub('^일$','',word.unlist)

word.unlist <- gsub('^집$','',word.unlist)

word.unlist <- gsub('^데$','',word.unlist)

word.unlist <- gsub('^가능$','',word.unlist)

word.unlist <- gsub('^미세$','',word.unlist)

word.unlist <- gsub('^공지$','',word.unlist)

word.unlist <- gsub('^빨래$','',word.unlist)

word.unlist <- gsub('^규정$','',word.unlist)

word.unlist <- gsub('^운영$','',word.unlist)

word.unlist <- gsub('^필독$','',word.unlist)

word.unlist <- gsub('^판별$','',word.unlist)

word.unlist <- gsub('^게시글작성$','',word.unlist)

word.unlist <- gsub('^시$','',word.unlist)

word.unlist <- gsub('^하나$','',word.unlist)

word.unlist <- gsub('^정도$','',word.unlist)

word.unlist <- gsub('^주방$','',word.unlist)

word.unlist <- gsub('^가지$','',word.unlist)

word.unlist <- gsub('^두$','',word.unlist)

word.unlist <- gsub('^유연$','',word.unlist)

word.unlist <- gsub('^아이들$','',word.unlist)

word.unlist <- gsub('^속$','',word.unlist)

word.unlist <- gsub('^선물$','',word.unlist)

word.unlist <- gsub('^이$','',word.unlist)

word.unlist <- gsub('^개$','',word.unlist)

word.unlist <- gsub('^원$','',word.unlist)

word.unlist <- gsub('^중$','',word.unlist)

word.unlist <- gsub('^이벤트$','',word.unlist)

------------------------------

word.unlist <- gsub('^옷$','',word.unlist)

word.unlist <- gsub('^추천$','',word.unlist)

word.unlist <- gsub('^양$','',word.unlist)

word.unlist <- gsub('^르주르$','',word.unlist)

word.unlist <- gsub('^가루$','',word.unlist)

word.unlist <- gsub('^섬유$','',word.unlist)

word.unlist <- gsub('^안전$','',word.unlist)

word.unlist <- gsub('^섬유유연제$','',word.unlist)

word.unlist <- gsub('^천연$','',word.unlist)

word.unlist <- gsub('^용$','',word.unlist)

word.unlist <- gsub('^듯$','',word.unlist)

word.unlist <- gsub('^지$','',word.unlist)

word.unlist <- word.unlist[nchar(word.unlist)>0]

wordcount<-table(word.unlist)

head(sort(wordcount, decreasing=T),80)

#wordcloud 만들기#

wordcount<-head(sort(wordcount, decreasing=T),80)

palete<-brewer.pal(5,"Set1")

windowsFonts(word\_font=windowsFont("맑은 고딕"))

wordcloud2(data=wordcount, size=1, col="random-dark", rotateRatio = 0)

#상위 80건 txt로 저장(나중에 워드 클라우드 사이트에서 사용하기 위해서 저장)

wordcount80<-head(sort(wordcount,decreasing=T),80)

names80<-names(wordcount80)

write(names80,"C:/Rdata/액체세제 wc80.txt")