################################################################

Preprocessing of Problems

##############################################################

install.packages("tm") # for text mining

install.package("SnowballC") # for text stemming

install.packages("wordcloud") # word-cloud generator

install.packages("RColorBrewer") # color palettes

library("tm")

library("SnowballC")

library("wordcloud")

library("RColorBrewer")

setwd("D:/Experimentjournal/problems/Security-patterns-shumacher/Problems")

source<- DirSource("D:/Experimentjournal/problems/Problem")

YourCorpus<- Corpus(source, readerControl=list(reader=readPlain))

docs<- Corpus(VectorSource(YourCorpus))

toSpace<- content\_transformer(function (x , pattern )gsub(pattern, " ", x))

docs<- tm\_map(docs, toSpace, "/")

docs<- tm\_map(docs, toSpace, "@")

docs<- tm\_map(docs,toSpace, "\\|")

docs<- tm\_map(docs, content\_transformer(tolower))

docs<- tm\_map(docs, removeNumbers)

docs<- tm\_map(docs, removeWords, stopwords("english"))

docs<- tm\_map(docs, removeWords, c("blabla1", "blabla2"))

docs<- tm\_map(docs, removePunctuation)

docs<- tm\_map(docs, stripWhitespace)

docs<- tm\_map(docs, stemDocument)

dtm<- TermDocumentMatrix(docs)

m <- as.matrix(dtm)

v <- sort(rowSums(m),decreasing=TRUE)

d <- data.frame(word = names(v),freq=v)

write.csv(d, "SACA-problem49.csv")