#Gene collection procedure using r

library(rentrez)

entrez\_dbs()

entrez\_db\_searchable("gene")

r\_search <- entrez\_search(db = "gene", term = "Heart[ALL] & homo sapiens[Human]", retmax=200)

r\_search

r\_search$ids

r\_seqs <- entrez\_fetch(db="gene", id= r\_search$ids, rettype="txt", retmode = "text")

r\_seqs

#Gene filtering procedure using r

r\_search <- entrez\_search(db="gene", term = "Heart[ALL] & homo sapiens[Human]", retmax = 200)

r\_search

#Gene sorting procedure using r

r\_seqs1 <- write(r\_seqs, file = "t.txt")

#Gene linkage procedure using r

r\_search <- entrez\_search(db="gene", term = "Heart[ALL] & schizophrenia[ALL]", retmax = 200)

r\_search

r\_search$ids

r\_seqs <- entrez\_fetch(db="gene", id=r\_search$ids, rettype = "txt", retmode = "text")

#Gene mining procedure using r

install.packages("tm")

library(tm)

text <- readLines("t.txt")

docs <- Corpus(VectorSource(text))

inspect(docs)

docs <- docs[-(1:3)]

docs <- tm\_map(docs, removeWords, c("dd", "dt", "class", "dl", "desig","details"))

docs <- tm\_map(docs, removePunctuation)

df <- head(docs, 100)

inspect(df)

df <- read.table("t.txt", fill = TRUE, header = FALSE)

df

df <- df[,-(3:227)]