Fibromyalgia PubMed

Janis Corona

12/9/2019

## This script takes ten articles from the abstracts on earache articles from NCBI’s PubMed

This creates a directory to stem the abstracts and preprocess from the csv file into a corpus of 20 files in a folder called Earache.

Auto <- read.csv('Fibromyalgia\_PubMed\_Abstracts.csv', sep=',',  
 header=TRUE, na.strings=c('',' '))  
  
auto <- Auto[complete.cases(Auto$abstract),]  
  
  
dir.create('./Fibromyalgias')  
  
ea <- as.character(auto$abstract)  
setwd('./Fibromyalgias')  
  
for (j in 1:length(ea)){  
 write(ea[j], paste(paste('EA',j, sep='.'), '.txt', sep=''))  
}  
setwd('../')

This code preprocesses and stems the corpus

library(tm)  
library(SnowballC)  
library(wordcloud)  
library(ggplot2)  
  
Fibromyalgias <- Corpus(DirSource("Fibromyalgias"))  
  
  
Fibromyalgias

## <<SimpleCorpus>>  
## Metadata: corpus specific: 1, document level (indexed): 0  
## Content: documents: 20

#Fibromyalgias <- tm\_map(Fibromyalgias, removePunctuation)  
#Fibromyalgias <- tm\_map(Fibromyalgias, removeNumbers)  
Fibromyalgias <- tm\_map(Fibromyalgias, tolower)  
Fibromyalgias <- tm\_map(Fibromyalgias, removeWords, stopwords("english"))  
Fibromyalgias <- tm\_map(Fibromyalgias, stripWhitespace)  
Fibromyalgias <- tm\_map(Fibromyalgias, stemDocument)  
  
dtmFibromyalgias <- DocumentTermMatrix(Fibromyalgias)  
  
freq <- colSums(as.matrix(dtmFibromyalgias))

This code orders words stemmed by frequency and finds input correlations

FREQ <- data.frame(freq)  
ord <- order(freq, decreasing=TRUE)  
  
freq[head(ord, 25)]

## pain patient fibromyalgia sleep studi   
## 60 54 51 36 28   
## group treatment use qualiti effect   
## 22 22 22 22 19   
## associ assess signific clinic differ   
## 19 18 18 17 17   
## week chronic symptom fms score   
## 15 15 15 14 14   
## measur mind includ diseas higher   
## 12 12 11 11 11

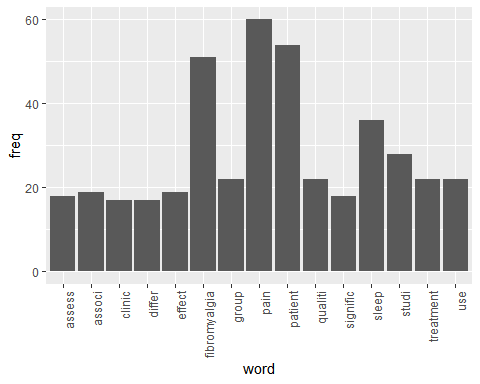
findAssocs(dtmFibromyalgias, "sleep", corlimit=0.7)

## $sleep  
## disturb anxieti mind (93%   
## 0.82 0.77 0.76 0.74   
## (five (hospit (pittsburgh (promi   
## 0.74 0.74 0.74 0.74   
## (psqi -.24, -.31, -.58,   
## 0.74 0.74 0.74 0.74   
## -0.23, -0.54, .0001), .0001).   
## 0.74 0.74 0.74 0.74   
## .001), .002) .002; 177   
## 0.74 0.74 0.74 0.74   
## 59% [promis-sd]), [psqi], aim:   
## 0.74 0.74 0.74 0.74   
## anxiety; asia australia, cultiv   
## 0.74 0.74 0.74 0.74   
## depression; disturbance. female; interfer   
## 0.74 0.74 0.74 0.74   
## interference), interference, john kg/m2   
## 0.74 0.74 0.74 0.74   
## leagu longitudin mediat mediation;   
## 0.74 0.74 0.74 0.74   
## method: pacif patent pearson   
## 0.74 0.74 0.74 0.74   
## peopl potenti previous problems,   
## 0.74 0.74 0.74 0.74   
## promi promis-sd questionnaire), relationship   
## 0.74 0.74 0.74 0.74   
## relationship. scale). sleep. son   
## 0.74 0.74 0.74 0.74   
## suffer unstudied. warrant white).   
## 0.74 0.74 0.74 0.74   
## wiley years; patients.   
## 0.74 0.74 0.73

findAssocs(dtmFibromyalgias, "pain", corlimit=0.5)

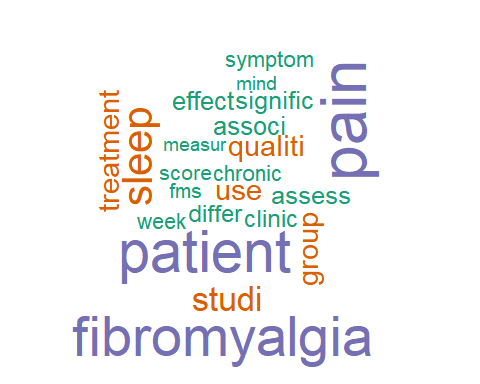
## $pain  
## (especi (myofasci   
## 0.81 0.81   
## .e., areas,   
## 0.81 0.81   
## attack bowel   
## 0.81 0.81   
## cervico-faci co-occur,   
## 0.81 0.81   
## co-occurr co-occurrence;   
## 0.81 0.81   
## concurr condition.   
## 0.81 0.81   
## districts. dysmenorrhea/endometriosi   
## 0.81 0.81   
## express give   
## 0.81 0.81   
## headach headache.   
## 0.81 0.81   
## headache; heart   
## 0.81 0.81   
## hyperalgesia, influenc   
## 0.81 0.81   
## input irrit   
## 0.81 0.81   
## ischem least   
## 0.81 0.81   
## locat messag   
## 0.81 0.81   
## migrain migraine;   
## 0.81 0.81   
## multifactorial; mutual   
## 0.81 0.81   
## myofasci organ   
## 0.81 0.81   
## pain) pain),   
## 0.81 0.81   
## paper part   
## 0.81 0.81   
## particular, pathophysiolog   
## 0.81 0.81   
## perpetu phenomena   
## 0.81 0.81   
## points, probabl   
## 0.81 0.81   
## process project   
## 0.81 0.81   
## refer rise   
## 0.81 0.81   
## sensori share   
## 0.81 0.81   
## strong suffering.   
## 0.81 0.81   
## summar syndrome,   
## 0.81 0.81   
## tend tension-typ   
## 0.81 0.81   
## trigger typic   
## 0.81 0.81   
## urinari various   
## 0.81 0.81   
## viscer viscero-viscer   
## 0.81 0.81   
## ways. intern   
## 0.81 0.74   
## central condit   
## 0.72 0.72   
## one interact   
## 0.72 0.70   
## modul can   
## 0.65 0.63   
## musculoskelet symptom   
## 0.61 0.59   
## enhanc hyperalgesia   
## 0.59 0.59   
## nocicept observed.   
## 0.59 0.59   
## under among   
## 0.59 0.53   
## will   
## 0.51

wf <- data.frame(word=names(freq), freq=freq)  
p <- ggplot(subset(wf, freq>16), aes(word, freq))  
p <- p + geom\_bar(stat= 'identity')   
p <- p + theme(axis.text.x=element\_text(angle=90, hjust=1))   
p



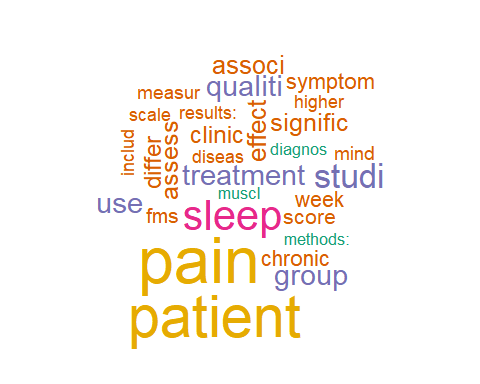
wordcloud(names(freq), freq, min.freq=12,colors=brewer.pal(1,'Dark2'))

## Warning in brewer.pal(1, "Dark2"): minimal value for n is 3, returning requested palette with 3 different levels



wordcloud(names(freq), freq, max.words=30,colors=brewer.pal(6,'Dark2'))

## Warning in wordcloud(names(freq), freq, max.words = 30, colors =  
## brewer.pal(6, : fibromyalgia could not be fit on page. It will not be  
## plotted.



### The above stemmed the corpus, this will lemmatize the original csv file

and add the field to the table and write out to csv, followed by plot the word count frequencies that were lemmatized and the word clouds

library(textstem)  
  
lemma <- lemmatize\_strings(auto$abstract, dictionary=lexicon::hash\_lemmas)  
  
Lemma <- as.data.frame(lemma)  
Lemma <- cbind(Lemma, auto)  
  
colnames(Lemma) <- c('lemmatizedAbstract','abstract', 'source')  
  
write.csv(Lemma, 'LemmatizedFibromyalgias.csv', row.names=FALSE)

dir.create('./Fibromyalgias-Lemma')  
  
ea <- as.character(Lemma$lemmatizedAbstract)  
setwd('./Fibromyalgias-Lemma')  
  
for (j in 1:length(ea)){  
 write(ea[j], paste(paste('EAL',j, sep='.'), '.txt', sep=''))  
}  
setwd('../')

library(tm)  
library(SnowballC)  
library(wordcloud)  
library(ggplot2)

Fibromyalgias <- Corpus(DirSource("Fibromyalgias-Lemma"))  
  
Fibromyalgias

## <<SimpleCorpus>>  
## Metadata: corpus specific: 1, document level (indexed): 0  
## Content: documents: 20

#Fibromyalgias <- tm\_map(Fibromyalgias, removePunctuation)  
#Fibromyalgias <- tm\_map(Fibromyalgias, removeNumbers)  
Fibromyalgias <- tm\_map(Fibromyalgias, tolower)  
Fibromyalgias <- tm\_map(Fibromyalgias, removeWords, stopwords("english"))  
Fibromyalgias <- tm\_map(Fibromyalgias, stripWhitespace)  
  
dtmFibromyalgias <- DocumentTermMatrix(Fibromyalgias)  
dtmFibromyalgias

## <<DocumentTermMatrix (documents: 20, terms: 1344)>>  
## Non-/sparse entries: 2242/24638  
## Sparsity : 92%  
## Maximal term length: 21  
## Weighting : term frequency (tf)

freq <- colSums(as.matrix(dtmFibromyalgias))  
  
FREQ <- data.frame(freq)  
ord <- order(freq, decreasing=TRUE)  
  
freq[head(ord, 25)]

## pain patient fibromyalgia sleep study   
## 60 55 52 36 29   
## group treatment quality use score   
## 23 23 23 22 18   
## fms high symptom much analysis   
## 17 17 16 15 14   
## clinical effect good chronic mdhaq   
## 14 14 14 14 13   
## scale significant week report mindfulness   
## 13 12 12 12 12

patient <- as.data.frame(findAssocs(dtmFibromyalgias, "patient", corlimit=0.70))  
  
pain <- as.data.frame(findAssocs(dtmFibromyalgias, "pain", corlimit=0.75))  
  
  
treatment <- as.data.frame(findAssocs(dtmFibromyalgias, "treatment", corlimit=0.55))  
  
patient

## patient  
## key 0.77  
## presence 0.74

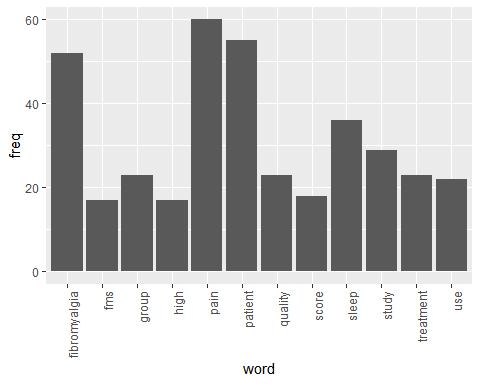
pain

## pain  
## area, 0.85  
## attack 0.85  
## bowel 0.85  
## cervico 0.85  
## concurrent 0.85  
## district. 0.85  
## dysmenorrhea 0.85  
## e., 0.85  
## endometriosis 0.85  
## enhancement 0.85  
## especially 0.85  
## expression 0.85  
## facial 0.85  
## headache 0.85  
## headache. 0.85  
## headache; 0.85  
## heart 0.85  
## hyperalgesia, 0.85  
## influence 0.85  
## input 0.85  
## internal 0.85  
## irritable 0.85  
## ischemic 0.85  
## less 0.85  
## locate 0.85  
## message 0.85  
## migraine 0.85  
## migraine; 0.85  
## multifactorial; 0.85  
## mutual 0.85  
## myofascial 0.85  
## occur, 0.85  
## occurrence 0.85  
## occurrence; 0.85  
## organ 0.85  
## paper 0.85  
## part 0.85  
## particular, 0.85  
## pathophysiology 0.85  
## perpetuate 0.85  
## phenomenon 0.85  
## point, 0.85  
## probably 0.85  
## process 0.85  
## projection 0.85  
## refer 0.85  
## rise 0.85  
## sensory 0.85  
## share 0.85  
## strong 0.85  
## suffer. 0.85  
## summarize 0.85  
## tend 0.85  
## tension 0.85  
## thorough 0.85  
## trigger 0.85  
## typical 0.85  
## urinary 0.85  
## various 0.85  
## visceral 0.85  
## viscero 0.85  
## way. 0.85  
## modulation 0.79  
## interaction 0.75

treatment

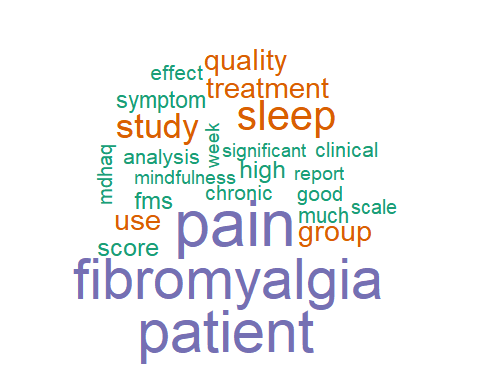
## treatment  
## eligible 0.84  
## outpatient 0.76  
## 1st 0.73  
## 60, 0.73  
## acr 0.73  
## back 0.73  
## balneological 0.73  
## balneotherapy; 0.73  
## blind, 0.73  
## blind. 0.73  
## delivery 0.73  
## except 0.73  
## fiq 0.73  
## fms. 0.73  
## full 0.73  
## hydrotherapy 0.73  
## hydrotherapy; 0.73  
## immersion 0.73  
## intensity 0.73  
## intention 0.73  
## intermittent 0.73  
## judgment 0.73  
## mcid 0.73  
## method. 0.73  
## min 0.73  
## month. 0.73  
## mud 0.73  
## pack 0.73  
## parallel 0.73  
## peloidotherapy 0.73  
## region 0.73  
## seem 0.73  
## statistical 0.73  
## tap 0.73  
## time 0.73  
## water 0.73  
## øc 0.73  
## øc. 0.73  
## consecutive 0.69  
## 2010 0.62  
## application 0.62  
## assign 0.62  
## clinic. 0.62  
## completion 0.62  
## difference 0.62  
## important 0.62  
## randomly 0.62  
## group. 0.58

wf <- data.frame(word=names(freq), freq=freq)  
p <- ggplot(subset(wf, freq>16), aes(word, freq))  
p <- p + geom\_bar(stat= 'identity')   
p <- p + theme(axis.text.x=element\_text(angle=90, hjust=1))   
p



wordcloud(names(freq), freq, min.freq=12,colors=brewer.pal(2,'Dark2'))

## Warning in brewer.pal(2, "Dark2"): minimal value for n is 3, returning requested palette with 3 different levels



wordcloud(names(freq), freq, max.words=28,colors=brewer.pal(1,'Dark2'))

## Warning in brewer.pal(1, "Dark2"): minimal value for n is 3, returning requested palette with 3 different levels

## Warning in wordcloud(names(freq), freq, max.words = 28, colors =  
## brewer.pal(1, : fibromyalgia could not be fit on page. It will not be  
## plotted.

