Sciatica 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('sciatic\_PubMed\_Abstracts.csv', sep=',',  
 header=TRUE, na.strings=c('',' '))  
  
auto <- Auto[complete.cases(Auto$abstract),]  
  
  
dir.create('./Sciatica')  
  
ea <- as.character(auto$abstract)  
setwd('./Sciatica')  
  
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)  
  
Sciatica <- Corpus(DirSource("Sciatica"))  
  
  
Sciatica

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

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

This code orders words stemmed by frequency and finds input correlations

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

## nerv sciatic patient pain studi muscl control   
## 68 59 56 41 29 22 20   
## postop piriformi clinic signific associ block measur   
## 18 17 16 16 16 16 14   
## sciatica group show produc effect assess am1710   
## 13 13 13 12 12 12 12   
## imag outcom receiv result   
## 11 11 11 10

findAssocs(dtmSciatica, "effect", corlimit=0.7)

## $effect  
## manag hour (cg).   
## 0.85 0.78 0.76   
## (f?=?16.26; (partial (pisp)   
## 0.76 0.76 0.76   
## (stens) (tens) (tg)   
## 0.76 0.76 0.76   
## 10?week 10th 19%   
## 0.76 0.76 0.76   
## 2018. 29th 2nd,   
## 0.76 0.76 0.76   
## 4th, 6th, 8th   
## 0.76 0.76 0.76   
## account administr africa   
## 0.76 0.76 0.76   
## allot analysi anova   
## 0.76 0.76 0.76   
## appli azikiw blind   
## 0.76 0.76 0.76   
## carri come comparison   
## 0.76 0.76 0.76   
## consent conveni counterpart   
## 0.76 0.76 0.76   
## current determin electr   
## 0.76 0.76 0.76   
## enrol eta hospital,   
## 0.76 0.76 0.76   
## howev inform intervent   
## 0.76 0.76 0.76   
## interventions. intramuscular landmark   
## 0.76 0.76 0.76   
## lasted. mani may,   
## 0.76 0.76 0.76   
## mean?ñ?sd, medicals/surg nnamdi   
## 0.76 0.76 0.76   
## nnewi nnewi. non-random   
## 0.76 0.76 0.76   
## none, obtain p?<?0.05.   
## 0.76 0.76 0.76   
## p?=?0.01); pactr201805003408271 pair   
## 0.76 0.76 0.76   
## pan physiotherapi physiotherapy,   
## 0.76 0.76 0.76   
## pisp pisp. post-inject   
## 0.76 0.76 0.76   
## prove recommend recruit   
## 0.76 0.76 0.76   
## registri reliev repeat   
## 0.76 0.76 0.76   
## respectively, result: services,   
## 0.76 0.76 0.76   
## session set sham   
## 0.76 0.76 0.76   
## squared)?=?0.19. sten stimul   
## 0.76 0.76 0.76   
## t-test. tens; test/experiment   
## 0.76 0.76 0.76   
## transcutan undertaken varianc   
## 0.76 0.76 0.76   
## variance. week, weeks)   
## 0.76 0.76 0.76   
## well wise written   
## 0.76 0.76 0.76   
## wrong ten particip   
## 0.76 0.74 0.73   
## test   
## 0.72

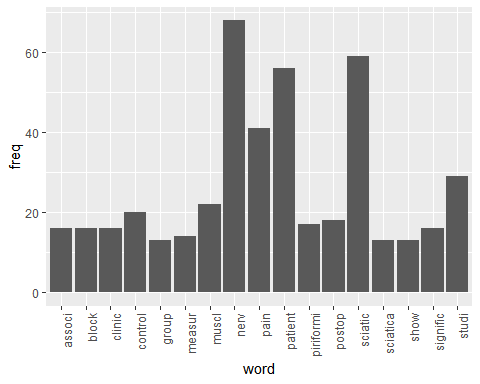
findAssocs(dtmSciatica, "assess", corlimit=0.5)

## $assess  
## intern less (32.9 (g.   
## 0.84 0.82 0.76 0.76   
## (ifis), (odi) (sf-36), /s.   
## 0.76 0.76 0.76 0.76   
## 0.01) 0.05). 159 16th   
## 0.76 0.76 0.76 0.76   
## 34th 36-item 4.7 agility;   
## 0.76 0.76 0.76 0.76   
## bodili bodily, cardiorespiratori compon   
## 0.76 0.76 0.76 0.76   
## course. disabl employ fit   
## 0.76 0.76 0.76 0.76   
## fitness. fitness; flexibility; flexibl   
## 0.76 0.76 0.76 0.76   
## gestat gestation; greater index   
## 0.76 0.76 0.76 0.76   
## john low-moder medium muscular   
## 0.76 0.76 0.76 0.76   
## old). oswestri overal pregnanc   
## 0.76 0.76 0.76 0.76   
## pregnancy. pregnant questionnair relev   
## 0.76 0.76 0.76 0.76   
## scale; self-report son speed-agil   
## 0.76 0.76 0.76 0.76   
## stage. strength strength; w.).   
## 0.76 0.76 0.76 0.76   
## w., women year analogu   
## 0.76 0.76 0.76 0.73   
## explor physic associ pain,   
## 0.73 0.72 0.71 0.69   
## pain wiley scale compris   
## 0.67 0.64 0.63 0.62   
## week health survey score   
## 0.62 0.62 0.62 0.59   
## ltd. need higher lumbar   
## 0.58 0.58 0.55 0.53

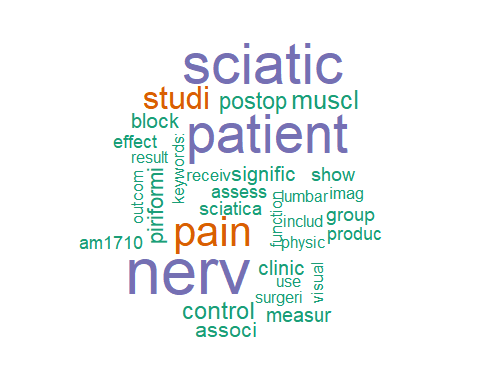
findAssocs(dtmSciatica, "piriformi", corlimit=0.5)

## $piriformi  
## muscle. twenti "nerv (5%), (70%),   
## 0.95 0.88 0.87 0.87 0.87   
## (85%), 32:282-286, 80% abandoned. anat.   
## 0.87 0.87 0.87 0.87 0.87   
## anatomi anatomy; anterior approach arthroplasty,   
## 0.87 0.87 0.87 0.87 0.87   
## cadav cadaver; clarifi clin. disorders,   
## 0.87 0.87 0.87 0.87 0.87   
## dissect exist fresh-frozen identified, identified.   
## 0.87 0.87 0.87 0.87 0.87   
## implic inc. inferior innerv joint;   
## 0.87 0.87 0.87 0.87 0.87   
## nerves. origin. periodicals, piriformis" plexus,   
## 0.87 0.87 0.87 0.87 0.87   
## plexus; posterior rami ramus s2.   
## 0.87 0.87 0.87 0.87 0.87   
## sacroiliac said sides, sides. singl   
## 0.87 0.87 0.87 0.87 0.87   
## sourc specimen study, suppli surround   
## 0.87 0.87 0.87 0.87 0.87   
## term vague. ventral side 2018   
## 0.87 0.87 0.87 0.85 0.80   
## syndrome; sacral superior one branch   
## 0.80 0.75 0.75 0.73 0.67   
## gluteal common iatrogen studied. origin   
## 0.65 0.64 0.64 0.64 0.62   
## thus nerve, descript specifically, trace   
## 0.57 0.57 0.57 0.57 0.57   
## proxim respectively. therefore, literatur   
## 0.57 0.57 0.57 0.56

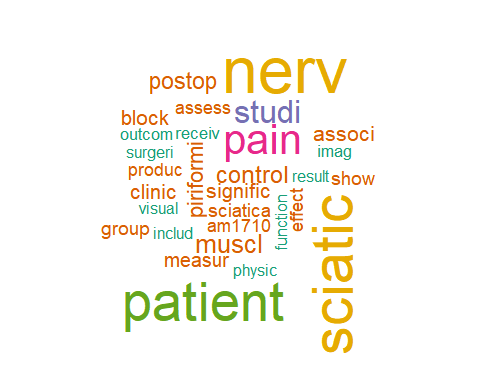
wf <- data.frame(word=names(freq), freq=freq)  
p <- ggplot(subset(wf, freq>12), aes(word, freq))  
p <- p + geom\_bar(stat= 'identity')   
p <- p + theme(axis.text.x=element\_text(angle=90, hjust=1))   
p



wordcloud(names(freq), freq, min.freq=10,colors=brewer.pal(3,'Dark2'))



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



### 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, 'LemmatizedSciatica.csv', row.names=FALSE)

dir.create('./Sciatica-Lemma')  
  
ea <- as.character(Lemma$lemmatizedAbstract)  
setwd('./Sciatica-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)

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

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

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

## <<DocumentTermMatrix (documents: 20, terms: 1425)>>  
## Non-/sparse entries: 2238/26262  
## Sparsity : 92%  
## Maximal term length: 21  
## Weighting : term frequency (tf)

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

## nerve sciatic patient pain study   
## 69 60 55 41 29   
## muscle control piriformis postoperative ca8   
## 22 21 18 17 17   
## group low block clinical sciatica   
## 16 16 16 14 14   
## show can much produce ultrasound   
## 13 12 12 12 12   
## tka associate am1710 conclusion: include   
## 12 12 12 11 11

piriformis <- as.data.frame(findAssocs(dtmSciatica, "piriformis", corlimit=0.70))  
  
ultrasound <- as.data.frame(findAssocs(dtmSciatica, "ultrasound", corlimit=0.5))  
  
  
am1710 <- as.data.frame(findAssocs(dtmSciatica, "am1710", corlimit=0.55))  
  
ca8 <- as.data.frame(findAssocs(dtmSciatica, "ca8", corlimit=0.85))  
  
piriformis

## piriformis  
## 2019.? 0.89  
## 282 0.89  
## 286, 0.89  
## 32: 0.89  
## abandon. 0.89  
## anat. 0.89  
## anatomy 0.89  
## anatomy; 0.89  
## anterior 0.89  
## approach 0.89  
## arthroplasty, 0.89  
## cadaver 0.89  
## cadaver; 0.89  
## clarify 0.89  
## clin. 0.89  
## description 0.89  
## disorder, 0.89  
## dissection 0.89  
## exist 0.89  
## freeze 0.89  
## fresh 0.89  
## identify, 0.89  
## identify. 0.89  
## implicate 0.89  
## inc. 0.89  
## inferior 0.89  
## innervate 0.89  
## innervation 0.89  
## joint; 0.89  
## majority 0.89  
## origin. 0.89  
## periodical, 0.89  
## plexus, 0.89  
## plexus; 0.89  
## posterior 0.89  
## proximally 0.89  
## rami 0.89  
## ramus 0.89  
## s2. 0.89  
## sacroiliac 0.89  
## say 0.89  
## side, 0.89  
## side. 0.89  
## source 0.89  
## specimen 0.89  
## study, 0.89  
## supply 0.89  
## surround 0.89  
## term 0.89  
## vague. 0.89  
## ventral 0.89  
## twenty 0.87  
## side 0.85  
## 2018 0.80  
## origin 0.80  
## sacral 0.77  
## superior 0.77  
## muscle. 0.71

ultrasound

## ultrasound  
## femoris, 0.72  
## medial 0.72  
## soleus 0.72  
## ultrasonography 0.72  
## july 0.72  
## contraction 0.70  
## area 0.68  
## bicep 0.68  
## change 0.68  
## low 0.68  
## ratio 0.65  
## 2017. 0.65  
## cross 0.64  
## sectional 0.64  
## evaluate 0.63  
## intensity 0.58  
## muscle 0.57  
## femoris 0.57  
## injection, 0.57  
## background: 0.55  
## copyright? 0.53  
## elsevier 0.53  
## pain; 0.53  
## structure 0.53  
## guide 0.53  
## image 0.52  
## measure 0.50  
## maximum 0.50

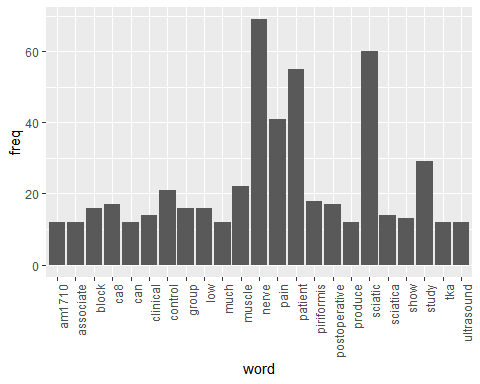
am1710

## am1710  
## activation 1.00  
## adjuvant 1.00  
## agent 1.00  
## agonist, 1.00  
## allodynia 1.00  
## antagonist 1.00  
## antiallodynic 1.00  
## antinociceptive 1.00  
## appropriate 1.00  
## attenuate 1.00  
## behave 1.00  
## benzo 1.00  
## broad 1.00  
## camp 1.00  
## cannabilactone 1.00  
## cannabinoid 1.00  
## cb1 1.00  
## cb2 1.00  
## cfa 1.00  
## characterize. 1.00  
## chromen 1.00  
## contrast, 1.00  
## day, 1.00  
## day? 1.00  
## delay 1.00  
## dependence. 1.00  
## development 1.00  
## dimethyl 1.00  
## dose 1.00  
## efficacy 1.00  
## embryonic 1.00  
## endure 1.00  
## extracellular 1.00  
## forskolin 1.00  
## freund's 1.00  
## gabapentin 1.00  
## hcb2. 1.00  
## hek 1.00  
## heptyl 1.00  
## hydroxy 1.00  
## incompletely 1.00  
## indication 1.00  
## inhibition 1.00  
## kg, 1.00  
## kidney 1.00  
## kinase 1.00  
## ligation 1.00  
## mcb2 1.00  
## mcb2. 1.00  
## mechanical 1.00  
## mediate 1.00  
## methoxy 1.00  
## model. 1.00  
## modest 1.00  
## mouse 1.00  
## mouse, 1.00  
## mouse. 1.00  
## paclitaxel 1.00  
## pathway 1.00  
## phosphorylation 1.00  
## precipitate 1.00  
## production 1.00  
## profile 1.00  
## prophylactic 1.00  
## psnl 1.00  
## receptor; 1.00  
## render 1.00  
## rodent 1.00  
## similarly, 1.00  
## species 1.00  
## spectrum 1.00  
## stably 1.00  
## stimulate 1.00  
## suppress 1.00  
## sustain 1.00  
## tetrahydrocannabinol, 1.00  
## tolerance 1.00  
## tolerant 1.00  
## translation. 1.00  
## underlie 1.00  
## vitro, 1.00  
## vivo, 1.00  
## vivo. 1.00  
## withdrawal 1.00  
## signal 0.94  
## morphine 0.89  
## receptor 0.89  
## identify 0.81  
## induce 0.79  
## suggest 0.70  
## unwanted 0.69  
## know 0.69  
## chemotherapy 0.69  
## neuropathic 0.69  
## partial 0.69  
## establish 0.69  
## express 0.69  
## human 0.69  
## inhibit 0.69  
## observe 0.69  
## therapeutic 0.69  
## whether 0.69

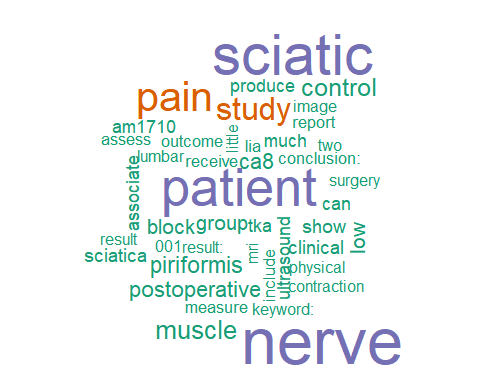
ca8

## ca8  
## 1,697bp 1.00  
## 201 1.00  
## 204 1.00  
## 204. 1.00  
## 204c 1.00  
## 204g 1.00  
## 3'utr 1.00  
## aav8 1.00  
## able 1.00  
## advance 1.00  
## advillin 1.00  
## allele 1.00  
## allosteric 1.00  
## also 1.00  
## anhydrase 1.00  
## ataxia, 1.00  
## barely 1.00  
## ca++ 1.00  
## calcium 1.00  
## carbonic 1.00  
## cell, 1.00  
## cell. 1.00  
## cerebellar 1.00  
## cis 1.00  
## construct 1.00  
## critical 1.00  
## cryptic 1.00  
## deliver 1.00  
## derive 1.00  
## develop 1.00  
## dorsal 1.00  
## drg 1.00  
## eqtl 1.00  
## eqtl. 1.00  
## evade 1.00  
## excitability 1.00  
## excitability. 1.00  
## exclusively 1.00  
## exon 1.00  
## explain 1.00  
## expression 1.00  
## extent 1.00  
## flag 1.00  
## g., 1.00  
## ganglion 1.00  
## gene 1.00  
## genomic 1.00  
## glial 1.00  
## greatly 1.00  
## hek293 1.00  
## homozygote 1.00  
## homozygous 1.00  
## hyperalgesic 1.00  
## ihc 1.00  
## immunohistochemistry 1.00  
## impact 1.00  
## imply 1.00  
## inhibitor 1.00  
## inositol 1.00  
## intracellular 1.00  
## itpr1 1.00  
## lesser 1.00  
## murine 1.00  
## mutation 1.00  
## naturally 1.00  
## nbl 1.00  
## neural 1.00  
## neuronal 1.00  
## nociception 1.00  
## null 1.00  
## occur 1.00  
## pathway, 1.00  
## peptide 1.00  
## phenotype. 1.00  
## pitpr1 1.00  
## population, 1.00  
## predominantly 1.00  
## previously 1.00  
## question, 1.00  
## release, 1.00  
## report, 1.00  
## restrict 1.00  
## reversion 1.00  
## rs6471859 1.00  
## rs6471859, 1.00  
## splice 1.00  
## splice, 1.00  
## stable 1.00  
## synaptic 1.00  
## tissue 1.00  
## transcript 1.00  
## transduce 1.00  
## transfer 1.00  
## trisphosphate 1.00  
## truncate 1.00  
## viral 1.00  
## vitro. 1.00  
## vivo 1.00  
## function 0.96  
## regulate 0.95  
## response 0.89  
## cell 0.87  
## produce 0.86

wf <- data.frame(word=names(freq), freq=freq)  
p <- ggplot(subset(wf, freq>11), 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=10,colors=brewer.pal(3,'Dark2'))



wordcloud(names(freq), freq, max.words=40,colors=brewer.pal(6,'Dark2'))

