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

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

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

This code orders words stemmed by frequency and finds input correlations

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

## morton neuroma patient nerv pain   
## 39 37 35 33 31   
## clinic forefoot inject keywords: neuroma;   
## 29 25 25 19 18   
## month foot treatment intermetatars studi   
## 18 17 17 17 14   
## use neuroma. plantar signific assess   
## 14 14 13 13 12   
## ankl diagnosi differ level results:   
## 12 12 11 11 11

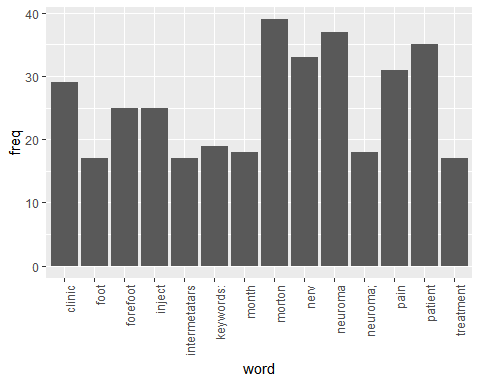
findAssocs(dtmneuromas, "plantar", corlimit=0.5)

## $plantar  
## may allow anatom foot;   
## 0.84 0.80 0.80 0.80   
## digit condit (34 (mppdn)   
## 0.78 0.75 0.72 0.72   
## (mppdn). 0.4-1.4). 0.8 1-8)   
## 0.72 0.72 0.72 0.72   
## 19-27) 2-9). adjac adult   
## 0.72 0.72 0.72 0.72   
## analysed. anatomy; applications. boni   
## 0.72 0.72 0.72 0.72   
## border branch clear consensus.   
## 0.72 0.72 0.72 0.72   
## consist cours course. cross-sect   
## 0.72 0.72 0.72 0.72   
## cuneiform. defined. depict detect   
## 0.72 0.72 0.72 0.72   
## direct distally, entir first   
## 0.72 0.72 0.72 0.72   
## hallux. hallux; healthi high-resolut   
## 0.72 0.72 0.72 0.72   
## joint, joplin known landmark   
## 0.72 0.72 0.72 0.72   
## long map materi medial   
## 0.72 0.72 0.72 0.72   
## metatarsophalang mm2 mppdn musculoskelet   
## 0.72 0.72 0.72 0.72   
## nerves) position, precis proper   
## 0.72 0.72 0.72 0.72   
## radiologist respect rise sesamoid   
## 0.72 0.72 0.72 0.72   
## specimen trauma, undertaken volunt   
## 0.72 0.72 0.72 0.72   
## along (rang locat head   
## 0.70 0.68 0.60 0.58   
## divid area key ultrasound   
## 0.53 0.53 0.53 0.51

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

## $pain  
## sever left (3/10). (nsaids). 1/10).   
## 0.82 0.80 0.79 0.79 0.79   
## 10. 50-year-old abduction, allodynia anti   
## 0.79 0.79 0.79 0.79 0.79   
## antidepress aspect away back block   
## 0.79 0.79 0.79 0.79 0.79   
## block; buttocks. central cm. contract   
## 0.79 0.79 0.79 0.79 0.79   
## cyst detected. dorsolater drug electric-lik   
## 0.79 0.79 0.79 0.79 0.79   
## eversion, exam experienc extension. flexion,   
## 0.79 0.79 0.79 0.79 0.79   
## foot, gluteal goe hip incis   
## 0.79 0.79 0.79 0.79 0.79   
## incision. inflammatori injected. intercourse. leg.   
## 0.79 0.79 0.79 0.79 0.79   
## length limp manner midlin mild   
## 0.79 0.79 0.79 0.79 0.79   
## mixtur motor movement. muscl neoplasm   
## 0.79 0.79 0.79 0.79 0.79   
## nervous neuropath nonsteroid nsaids, observed.   
## 0.79 0.79 0.79 0.79 0.79   
## ovarian palpat parallel pharmacolog physic   
## 0.79 0.79 0.79 0.79 0.79   
## pinprick plexus point pregabalin. prescrib   
## 0.79 0.79 0.79 0.79 0.79   
## rare. reappear refer removed. repetit   
## 0.79 0.79 0.79 0.79 0.79   
## rest reveal root root. s2.   
## 0.79 0.79 0.79 0.79 0.79   
## sacral schwannoma. schwanomma sciatic sensat   
## 0.79 0.79 0.79 0.79 0.79   
## shown sit start stimul surgery,   
## 0.79 0.79 0.79 0.79 0.79   
## systems. take thigh touch, trigger   
## 0.79 0.79 0.79 0.79 0.79   
## upon woman persist lower surgeri   
## 0.79 0.79 0.78 0.77 0.77   
## leg reduct temperatur toe excis   
## 0.75 0.69 0.69 0.69 0.69   
## (vas day patient analog despit   
## 0.69 0.69 0.66 0.66 0.65   
## later visual peripher histori old   
## 0.65 0.64 0.63 0.61 0.61   
## previous sensori year (vas) free   
## 0.61 0.61 0.61 0.61 0.61   
## nerv pain.   
## 0.57 0.53

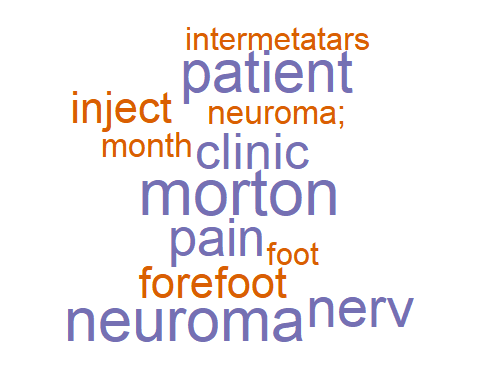
wf <- data.frame(word=names(freq), freq=freq)  
p <- ggplot(subset(wf, freq>16), aes(word, freq))  
p <- p + geom\_bar(stat= 'identity')   
p <- p + theme(axis.text.x=element\_text(angle=90, hjust=1))   
p



wordcloud(names(freq), freq, min.freq=15,colors=brewer.pal(3,'Dark2'))

## Warning in wordcloud(names(freq), freq, min.freq = 15, colors =  
## brewer.pal(3, : keywords: could not be fit on page. It will not be plotted.

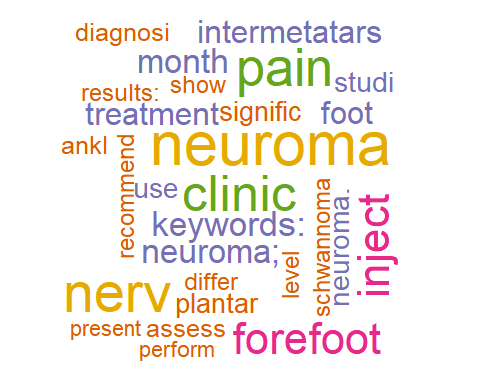
## Warning in wordcloud(names(freq), freq, min.freq = 15, colors =  
## brewer.pal(3, : treatment could not be fit on page. It will not be plotted.



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, : morton could not be fit on page. It will not be plotted.

## Warning in wordcloud(names(freq), freq, max.words = 30, colors =  
## brewer.pal(6, : patient 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, 'Lemmatizedneuromas.csv', row.names=FALSE)

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

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

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

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

## <<DocumentTermMatrix (documents: 20, terms: 1087)>>  
## Non-/sparse entries: 1883/19857  
## Sparsity : 91%  
## Maximal term length: 19  
## Weighting : term frequency (tf)

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

## neuroma nerve patient morton's   
## 38 35 33 29   
## pain forefoot injection clinical   
## 27 25 25 24   
## month ultrasound keyword: neuroma;   
## 20 20 19 18   
## foot treatment intermetatarsal follow   
## 17 17 17 16   
## study plantar significant use   
## 14 13 13 13   
## neuroma. guide case control   
## 13 13 12 12   
## group   
## 12

pain <- as.data.frame(findAssocs(dtmneuromas, "pain", corlimit=0.70))  
  
ultrasounds <- as.data.frame(findAssocs(dtmneuromas, "ultrasound", corlimit=0.75))  
  
  
treatment <- as.data.frame(findAssocs(dtmneuromas, "treatment", corlimit=0.55))  
  
pain

## pain  
## leave 0.86  
## severe 0.85  
## 10. 0.85  
## abduction, 0.85  
## allodynia 0.85  
## anti 0.85  
## antidepressant 0.85  
## aspect 0.85  
## away 0.85  
## back 0.85  
## block 0.85  
## block; 0.85  
## buttock. 0.85  
## central 0.85  
## cm. 0.85  
## contraction 0.85  
## detect. 0.85  
## dorsolateral 0.85  
## drug 0.85  
## electric 0.85  
## eversion, 0.85  
## exam 0.85  
## extension. 0.85  
## flexion, 0.85  
## foot, 0.85  
## gluteal 0.85  
## hip 0.85  
## incision 0.85  
## incision. 0.85  
## inflammatory 0.85  
## inject. 0.85  
## intercourse. 0.85  
## late 0.85  
## leg. 0.85  
## length 0.85  
## limp 0.85  
## manage 0.85  
## manner 0.85  
## midline 0.85  
## mild 0.85  
## mixture 0.85  
## motor 0.85  
## movement. 0.85  
## muscle 0.85  
## neoplasm 0.85  
## nervous 0.85  
## neurologic 0.85  
## neuropathic 0.85  
## nonsteroidal 0.85  
## nsaids 0.85  
## nsaids, 0.85  
## observe. 0.85  
## ovarian 0.85  
## palpation 0.85  
## parallel 0.85  
## patient's 0.85  
## persistent 0.85  
## pharmacological 0.85  
## physical 0.85  
## pinprick 0.85  
## plexus 0.85  
## point 0.85  
## pregabalin. 0.85  
## prescribe 0.85  
## rare. 0.85  
## reappear 0.85  
## refer 0.85  
## remove. 0.85  
## repetitive 0.85  
## rest 0.85  
## reveal 0.85  
## root 0.85  
## root. 0.85  
## s2. 0.85  
## sacral 0.85  
## schwannoma. 0.85  
## schwanommas 0.85  
## sciatic 0.85  
## sensation 0.85  
## severity 0.85  
## sit 0.85  
## start 0.85  
## stimulation 0.85  
## stimulator 0.85  
## surgery, 0.85  
## system. 0.85  
## take 0.85  
## thigh 0.85  
## touch, 0.85  
## trigger 0.85  
## upon 0.85  
## woman 0.85  
## leg 0.81  
## old 0.81  
## low 0.80  
## excision 0.75  
## day 0.75  
## several 0.75  
## despite 0.71

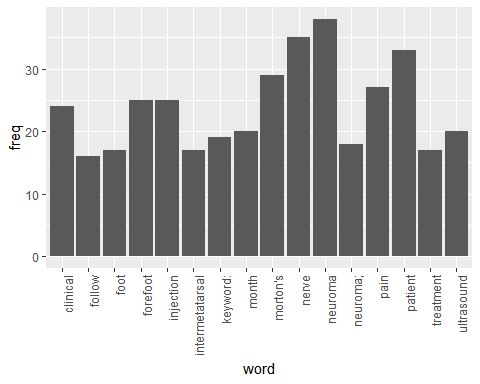
ultrasounds

## ultrasound  
## point:? 0.87  
## work 0.87  
## include 0.83  
## guide 0.79  
## ultrasonography 0.79  
## 0.5 0.79  
## 0.5, 0.79  
## 001; 0.79  
## 002; 0.79  
## 010; 0.79  
## 018 0.79  
## 020 0.79  
## 047 0.79  
## 1.8 0.79  
## 1.9 0.79  
## 2.0, 0.79  
## 3.0 0.79  
## 3.1 0.79  
## 31.2 0.79  
## 31.5 0.79  
## 37.7 0.79  
## 38.5 0.79  
## 38.8 0.79  
## 5.2 0.79  
## 5.5 0.79  
## 5.6 0.79  
## age 0.79  
## appropriate 0.79  
## day, 0.79  
## disability 0.79  
## evaluator 0.79  
## initial 0.79  
## injection.? 0.79  
## interventional 0.79  
## manchester 0.79  
## mepivacaine 0.79  
## mfpds 0.79  
## mfpds: 0.79  
## patient, 0.79  
## patient.? 0.79  
## randomise 0.79  
## stage 0.79  
## triamcinolone 0.79  
## va: 0.79  
## versus 0.79  
## web 0.79  
## blind 0.77

treatment

## treatment  
## randomize 0.80  
## last 0.73  
## case 0.65  
## anesthetic 0.59  
## 2015. 0.58  
## 283 0.58  
## august 0.58  
## benefit 0.58  
## bibliographic 0.58  
## cochrane 0.58  
## collect. 0.58  
## conservative, 0.58  
## conservative; 0.58  
## criterion. 0.58  
## dare. 0.58  
## different 0.58  
## english 0.58  
## evaluation 0.58  
## find, 0.58  
## include. 0.58  
## inclusion 0.58  
## independently. 0.58  
## infiltrative 0.58  
## infiltrative. 0.58  
## infiltrative; 0.58  
## library, 0.58  
## medline, 0.58  
## meet 0.58  
## operative 0.58  
## operative, 0.58  
## outcome 0.58  
## patients' 0.58  
## primary 0.58  
## quality, 0.58  
## ratio 0.58  
## rcts 0.58  
## result. 0.58  
## retrieve 0.58  
## reviewer 0.58  
## risk 0.58  
## satisfaction 0.58  
## series 0.58  
## surgery; 0.58  
## title 0.58  
## trial 0.58  
## type 0.58  
## 48. 0.58  
## alone 0.58  
## alone. 0.58  
## baseline 0.58  
## baseline, 0.58  
## collect 0.58  
## comparison 0.58  
## corticosteroid; 0.58  
## experimental 0.58  
## forty 0.58  
## injection, 0.58  
## interphalangeal 0.58  
## local 0.58  
## neuroma: 0.58  
## obtain 0.58  
## orthopaedic 0.58  
## placebo 0.58  
## plus 0.58  
## receive 0.58  
## request 0.58  
## researcher 0.58  
## superior 0.58  
## unclear. 0.58  
## value. 0.58

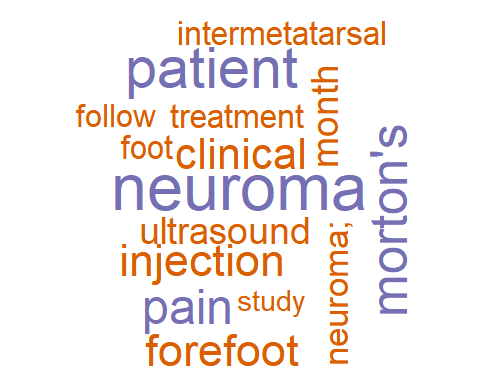
wf <- data.frame(word=names(freq), freq=freq)  
p <- ggplot(subset(wf, freq>15), aes(word, freq))  
p <- p + geom\_bar(stat= 'identity')   
p <- p + theme(axis.text.x=element\_text(angle=90, hjust=1))   
p



wordcloud(names(freq), freq, min.freq=14,colors=brewer.pal(3,'Dark2'))

## Warning in wordcloud(names(freq), freq, min.freq = 14, colors =  
## brewer.pal(3, : nerve could not be fit on page. It will not be plotted.

## Warning in wordcloud(names(freq), freq, min.freq = 14, colors =  
## brewer.pal(3, : keyword: could not be fit on page. It will not be plotted.



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

## Warning in wordcloud(names(freq), freq, max.words = 40, colors =  
## brewer.pal(6, : neuroma could not be fit on page. It will not be plotted.

## Warning in wordcloud(names(freq), freq, max.words = 40, colors =  
## brewer.pal(6, : injection could not be fit on page. It will not be plotted.

## Warning in wordcloud(names(freq), freq, max.words = 40, colors =  
## brewer.pal(6, : nerve could not be fit on page. It will not be plotted.

## Warning in wordcloud(names(freq), freq, max.words = 40, colors =  
## brewer.pal(6, : recommendation could not be fit on page. It will not be  
## plotted.

## Warning in wordcloud(names(freq), freq, max.words = 40, colors =  
## brewer.pal(6, : metatarsal could not be fit on page. It will not be  
## plotted.

## Warning in wordcloud(names(freq), freq, max.words = 40, colors =  
## brewer.pal(6, : group could not be fit on page. It will not be plotted.

## Warning in wordcloud(names(freq), freq, max.words = 40, colors =  
## brewer.pal(6, : clinical could not be fit on page. It will not be plotted.

## Warning in wordcloud(names(freq), freq, max.words = 40, colors =  
## brewer.pal(6, : ultrasound could not be fit on page. It will not be  
## plotted.

## Warning in wordcloud(names(freq), freq, max.words = 40, colors =  
## brewer.pal(6, : neuroma; could not be fit on page. It will not be plotted.

