iemisctext: Example Analysis using the tm package

Irucka Embry

2024-05-31

Contents

Examples

```r
install.load::load_package("iemisctext", "tm", "data.table", "ggwordcloud")
# load needed packages using the load_package function from the install.load
# package (it is assumed that you have already installed these packages)


# appeal_womanhood_world
data(appeal_womanhood_world)

# create a document term matrix of appeal_womanhood_world
aww <- DocumentTermMatrix(appeal_womanhood_world, control = list(removePunctuation = TRUE,
    stopwords = TRUE, weighting = function(x) weightSMART(x, spec = "ntc")))

aww

## <<DocumentTermMatrix (documents: 1, terms: 170)>>
## Non-/sparse entries: 0/170
## Sparsity           : 100%
## Maximal term length: 13
## Weighting          : SMART ntc (SMART)
# get the word frequency table
aww_tf <- as.data.table(termFreq(appeal_womanhood_world$content[[1]]$content, control = list(removePunct
    stopwords = TRUE, weighting = function(x) weightSMART(x, spec = "ntc"))), keep.rownames = TRUE)

# create a word cloud
ggwordcloud(aww_tf$V1, aww_tf$V2)
```

women

day let men will

sacred great

sons questions

```
# anarchy
data(anarchy)

# create a document term matrix of anarchy
ancy <- DocumentTermMatrix(anarchy, control = list(removePunctuation = TRUE, stopwords = TRUE,
    weighting = function(x) weightSMART(x, spec = "ntc")))

ancy

## <<DocumentTermMatrix (documents: 1, terms: 785)>>
## Non-/sparse entries: 0/785
## Sparsity           : 100%
## Maximal term length: 14
## Weighting          : SMART ntc (SMART)
# get the word frequency table
ancy_tf <- as.data.table(termFreq(anarchy$content[[1]]$content, control = list(removePunctuation = TRUE
    stopwords = TRUE, weighting = function(x) weightSMART(x, spec = "ntc")), keep.rownames = TRUE)

# create a word cloud
ggwordcloud(ancy_tf$V1, ancy_tf$V2)
```
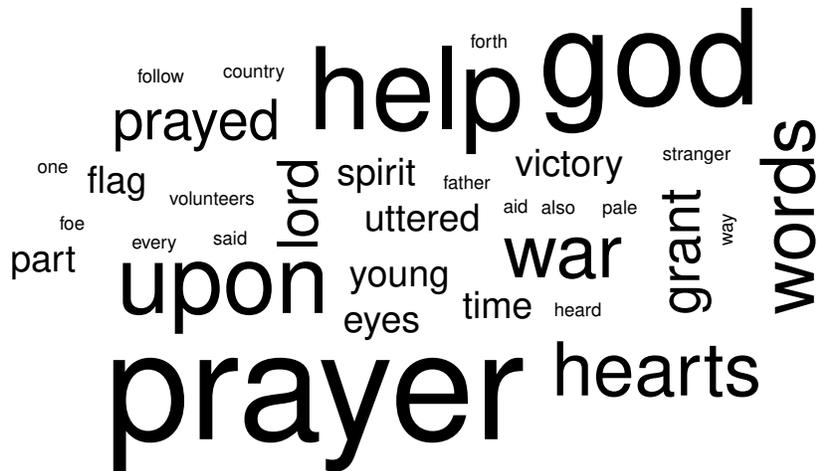
thou

like horses stand earth

see wealth strong

words cried many wide dead daily feet

food will till let weak heard rich

make blood rest rode lay pale passed foot following

english find multitude wind tis laws

old come whose knew great one gold around day war

well away fast name every anarchy arms heart never speak art

must dew death mighty made light upon came home looked trampling feel toil god law

children land men sea king

tyrants even misery first air free felt thee
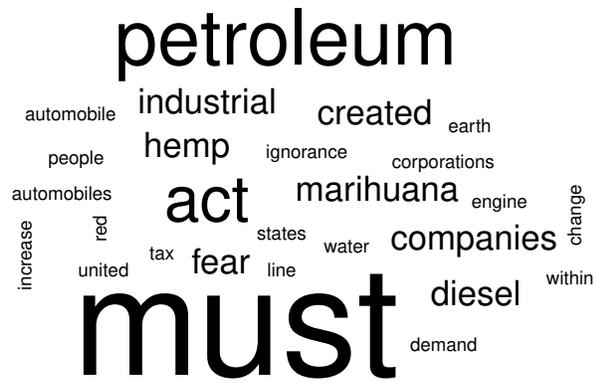
england thy

```
# war_prayer
data(war_prayer)

# create a document term matrix of war_prayer
wp <- DocumentTermMatrix(war_prayer, control = list(removePunctuation = TRUE, stopwords = TRUE,
    weighting = function(x) weightSMART(x, spec = "ntc")))

wp

## <<DocumentTermMatrix (documents: 1, terms: 523)>>
```

```
## Non-/sparse entries: 0/523
## Sparsity           : 100%
## Maximal term length: 13
## Weighting          : SMART ntc (SMART)
```

```r
# get the word frequency table
wp_tf <- as.data.table(termFreq(war_prayer$content[[1]]$content, control = list(removePunctuation = TRUE,
    stopwords = TRUE, weighting = function(x) weightSMART(x, spec = "ntc"))), keep.rownames = TRUE)

# create a word cloud
ggwordcloud(wp_tf$V1, wp_tf$V2)
```

forth

follow   country

prayed   help   god

one flag   spirit   father   victory   stranger

volunteers

foe   lord   aid also   pale   grant   words

part   every   said   uttered   way

upon   young   war

time   heard

eyes

prayer   hearts

```r
# war_racket
data(war_racket)

# create a document term matrix of war_racket
wr <- DocumentTermMatrix(war_racket, control = list(removePunctuation = TRUE, stopwords = TRUE,
    weighting = function(x) weightSMART(x, spec = "ntc")))

wr
```

```
## <<DocumentTermMatrix (documents: 1, terms: 1475)>>
## Non-/sparse entries: 0/1475
## Sparsity           : 100%
## Maximal term length: 15
## Weighting          : SMART ntc (SMART)
```

```r
# get the word frequency table
wr_tf <- as.data.table(termFreq(war_racket$content[[1]]$content, control = list(removePunctuation = TRUE,
    stopwords = TRUE, weighting = function(x) weightSMART(x, spec = "ntc"))), keep.rownames = TRUE)

# create a word cloud
ggwordcloud(wr_tf$V1, wr_tf$V2)
```
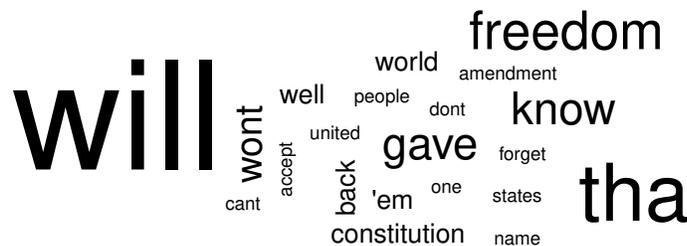
```r
# connect_dots
data(connect_dots)

# create a document term matrix of connect_dots
cds <- DocumentTermMatrix(connect_dots, control = list(removePunctuation = TRUE,
    stopwords = TRUE, weighting = function(x) weightSMART(x, spec = "ntc")))

cds

## <<DocumentTermMatrix (documents: 1, terms: 337)>>
## Non-/sparse entries: 0/337
## Sparsity           : 100%
## Maximal term length: 19
## Weighting          : SMART ntc (SMART)
# get the word frequency table
cds_tf <- as.data.table(termFreq(connect_dots$content[[1]]$content, control = list(removePunctuation = T
    stopwords = TRUE, weighting = function(x) weightSMART(x, spec = "ntc")), keep.rownames = TRUE)

# create a word cloud
ggwordcloud(cds_tf$V1, cds_tf$V2)
```

petroleum

automobile industrial created earth

people hemp ignorance corporations

automobiles marihuana engine

act states companies change

increase red water

united tax fear line

must diesel within

demand

```r
# us_them
data(us_them)

# create a document term matrix of us_them
ut <- DocumentTermMatrix(us_them, control = list(removePunctuation = TRUE, stopwords = TRUE,
    weighting = function(x) weightSMART(x, spec = "ntc")))

ut

## <<DocumentTermMatrix (documents: 1, terms: 247)>>
## Non-/sparse entries: 0/247
## Sparsity           : 100%
## Maximal term length: 71
## Weighting          : SMART ntc (SMART)

# get the word frequency table
ut_tf <- as.data.table(termFreq(us_them$content[[1]]$content, control = list(removePunctuation = TRUE,
    stopwords = TRUE, weighting = function(x) weightSMART(x, spec = "ntc"))), keep.rownames = TRUE)

# create a word cloud
ggwordcloud(ut_tf$V1, ut_tf$V2)
```

freedom

world amendment

will well people dont know

wont united

accept gave forget

cant back 'em one states tha

constitution name

```r
# climate_strange
data(climate_strange)

# create a document term matrix of climate_strange
cs <- DocumentTermMatrix(climate_strange, control = list(removePunctuation = TRUE,
    stopwords = TRUE, weighting = function(x) weightSMART(x, spec = "ntc")))

cs

## <<DocumentTermMatrix (documents: 1, terms: 525)>>
## Non-/sparse entries: 0/525
## Sparsity           : 100%
## Maximal term length: 24
```
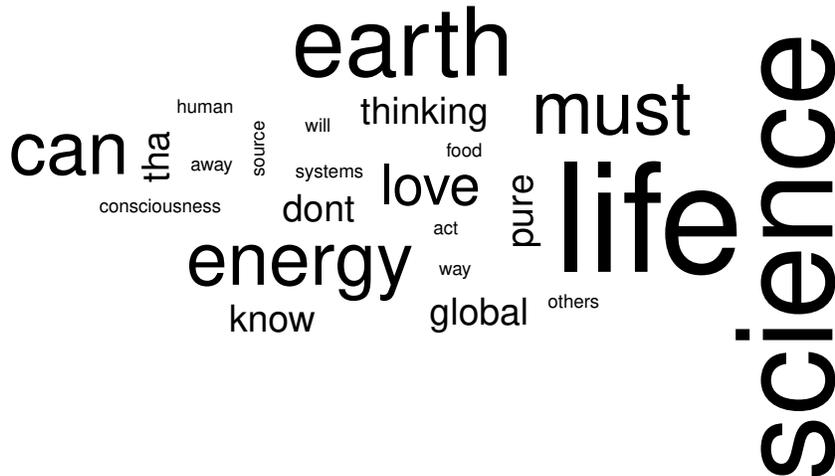
```
## Weighting          : SMART ntc (SMART)
# get the word frequency table
cs_tf <- as.data.table(termFreq(climate_strange$content[[1]]$content, control = list(removePunctuation =
    stopwords = TRUE, weighting = function(x) weightSMART(x, spec = "ntc"))), keep.rownames = TRUE)

# create a word cloud
ggwordcloud(cs_tf$V1, cs_tf$V2)
```


EcoC²S Links

EcoC²S Home – https://www.ecoccs.com/
About EcoC²S – https://www.ecoccs.com/about-ecoc2s.html
Services – https://www.ecoccs.com/services.html
1 Stop Shop – https://www.ecoccs.com/other-biz.html
Products – https://www.questionuniverse.com/products.html
Media – https://www.ecoccs.com/media.html
Resources – https://www.ecoccs.com/resources.html
R Trainings and Resources provided by EcoC²S (Irucka Embry, E.I.T.) – https://www.ecoccs.com/rtraining.
html

Copyright and License

All R code written by Irucka Embry is distributed under the GPL-3 (or later) license, see the GNU General
Public License {GPL} page.

All written content originally created by Irucka Embry is copyrighted under the Creative Commons
Attribution-ShareAlike 4.0 International license. All other written content retains the copyright of the
original author(s).

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International license.