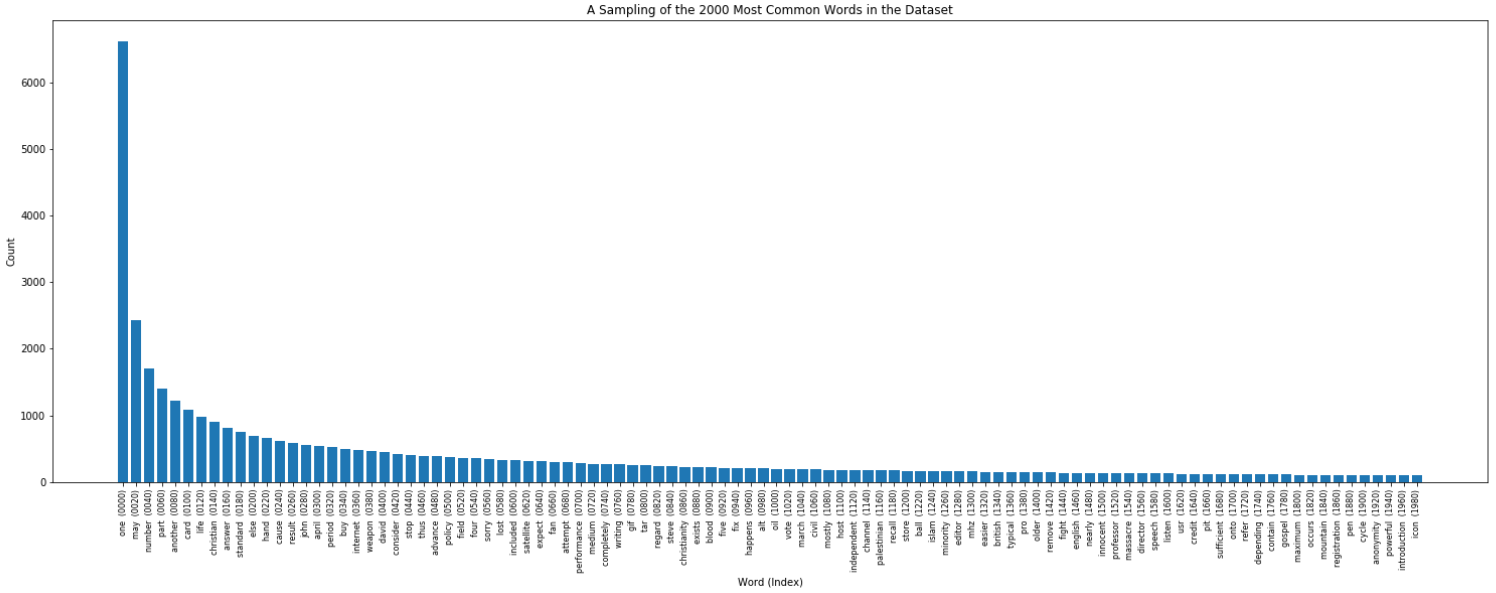
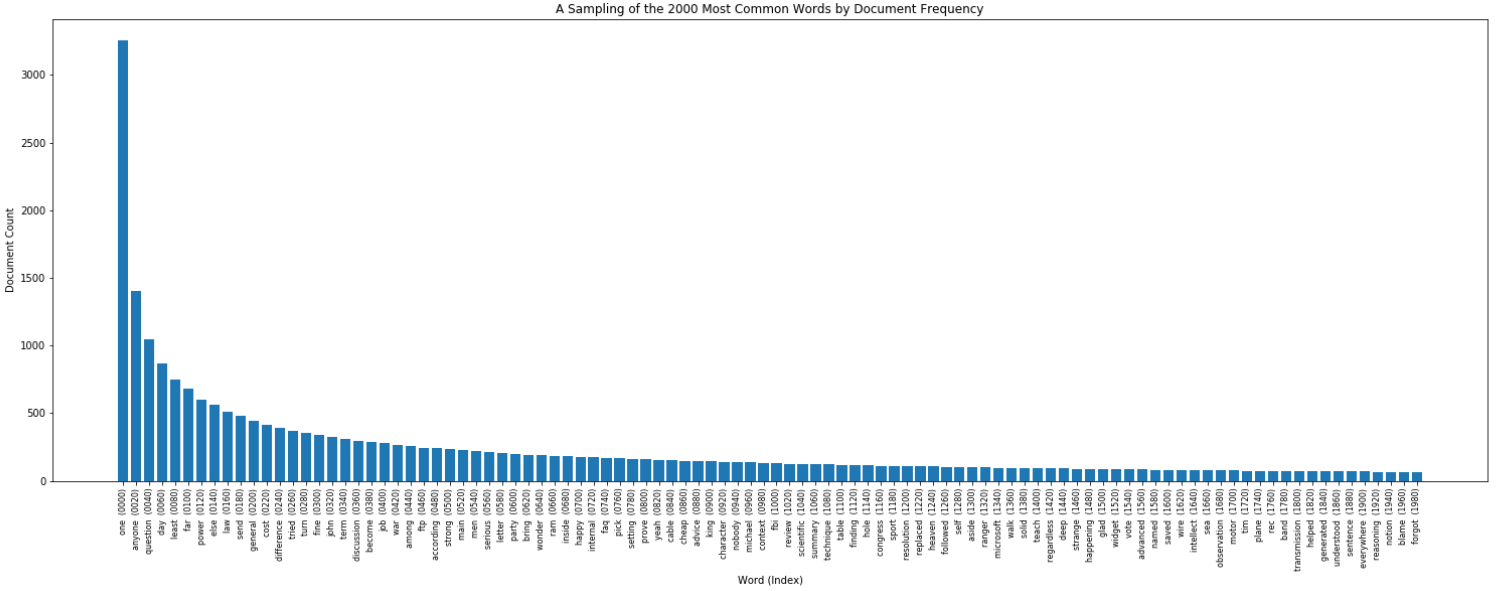
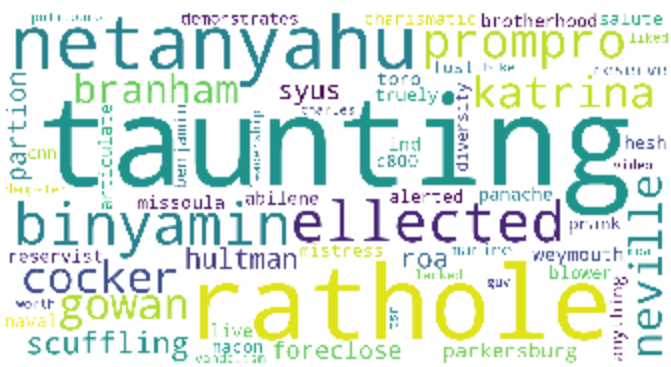
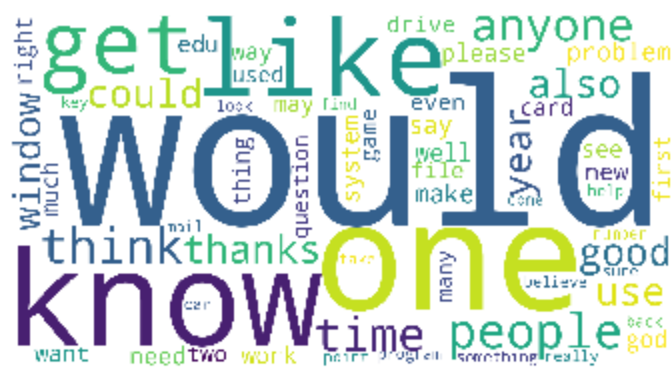
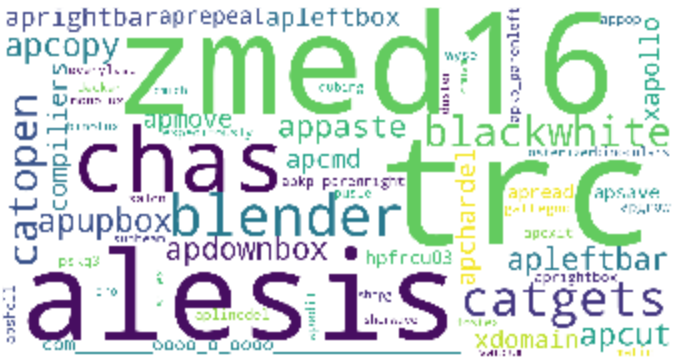
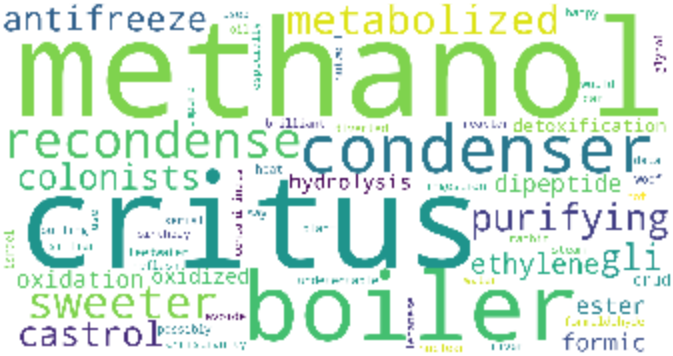

A Sampling of the 2000 Most Common Words in the Dataset
Count
Word (Index)


A Sampling of the 2000 Most Common Words by Document Frequency
Document Count
Word (Index)


melittin
ihr
pads
tinnitus
precompiled
mike_johnson
preferably
hp2xx
copro
perret
mentorg
alle
haufen
jetzt
ingénieur
aroundе
hoffnung
klebt
angeben
voic
zeuge
wollt
suisse
sind
nur
immerfort
verteilt
koph
locle
daten
aoutput
schwindet
xssr
muessen
schalem
bauteile
einem
wie
ich

taunting
netanyahu
prompro
ratHole
binyamin
ellected
katrina
neville
branham
syus
cocker
gowan
scuffling
hultman
roa
weymouth
foreclose
parkersburg
partion
missoula
abilene
alerted
panache
diversity
demonstrates
charismatic
brotherhood
salute
reservist
naval
mistress
blower
prank
hesh
truely
toro

would
know
one
like
get
anyone
could
also
think
thanks
time
people
good
use
year
window
right
drive
please
problem
used
way
edu
card
even
say
well
file
make
see
new
first
system
game
question
thing
much
many
want
need
two
work
god
something
really


op_cols
rawley
aantal_snijpunten
tittle
cindy
vector2d
op_rows
hcl
puntbinnenpolygoon
polygoon
jad
eastwick
clibsucces


zmed16
trc
chas
blackwhite
blender
alesis
catgets
apcut

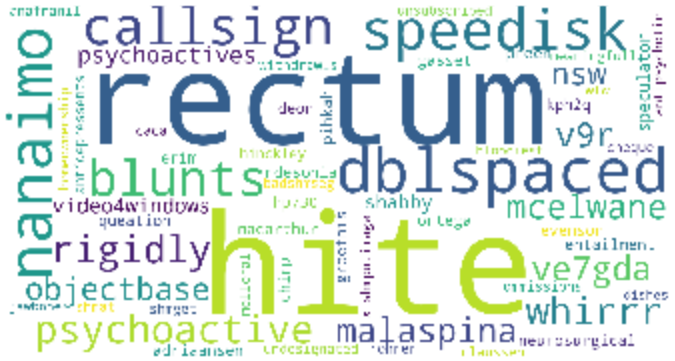

antifreeze
methanol
metabolized
recondense
condenser
colonists
critus
purifying
boiler
oxidation
sweeter
castrol


een
dit
niet
horloges
kaldilar
sonra
diye
meer
deze
rolex
ordularinin
alasini
yok
yuru
plaka
herkesi
seferlik


powerkey
inhis
m2c
excedrin
sophisicated
did1
backorder
spdcc
sittler
ima
olcyzk
ghq
vaive
p3y
cartlidge
rayssd
dalva
tis
pu9
linus
synergism
glenwood
az2tv6l

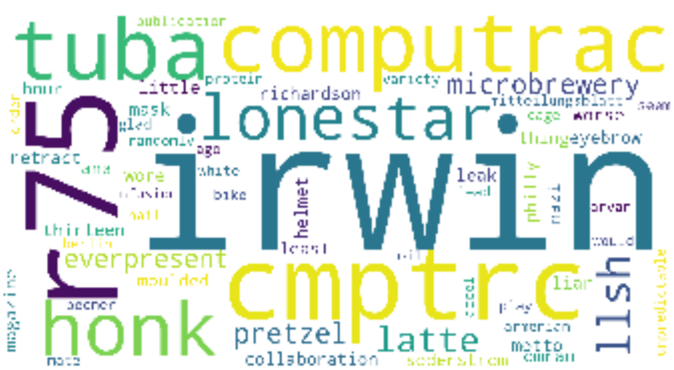

christie
recension
bilinsky
kille
resaon
essene
claromontanus
bezae
reponded


same_name
sparcclassic
steinn
kuiper
ctritest
annotation
aftershave
fairings
bogota
bwc
karr
mjs
bloodline
cg3
fallahin
paz
karla
mfc
qb1
shafting
yip
freehand
sharen
levitical
sigurdsson

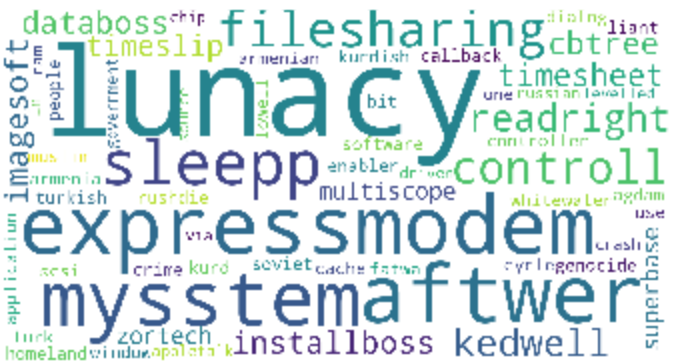

callsign
speedisk
psychoactives
rectum
nsw
v9r
nanaimo
blunts
dblspaced
video4windows
mcelwane
rigidly
hite
ve7gda
objectbase
whirrr
psychoactive
malaspina

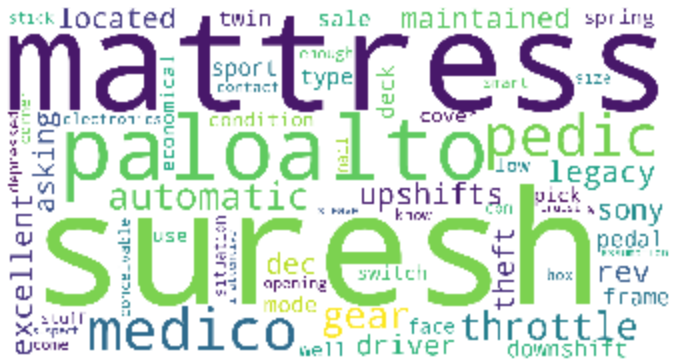

max
g9v
b8f
a86
razzing
stafford
hfe
chz
cb750
geeky
bhj
giz
vce
winona
handson
thruput
sckz1300
ican
b9r
kjz
den
rlk
x_s
scx
bxn
ahf
jur
w47
lt1839
lt5839
cj1
lhz
froumentin


xew
tesrt
trinomials
scand
zbl
tverberg
mpm
irreducibility
kurtaran
thermocouple
kubilay
kuvvetli
ruhlari
kirden
indeterminacy
kultigir
decoupling
ruzgardir
boucher
brenner
motherland


irwin
computrac
tuba
r75
lonestar
cmptrc
honk
microbrewery
richardson
everpresent
pretzel
latte
llsh
collaboration
helmet
thirteen
retract
eyebrow
philly
leak
little
moulded
magazine


goggles
retriever
sidecar
scope
onedown
watched
camera
oneup
playback
record
xtrap
joke
dog
extension
cmaeras
answer
wwii
wearing
modify
would
confrontation
always
project
testing
sampling
style
golden
event
mhz
boulder
tricky
mit
time

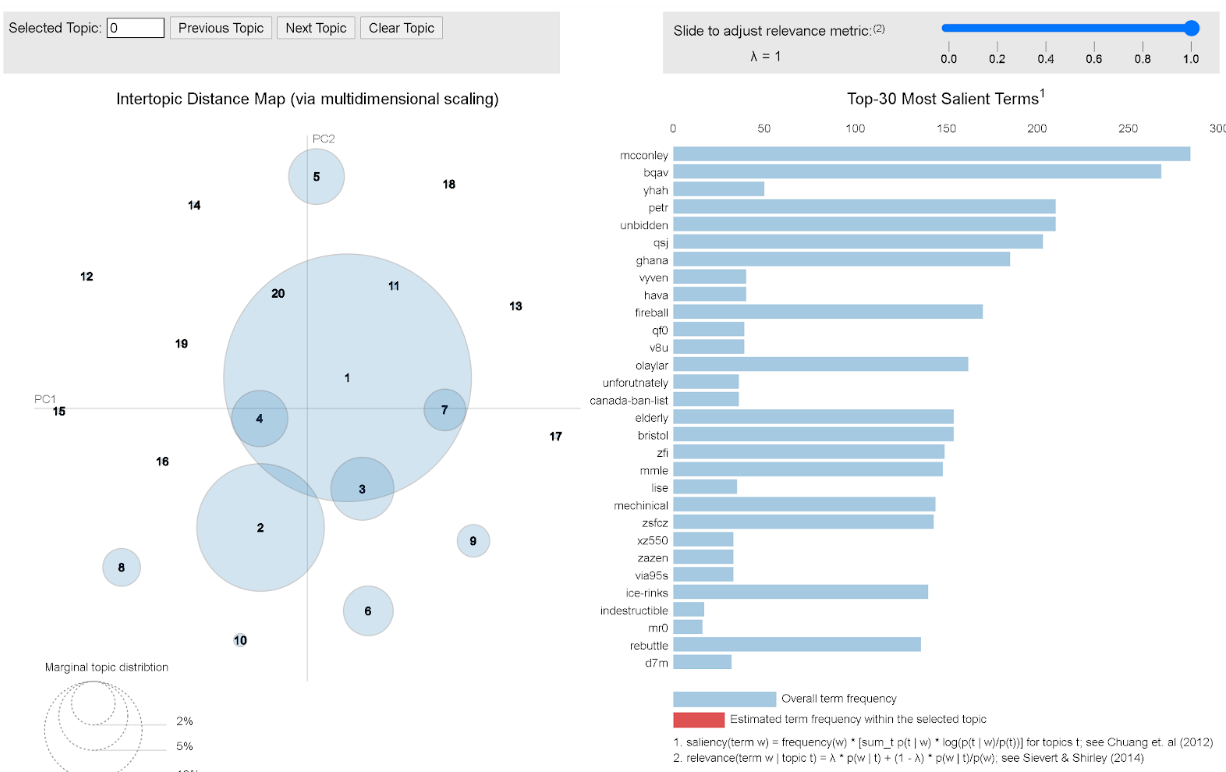

databoss
filesharing
timeslip
cbtree
imagesoft
lunacy
timesheet
readright
sleepp
controll
multiscope
expressmodem
mysstemaftwer
installboss
kedwell
zorlech
superbase

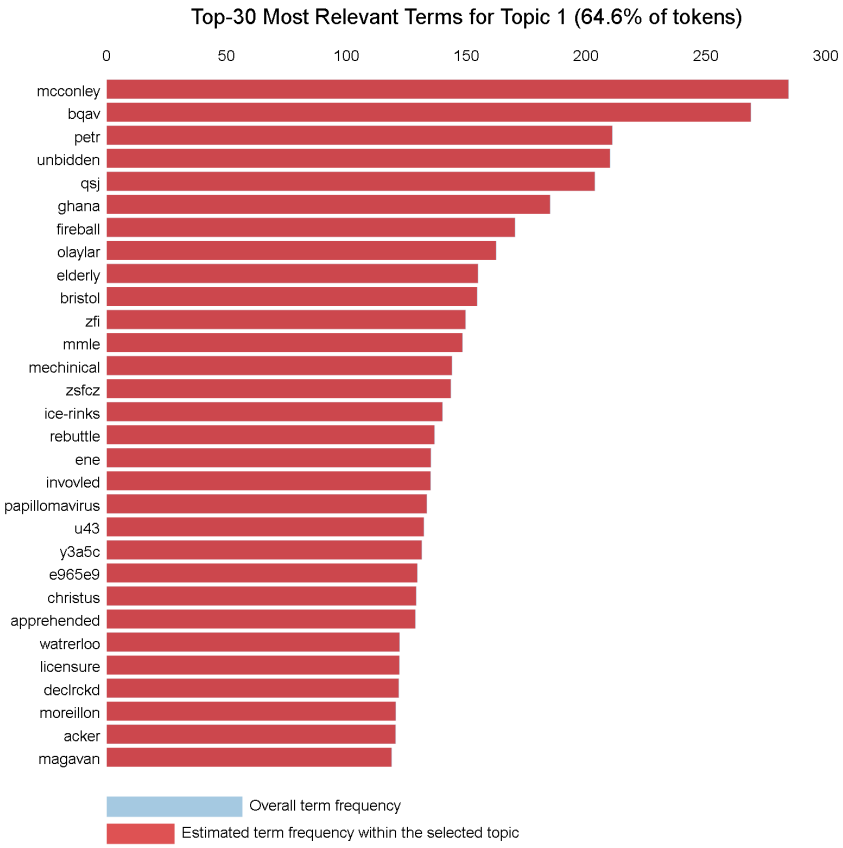

mattress
located
twin
sale
maintained
paloalto
pedic
asking
automatic
upshifts
legacy
suresh
sony
excellent
medico
gear
throttle
theft
rev
dec
driver
downshift

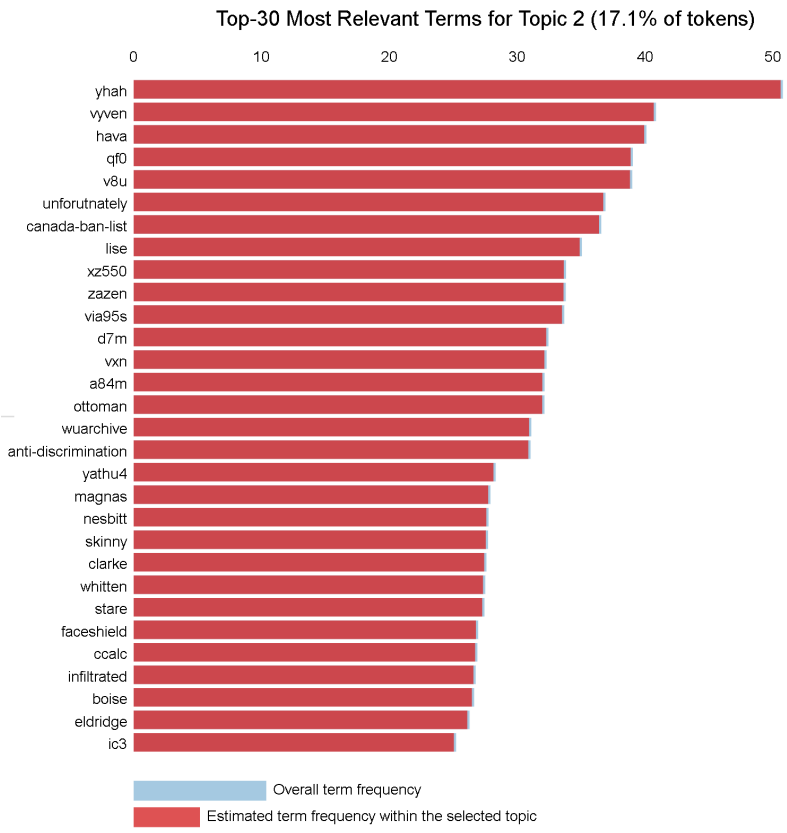

jurgen
whadda
durant
genashor
hypoglycemic
kssimon
bloodcount
anello
vell
glycemic
vit
fazah
nuetzman
iubacs
piney
leefing
ballclub
easytalk
chevrolet
gendel
indy
nuet_ke
muzzah
mearly
brando
whatching
pancreatic
lois

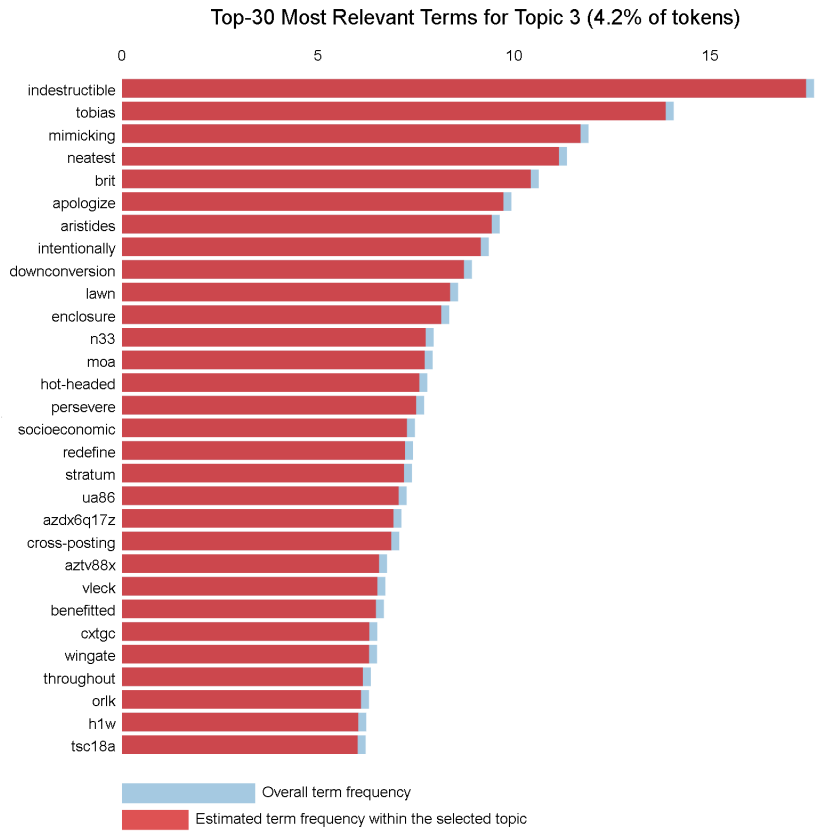

decided
annulled
commandment
darius
cockroach
oppose
soc
argument
chz
grafted
talkreligion
clh
thud
sabbath
israel
ten
mean
present
portion
scripture
applicable
biblical
amazed
since
ceremonial
seventh
requirement
answered
gentile
christian
commonwealth
talent
bible

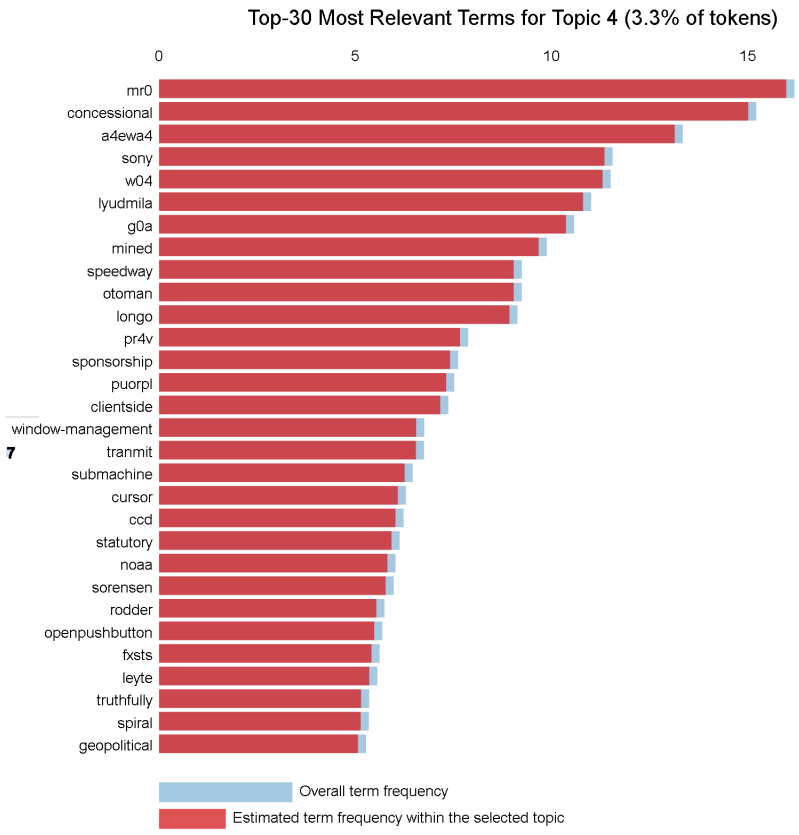

melido
batchfile
acne
dadies
scaliness
shampoo
sterilize
mailman
rendu
osler
sacremento
neurovascular
hydrophilic
iodine
n3p
dalacin
escalate
rinse
dfw
massachussetts
clearasil
overdose
doco
nosebleed
pimple
bathwater
ascendent
soften
submerging
opportunist


Selected Topic: 0
Previous Topic
Next Topic
Clear Topic
Slide to adjust relevance metric:(2)
λ = 1
0.0
0.2
0.4
0.6
0.8
1.0
Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
Marginal topic distribtion
2%
5%
10%
Top-30 Most Salient Terms1
0
50
100
150
200
250
300
mcconley
bqav
yhah
petr
unbidden
qsj
ghana
vyven
hava
fireball
qf0
v8u
olaylar
unforutnately
canada-ban-list
elderly
bristol
zfi
mmle
lise
mechinical
zsfcz
xz550
zazen
via95s
ice-rinks
indestructible
mr0
rebuttle
d7m
Overall term frequency
Estimated term frequency within the selected topic
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

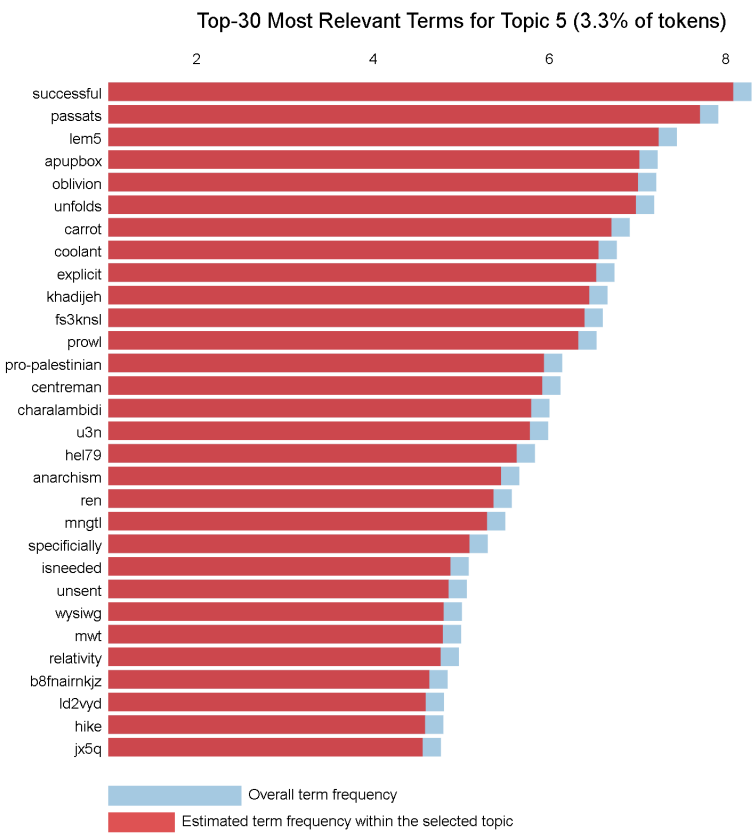

Top-30 Most Relevant Terms for Topic 1 (64.6% of tokens)
0
50
100
150
200
250
300
mcconley
bqav
petr
unbidden
qsj
ghana
fireball
olaylar
elderly
bristol
zfi
mmle
mechinical
zsfcz
ice-rinks
rebuttle
ene
invovled
papillomavirus
u43
y3a5c
e965e9
christus
apprehended
watrerloo
licensure
declrckd
moreillon
acker
magavan
Overall term frequency
Estimated term frequency within the selected topic

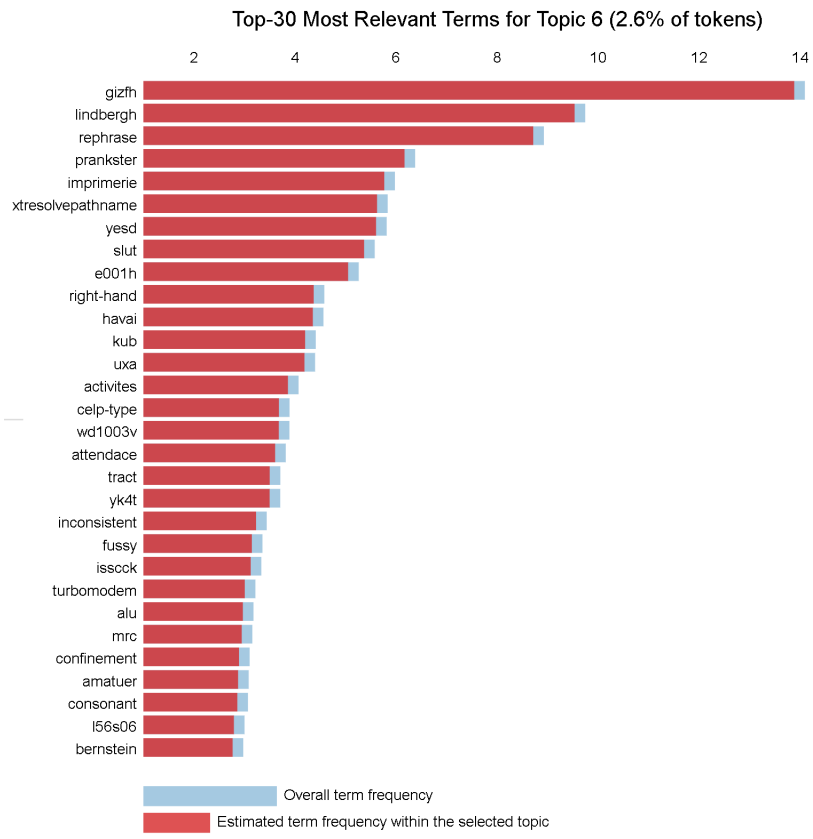

Top-30 Most Relevant Terms for Topic 2 (17.1% of tokens)
0
10
20
30
40
50
yhah
vyven
hava
qf0
v8u
unforutnately
canada-ban-list
lise
xz550
zazen
via95s
d7m
vxn
a84m
ottoman
wuarchive
anti-discrimination
yathu4
magnas
nesbitt
skinny
clarke
whitten
stare
faceshield
ccalc
infiltrated
boise
eldridge
ic3
Overall term frequency
Estimated term frequency within the selected topic

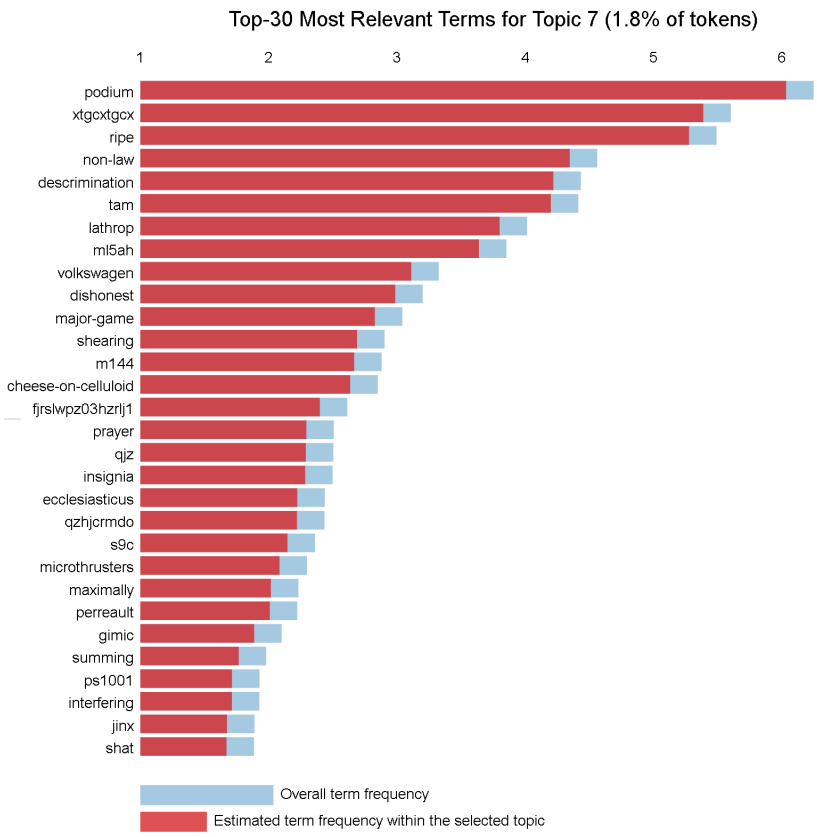

Top-30 Most Relevant Terms for Topic 3 (4.2% of tokens)
0
5
10
15
indestructible
tobias
mimicking
neatest
brit
apologize
aristides
intentionally
downconversion
lawn
enclosure
n33
moa
hot-headed
persevere
socioeconomic
redefine
stratum
ua86
azdx6q17z
cross-posting
aztv88x
vleck
benefitted
cxtgc
wingate
throughout
orlk
h1w
tsc18a
Overall term frequency
Estimated term frequency within the selected topic

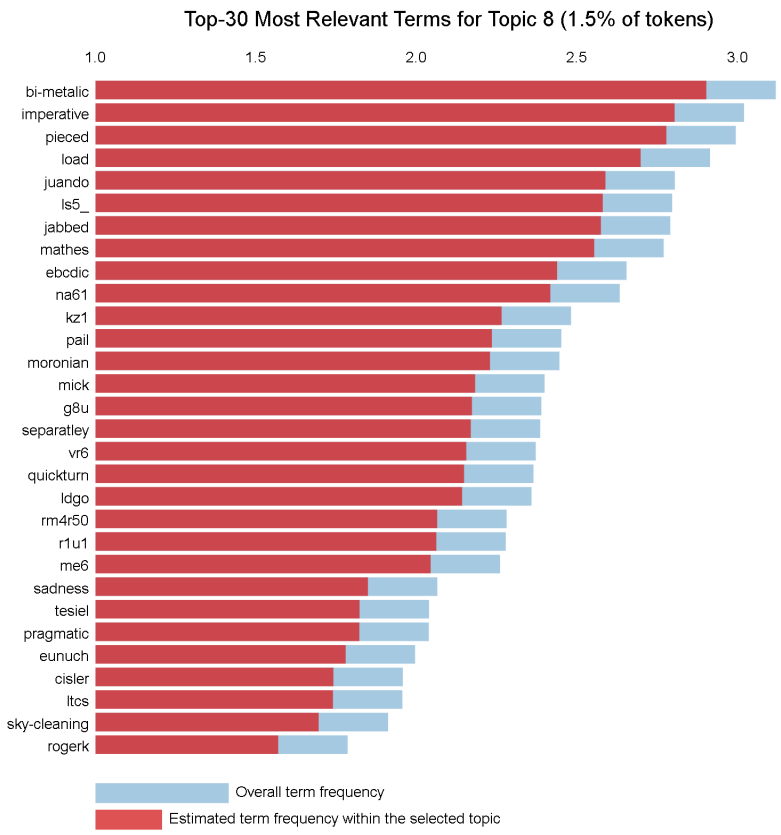

Top-30 Most Relevant Terms for Topic 4 (3.3% of tokens)
0
5
10
15
mr0
concessional
a4ewa4
sony
w04
lyudmila
g0a
mined
speedway
otoman
longo
pr4v
sponsorship
puorpl
clientside
window-management
tranmit
submachine
cursor
ccd
statutory
noaa
sorensen
rodder
openpushbutton
fxsts
leyte
truthfully
spiral
geopolitical
Overall term frequency
Estimated term frequency within the selected topic

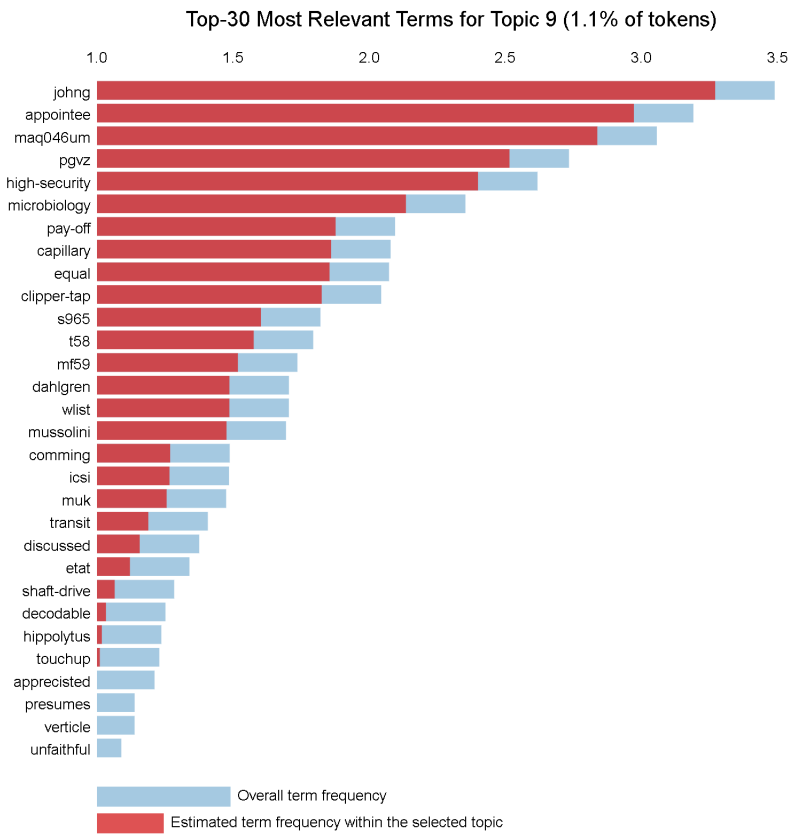
7

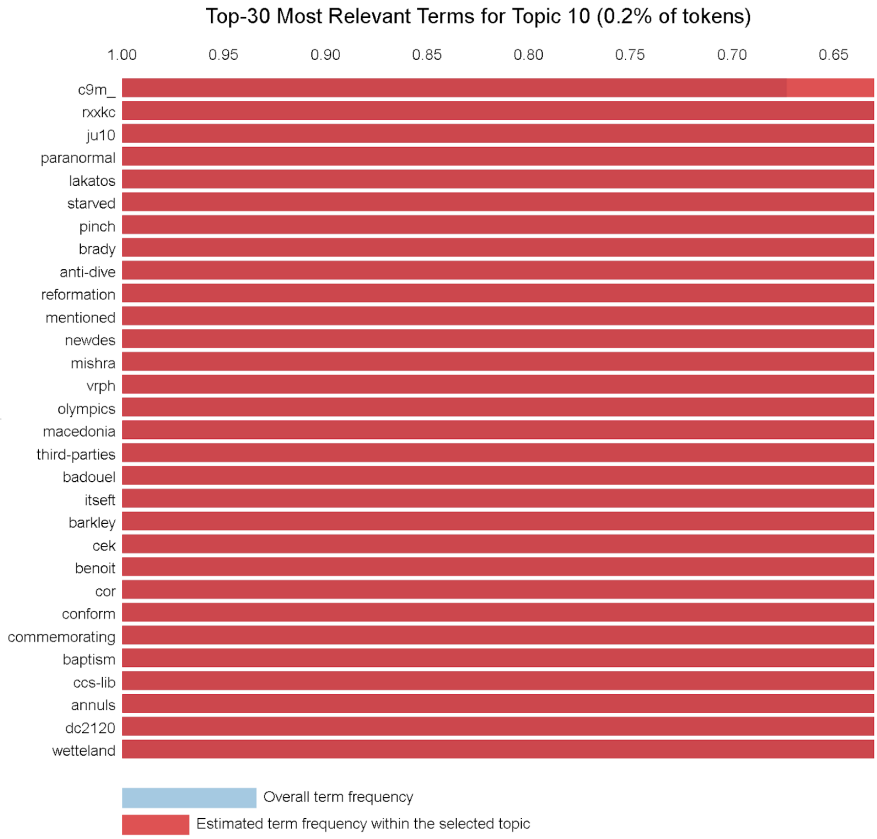

Top-30 Most Relevant Terms for Topic 5 (3.3% of tokens)
2
4
6
8
successful
passats
lem5
apupbox
oblivion
unfolds
carrot
coolant
explicit
khadijeh
fs3knsl
prowl
pro-palestinian
centreman
charalambidi
u3n
hel79
anarchism
ren
mngtl
specificially
isneeded
unsent
wysiwg
mwt
relativity
b8fnairnkjz
ld2vyd
hike
jx5q
Overall term frequency
Estimated term frequency within the selected topic

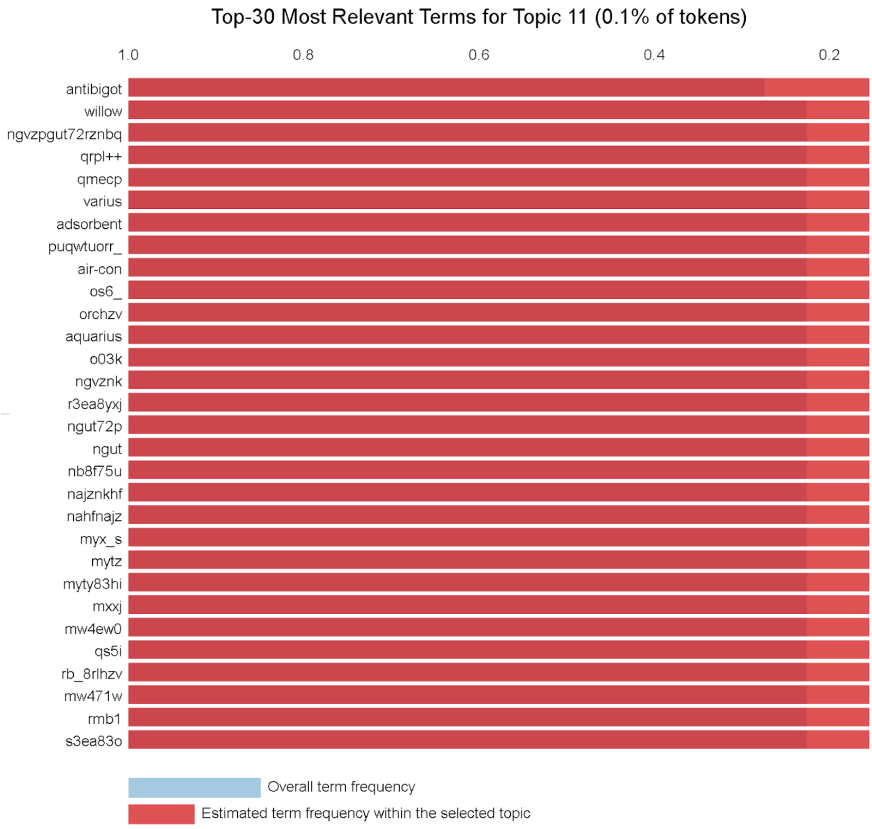

Top-30 Most Relevant Terms for Topic 6 (2.6% of tokens)
2
4
6
8
10
12
14
gizfh
lindbergh
rephrase
prankster
imprimerie
xtresolvepathname
yesd
slut
e001h
right-hand
havai
kub
uxa
activites
celp-type
wd1003v
attendace
tract
yk4t
inconsistent
fussy
isscck
turbomodem
alu
mrc
confinement
amatuer
consonant
l56s06
bernstein
Overall term frequency
Estimated term frequency within the selected topic

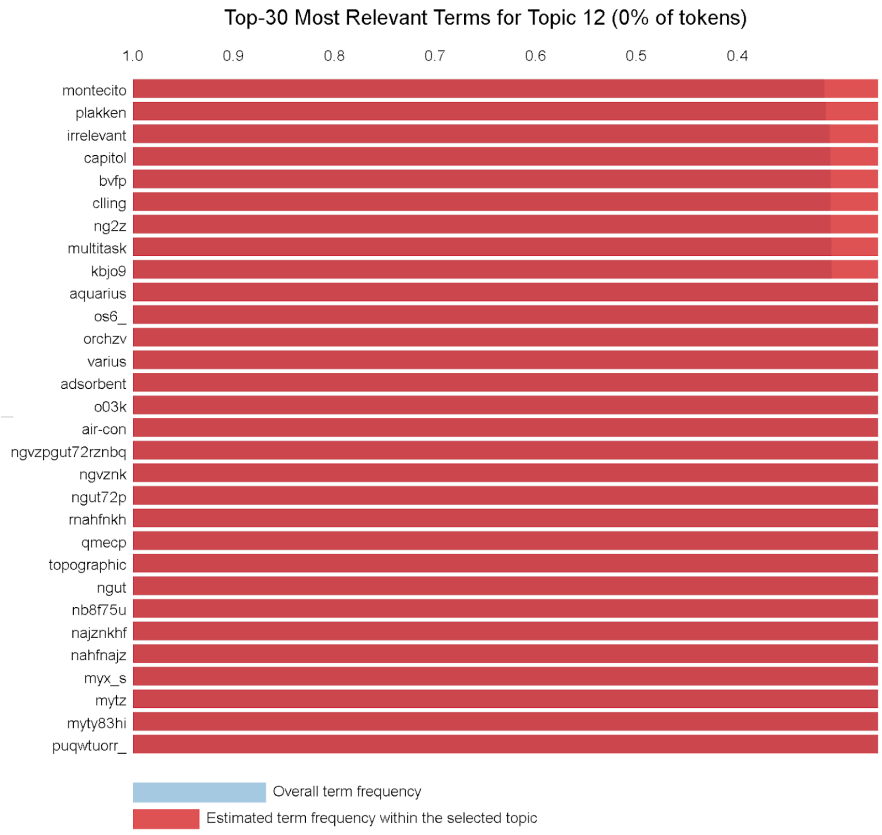

Top-30 Most Relevant Terms for Topic 7 (1.8% of tokens)
1
2
3
4
5
6
podium
xtgcxtgcx
ripe
non-law
descrimination
tam
lathrop
ml5ah
volkswagen
dishonest
major-game
shearing
m144
cheese-on-celluloid
fjrslwpz03hzrlj1
prayer
qjz
insignia
ecclesiasticus
qzhjcrmdo
s9c
microthrusters
maximally
perreault
gimic
summing
ps1001
interfering
jinx
shat
Overall term frequency
Estimated term frequency within the selected topic

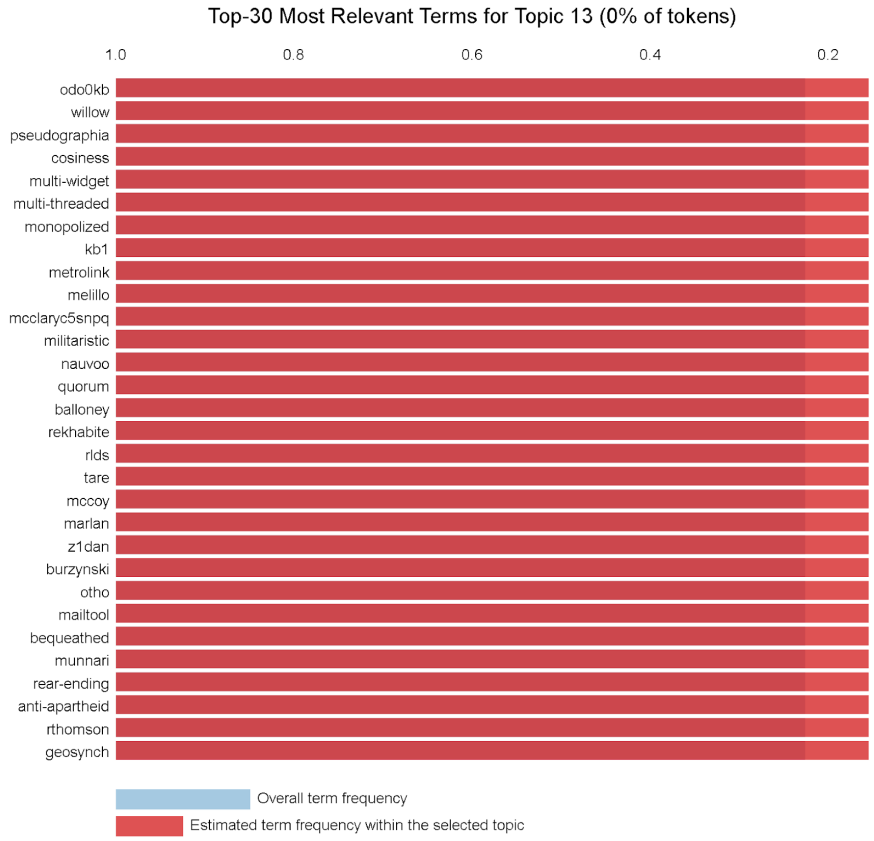

Top-30 Most Relevant Terms for Topic 8 (1.5% of tokens)
1.0
1.5
2.0
2.5
3.0
bi-metalic
imperative
pieced
load
juando
ls5_
jabbed
mathes
ebcdic
na61
kz1
pail
moronian
mick
g8u
separatley
vr6
quickturn
ldgo
rm4r50
r1u1
me6
sadness
tesiel
pragmatic
eunuch
cisler
ltcs
sky-cleaning
rogerk
Overall term frequency
Estimated term frequency within the selected topic

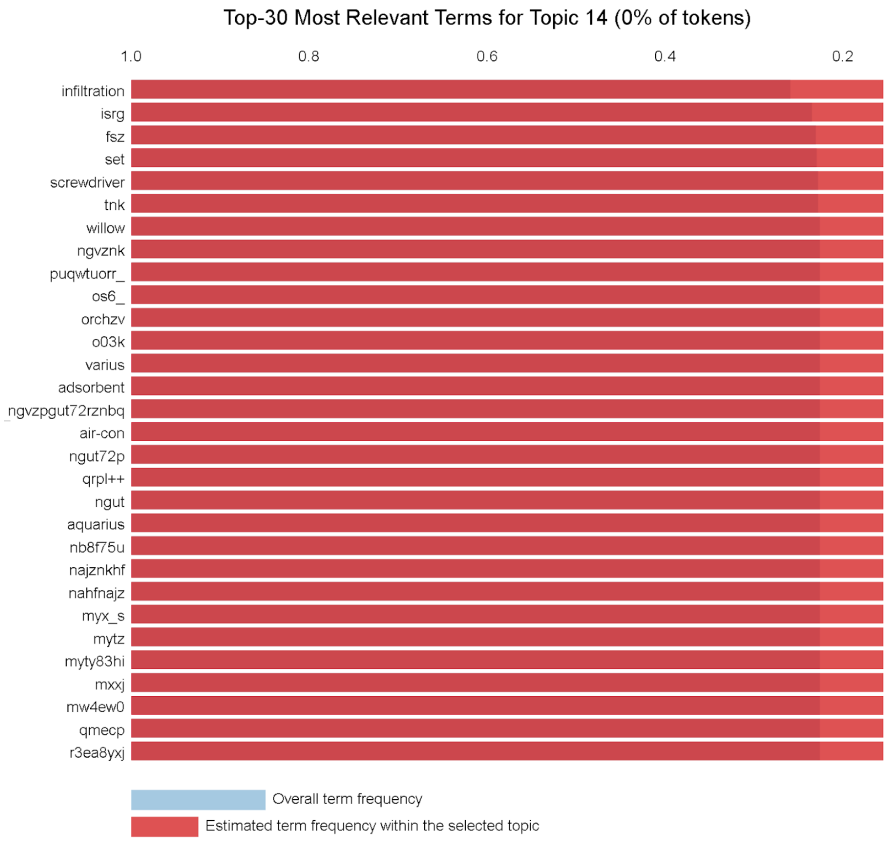

Top-30 Most Relevant Terms for Topic 9 (1.1% of tokens)
1.0
1.5
2.0
2.5
3.0
3.5
johng
appointee
maq046um
pgvz
high-security
microbiology
pay-off
capillary
equal
clipper-tap
s965
t58
mf59
dahlgren
wlist
mussolini
comming
icsi
muk
transit
discussed
etat
shaft-drive
decodable
hippolytus
touchup
appreciated
presumes
verticle
unfaithful
Overall term frequency
Estimated term frequency within the selected topic

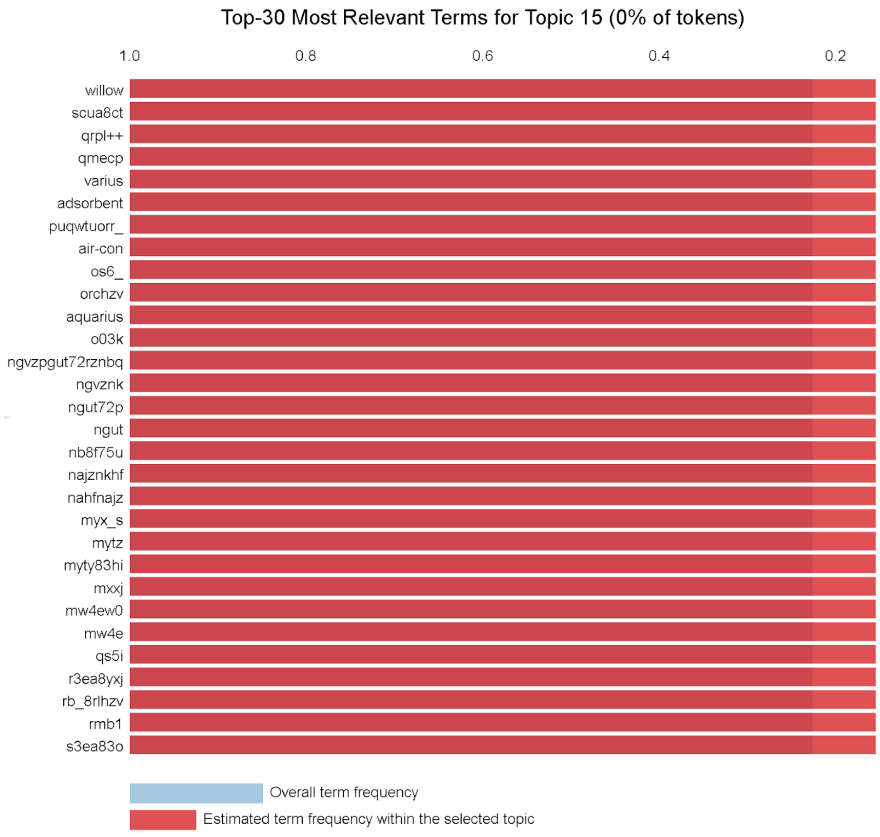

Top-30 Most Relevant Terms for Topic 10 (0.2% of tokens)
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
c9m_
rxxkc
ju10
paranormal
lakatos
starved
pinch
brady
anti-dive
reformation
mentioned
newdes
mishra
vrph
olympics
macedonia
third-parties
badouel
itseft
barkley
cek
benoit
cor
conform
commemorating
baptism
ccs-lib
annuls
dc2120
wetteland
Overall term frequency
Estimated term frequency within the selected topic

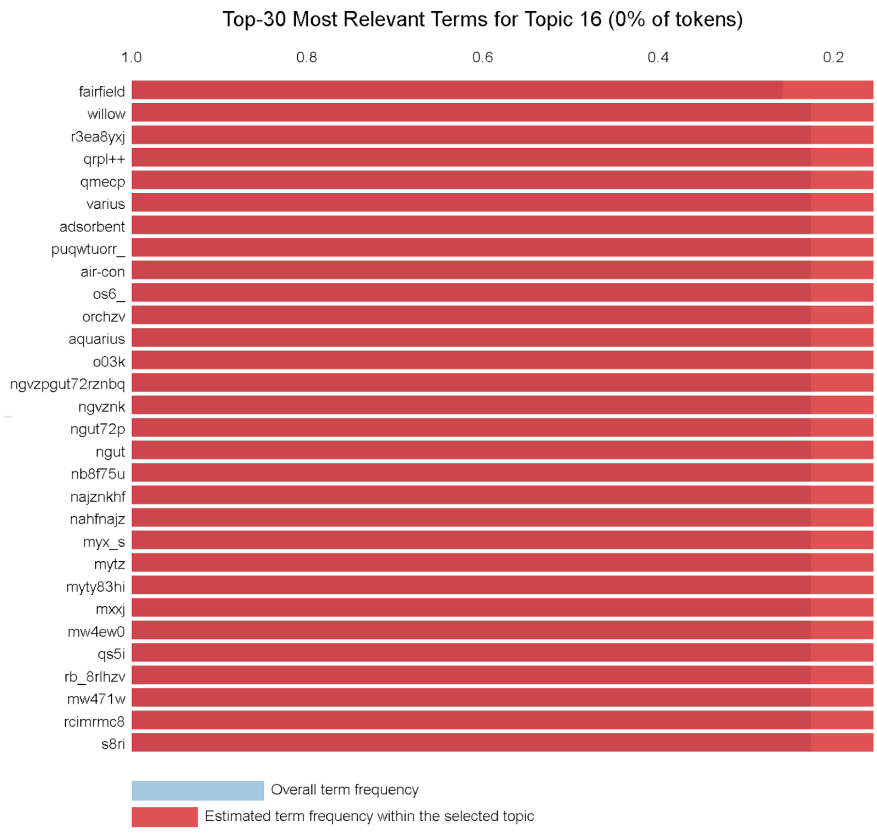

Top-30 Most Relevant Terms for Topic 11 (0.1% of tokens)
1.0
0.8
0.6
0.4
0.2
antibigot
willow
ngvzpgut72rznbq
qrpl++
qmecp
varius
adsorbent
puqwtuorr_
air-con
os6_
orchzv
aquarius
o03k
ngvznk
r3ea8yxj
ngut72p
ngut
nb8f75u
najznkhf
nahfnajz
myx_s
mytz
myty83hi
mxxj
mw4ew0
qs5i
rb_8rlhzv
mw471w
rmb1
s3ea83o
Overall term frequency
Estimated term frequency within the selected topic

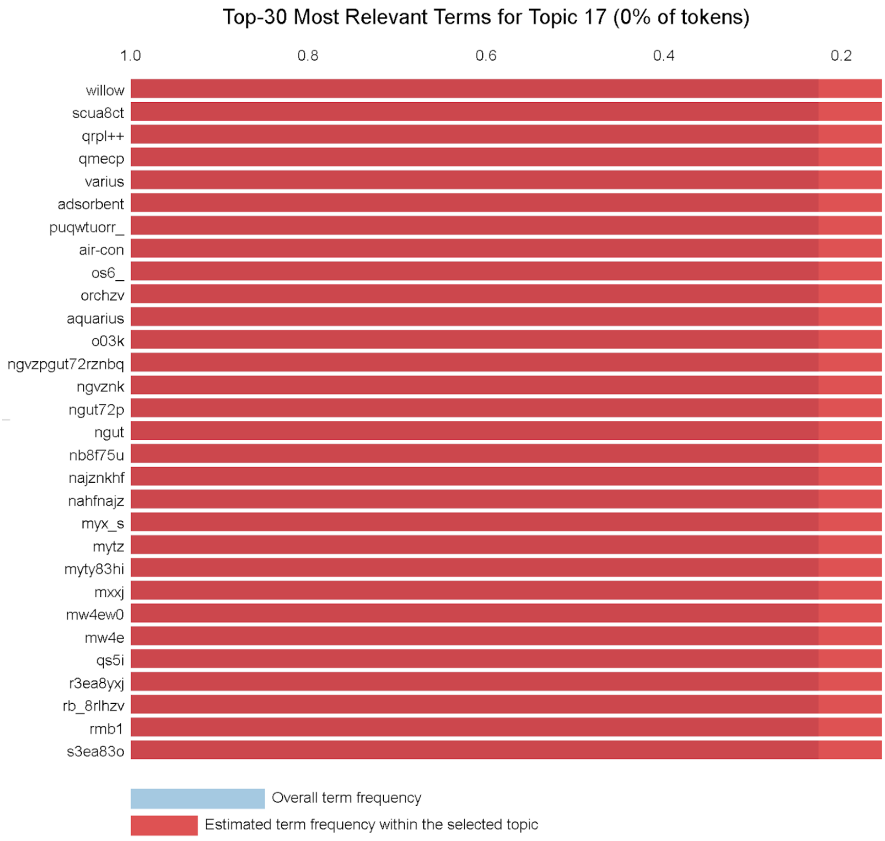

Top-30 Most Relevant Terms for Topic 12 (0% of tokens)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
montecito
plakken
irrelevant
capitol
bvfp
clling
ng2z
multitask
kbjo9
aquarius
os6_
orchzv
varius
adsorbent
o03k
air-con
ngvzpgut72rznbq
ngvznk
ngut72p
rnahfnkh
qmecp
topographic
ngut
nb8f75u
najznkhf
nahfnajz
myx_s
mytz
myty83hi
puqwtuorr_
Overall term frequency
Estimated term frequency within the selected topic

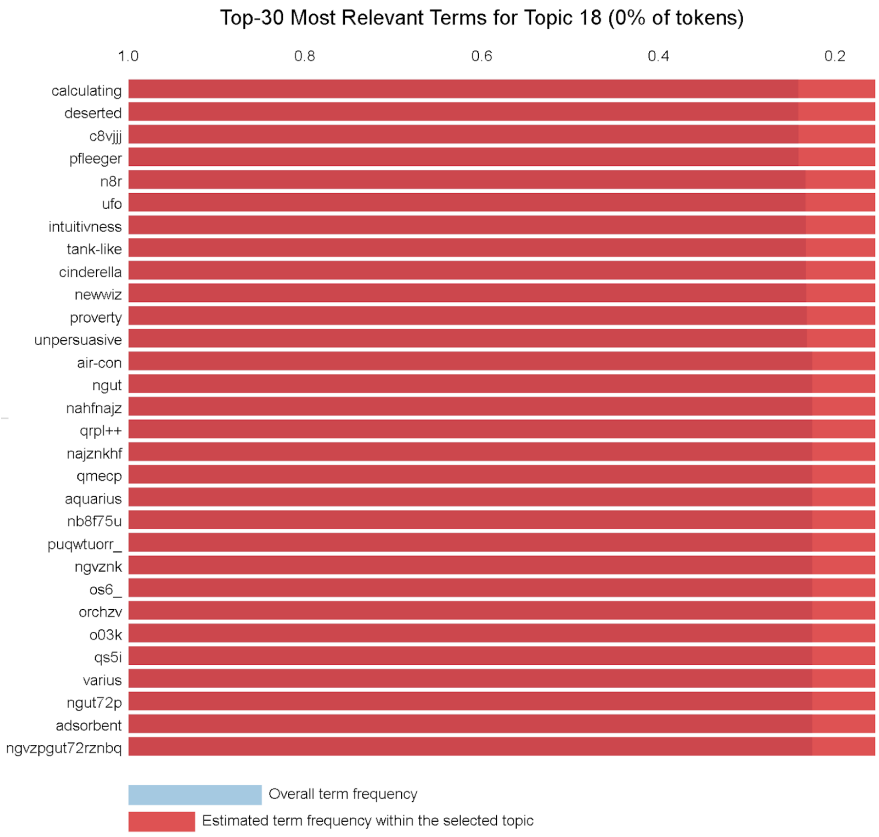

Top-30 Most Relevant Terms for Topic 13 (0% of tokens)
1.0
0.8
0.6
0.4
0.2
odo0kb
willow
pseudographia
cosiness
multi-widget
multi-threaded
monopolized
kb1
metrolink
melillo
mcclaryc5snpq
militaristic
nauvoo
quorum
balloney
rekhabite
rlds
tare
mccoy
marlan
z1dan
burzynski
otho
mailtool
bequeathed
munnari
rear-ending
anti-apartheid
rthomson
geosynch
Overall term frequency
Estimated term frequency within the selected topic

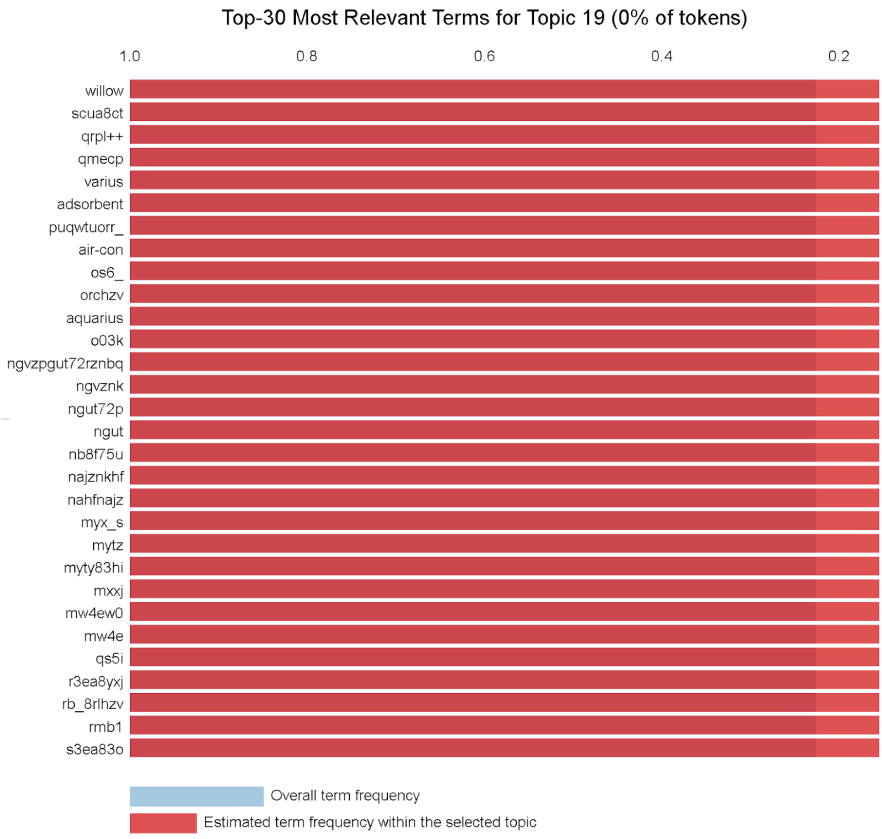

Top-30 Most Relevant Terms for Topic 14 (0% of tokens)
1.0
0.8
0.6
0.4
0.2
infiltration
isrg
fsz
set
screwdriver
tnk
willow
ngvznk
puqwtuorr_
os6_
orchzv
o03k
varius
adsorbent
ngvzpgut72rznbq
air-con
ngut72p
qrpl++
ngut
aquarius
nb8f75u
najznkhf
nahfnajz
myx_s
mytz
myty83hi
mxxj
mw4ew0
qmecp
r3ea8yxj
Overall term frequency
Estimated term frequency within the selected topic

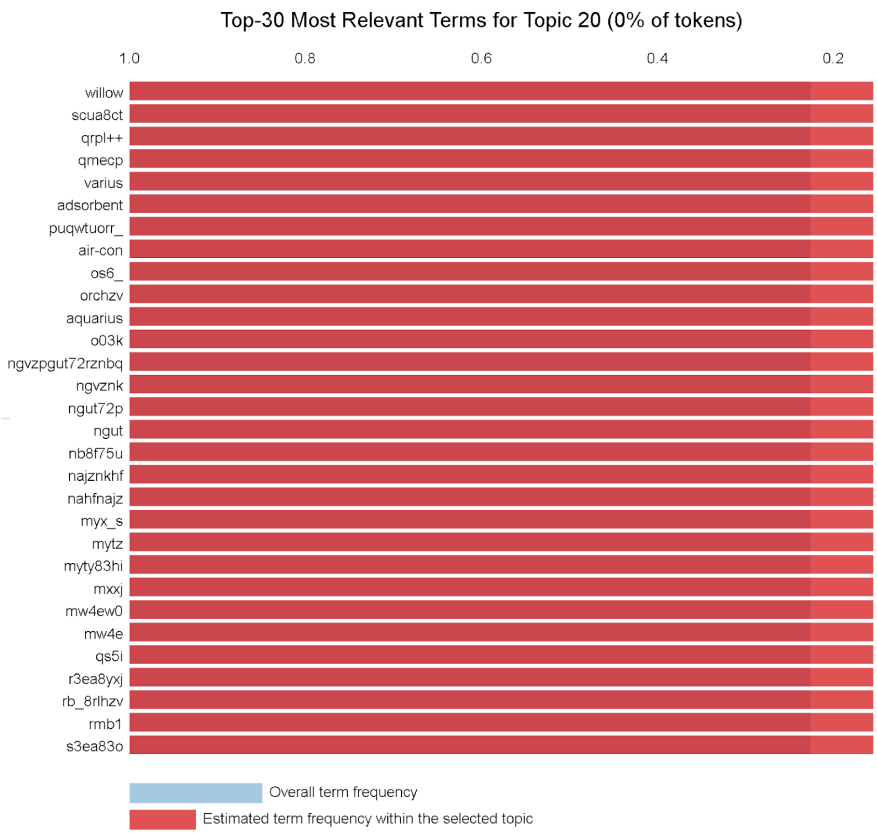

Top-30 Most Relevant Terms for Topic 15 (0% of tokens)
1.0
0.8
0.6
0.4
0.2
willow
scua8ct
qrpl++
qmecp
varius
adsorbent
puqwtuorr_
air-con
os6_
orchzv
aquarius
o03k
ngvzpgut72rznbq
ngvznk
ngut72p
ngut
nb8f75u
najznkhf
nahfnajz
myx_s
mytz
myty83hi
mxxj
mw4ew0
mw4e
qs5i
r3ea8yxj
rb_8rlhzv
rmb1
s3ea83o
Overall term frequency
Estimated term frequency within the selected topic

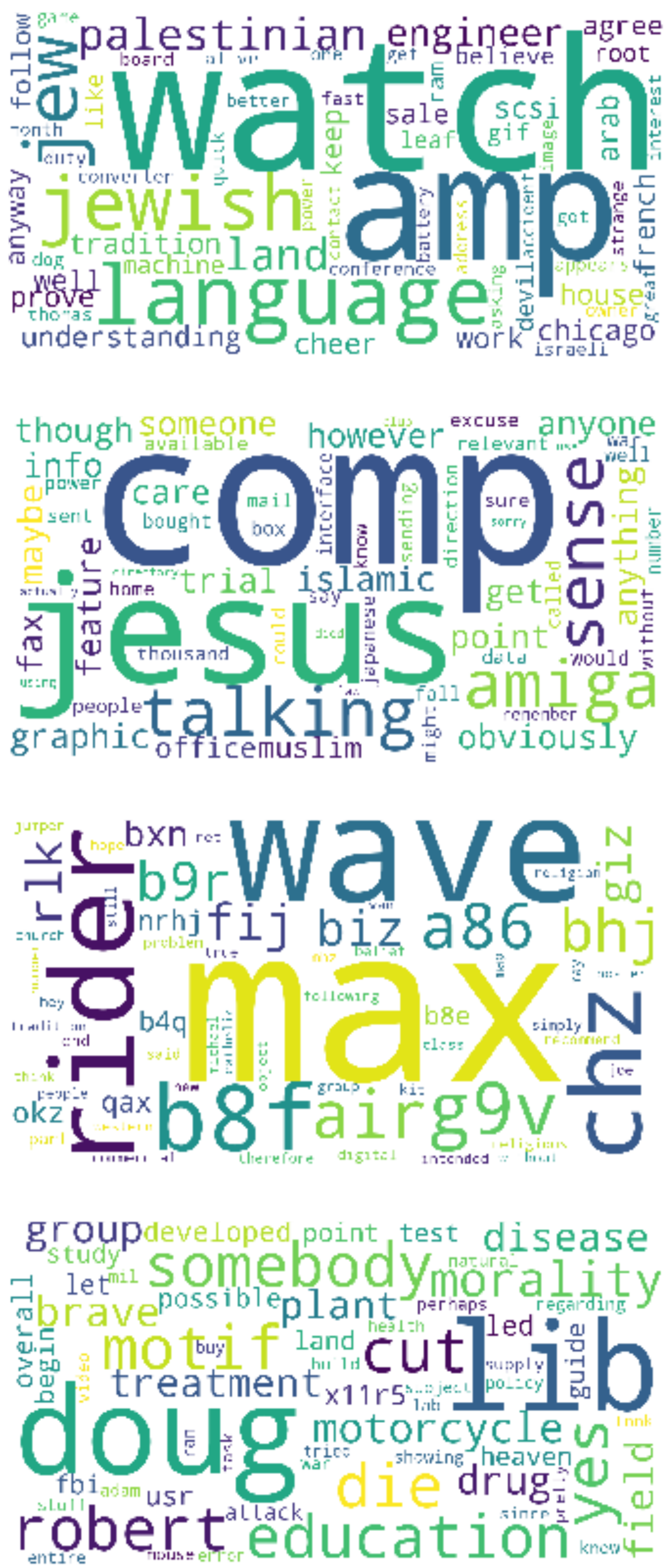

Top-30 Most Relevant Terms for Topic 16 (0% of tokens)
1.0
0.8
0.6
0.4
0.2
fairfield
willow
r3ea8yxj
qrpl++
qmecp
varius
adsorbent
puqwtuorr_
air-con
os6_
orchzv
aquarius
o03k
ngvzpgut72rznbq
ngvznk
ngut72p
ngut
nb8f75u
najznkhf
nahfnajz
myx_s
mytz
myty83hi
mxxj
mw4ew0
qs5i
rb_8rlhzv
mw471w
rcimrmc8
s8ri
Overall term frequency
Estimated term frequency within the selected topic

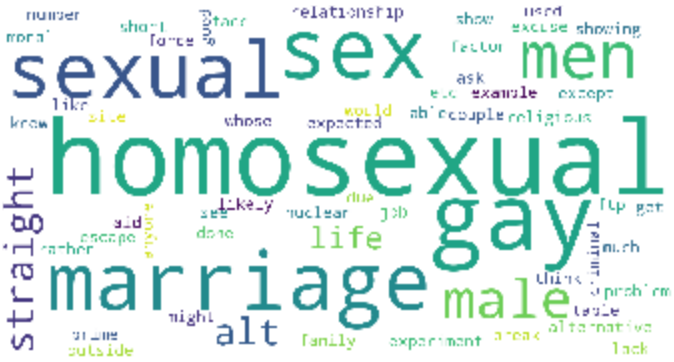

Top-30 Most Relevant Terms for Topic 17 (0% of tokens)
1.0
0.8
0.6
0.4
0.2
willow
scua8ct
qrpl++
qmecp
varius
adsorbent
puqwtuorr_
air-con
os6_
orchzv
aquarius
o03k
ngvzpgut72rznbq
ngvznk
ngut72p
ngut
nb8f75u
najznkhf
nahfnajz
myx_s
mytz
myty83hi
mxxj
mw4ew0
mw4e
qs5i
r3ea8yxj
rb_8rlhzv
rmb1
s3ea83o
Overall term frequency
Estimated term frequency within the selected topic


Top-30 Most Relevant Terms for Topic 18 (0% of tokens)
1.0
0.8
0.6
0.4
0.2
calculating
deserted
c8vjjj
pfleeger
n8r
ufo
intuitivness
tank-like
cinderella
newwiz
proverty
unpersuasive
air-con
ngut
nahfnajz
qrpl++
najznkhf
qmecp
aquarius
nb8f75u
puqwtuorr_
ngvznk
os6_
orchzv
o03k
qs5i
varius
ngut72p
adsorbent
ngvzpgut72rznbq
Overall term frequency
Estimated term frequency within the selected topic


Top-30 Most Relevant Terms for Topic 19 (0% of tokens)
1.0
0.8
0.6
0.4
0.2
willow
scua8ct
qrpl++
qmecp
varius
adsorbent
puqwtuorr_
air-con
os6_
orchzv
aquarius
o03k
ngvzpgut72rznbq
ngvznk
ngut72p
ngut
nb8f75u
najznkhf
nahfnajz
myx_s
mytz
myty83hi
mxxj
mw4ew0
mw4e
qs5i
r3ea8yxj
rb_8rlhzv
rmb1
s3ea83o
Overall term frequency
Estimated term frequency within the selected topic

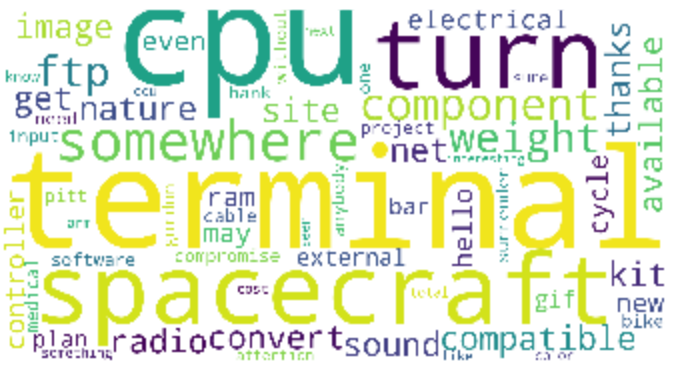

Top-30 Most Relevant Terms for Topic 20 (0% of tokens)
1.0
0.8
0.6
0.4
0.2
willow
scua8ct
qrpl++
qmecp
varius
adsorbent
puqwtuorr_
air-con
os6_
orchzv
aquarius
o03k
ngvzpgut72rznbq
ngvznk
ngut72p
ngut
nb8f75u
najznkhf
nahfnajz
myx_s
mytz
myty83hi
mxxj
mw4ew0
mw4e
qs5i
r3ea8yxj
rb_8rlhzv
rmb1
s3ea83o
Overall term frequency
Estimated term frequency within the selected topic


watch
amp
language
jewish
jew
palestinian
engineer
agree
root
believe
land
tradition
arab
french
house
chicago
understanding
cheer
work
israeli
machine
prove
well
anyway
follow
like
keep
sale
scsi
gif
leaf
better
fast
board
conference
devil
strange
comp
jesus
talking
sense
amiga
anything
feature
graphic
officemuslim
obviously
point
islamic
trial
care
info
maybe
fax
though
someone
available
however
anyone
excuse
relevant
interface
direction
bought
box
mail
sure
thousand
people
home
power
sent
get
would
without
number
data
max
wave
rider
chz
b8fairg9v
a86
bhj
biz
fij
b9r
rlk
bxn
nrhj
giz
b4q
b8e
qax
okz
doug
lib
education
robert
motorcycle
somebody
morality
motif
cut
die
drug
yes
field
treatment
plant
brave
group
developed
point
test
disease
study
let
possible
land
x11r5
heaven
fbi
usr
attack
guide
overall
begin
led
perhaps
regarding
health
buy
supply
policy
showing
entire

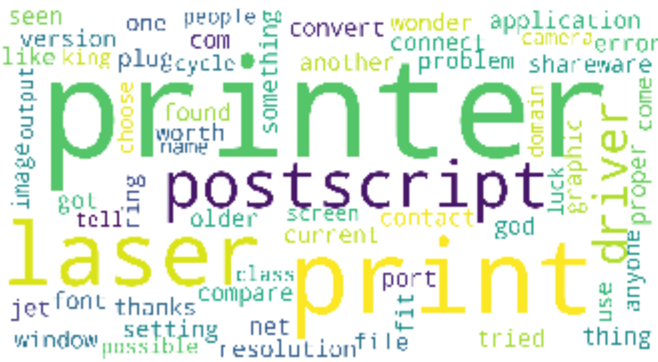

sexual
sex
men
homosexual
straight
gay
marriage
male
life
alt

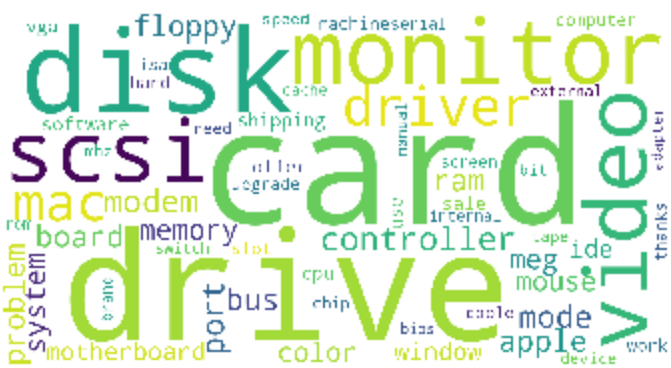

key
chip
encryption
escrow
algorithm
government
clipper
nsa
phone
secure
crypto
encrypted
warrant
wiretap
secret
system
pgp


armenia
genocide
germany
turkish
serdar
massacre
republic
greek
nazi
turkey
armenian
russian
killed
muslim
turk
village
azerbaijan
soldier
population

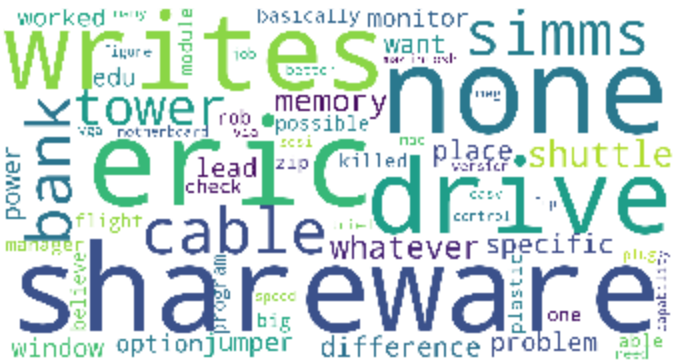

cpu
turn
image
electrical
ftp
component
somewhere
weight
terminal
spacecraft
radioconvert
compatible
sound
kit
thanks
available

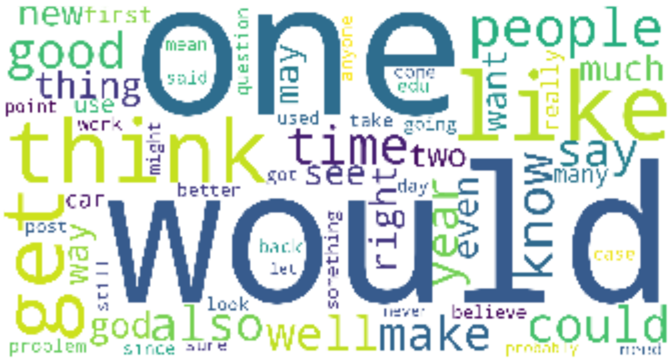

window
file
thanks
please
program
mail
advance
graphic
email
version
information
know
looking
anyone
ftp
server
address
edu
directory
anybody
image
help
info
site
use
using
get
available
code
com
software
appreciated
format

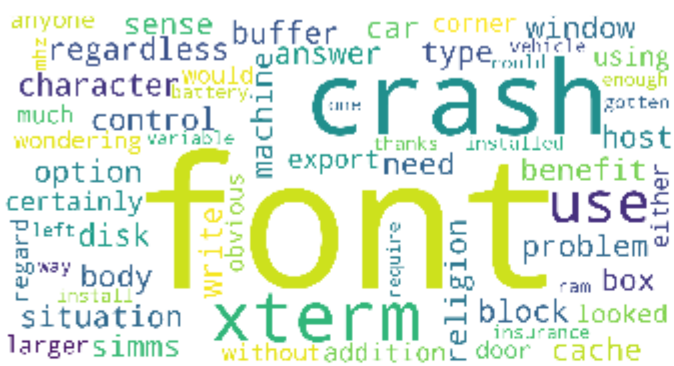

printer
postscript
print
laser
driver
seen
version
one
people
convert
wonder
application
connect
error
like
plug
cycle
another
problem
shareware
something
found
worth
output
image
choose
domain
graphic
come
proper
older
screen
contact
current
god
luck
anyone
class
port
compare
font
thanks
jet
setting
net
window
possible
resolution
file
tried
thing
use
fit

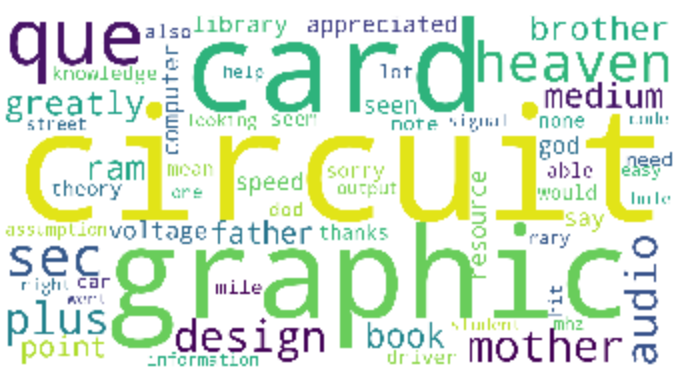

disk
monitor
card
drive
scsi
video
driver
floppy
mac
modem
memory
board
controller
ram
problem
system
port
bus
mode
apple
mouse
ide
meg
motherboard
color
window
software
shipping
computer
screen
sale
thanks


test
message
information
nice
reality
real
quality
post
number
universe
like
larry
third
see
king
get
iii
phone
address
certain
chip
engine
concept
guess
know
politics
talking
configuration
explain
little
sell
consideration
virtual
wish
compound
nothing
thing
agree
fired
turbo
obvious
great
ibm
guy
bug
contract
large


shareware
eric
drive
none
writes
bank
tower
cable
simms
shuttle
memory
monitor
worked
whatever
specific
difference
problem
window
option
jumper
place
possible
power
lead
check
killed
want
flight

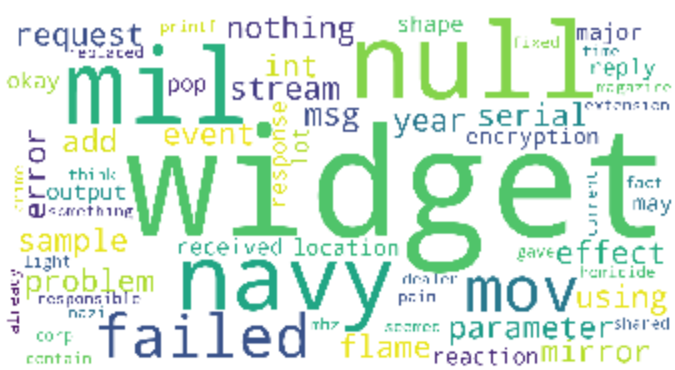

would
one
think
like
get
people
good
thing
time
know
year
right
even
say
much
make
could
well
also
god
see
two
new
first
may
want
many
way


font
crash
xterm
use
machine
religion
write
regardless
character
control
buffer
sense
answer
type
window
host
benefit
problem
box
block
looked
situation
simms
addition
cache
option
certainly
disk
body
export
need

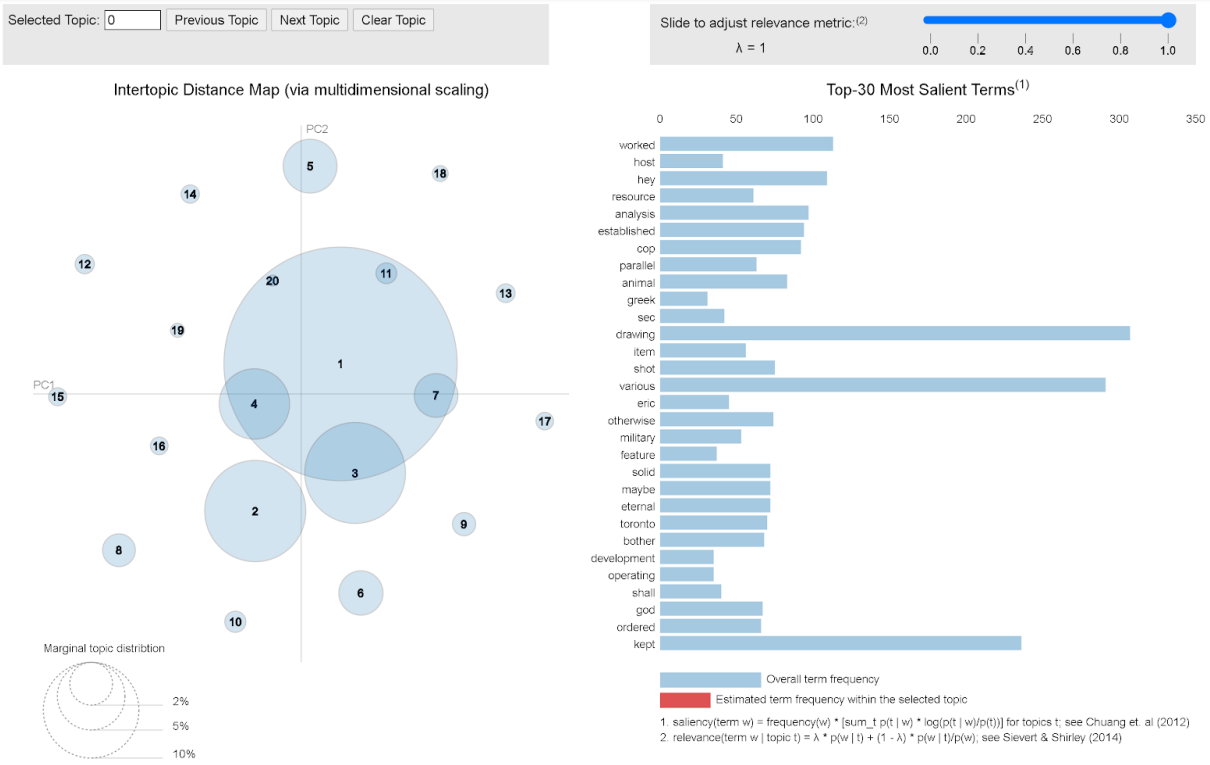

circuit
graphic
card
que
heaven
brother
medium
greatly
sec
plus
design
mother
book
audio
father
voltage
ram
speed
resource
appreciated
library

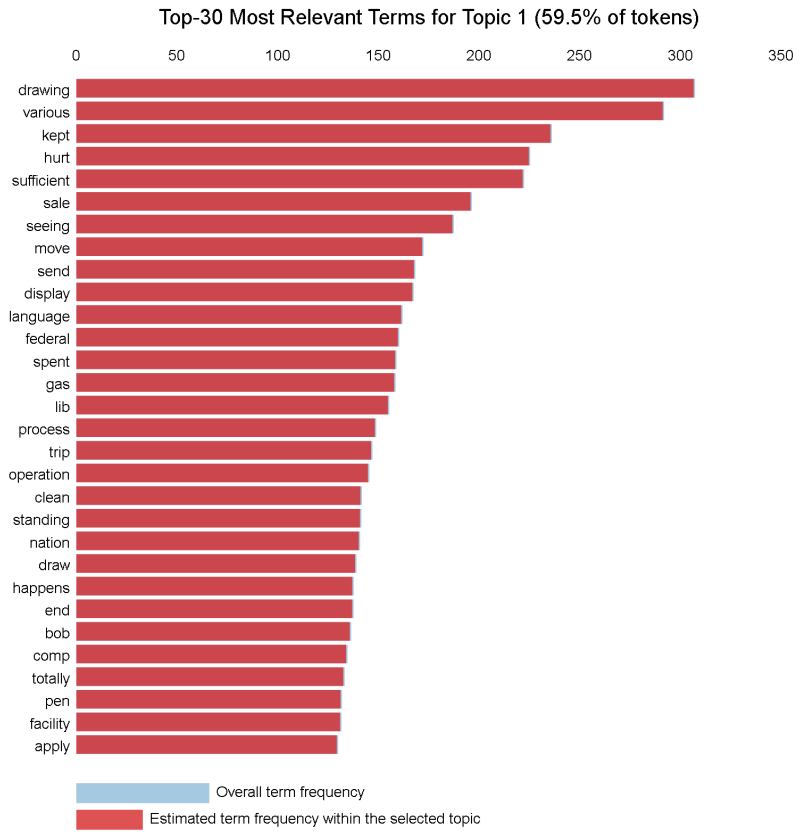

game
team
player
season
play
win
nhl
playoff
chicago
league
toronto
ranger
goal
year
wing
detroit
leaf
cub
pitcher
pen
score
last
boston
montreal
division
pittsburgh
flyer
ice
record
red
final
played
series
stats
pick
minnesota
louis
first
period

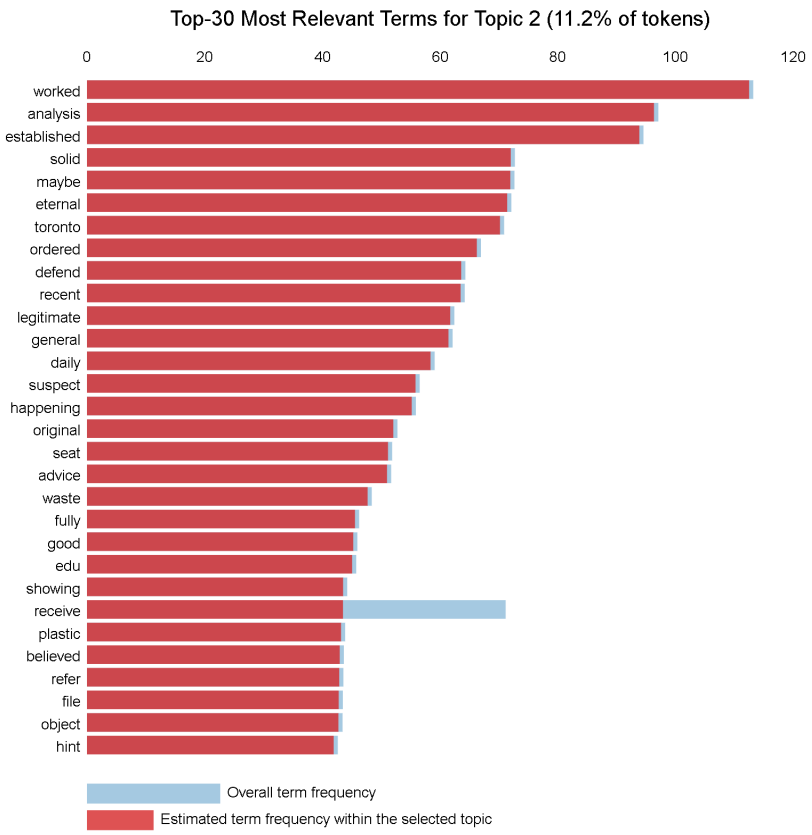

cycle
loaded
released
product
insurance
word
child
bob
access
dog
post
always
screen
based
window
two
style
could
enough
call
time
response
case
mark
boy
single
boot
john
org
matthew
data
yeah
description
telling
would
paper
three
article
baseball
missing
ask

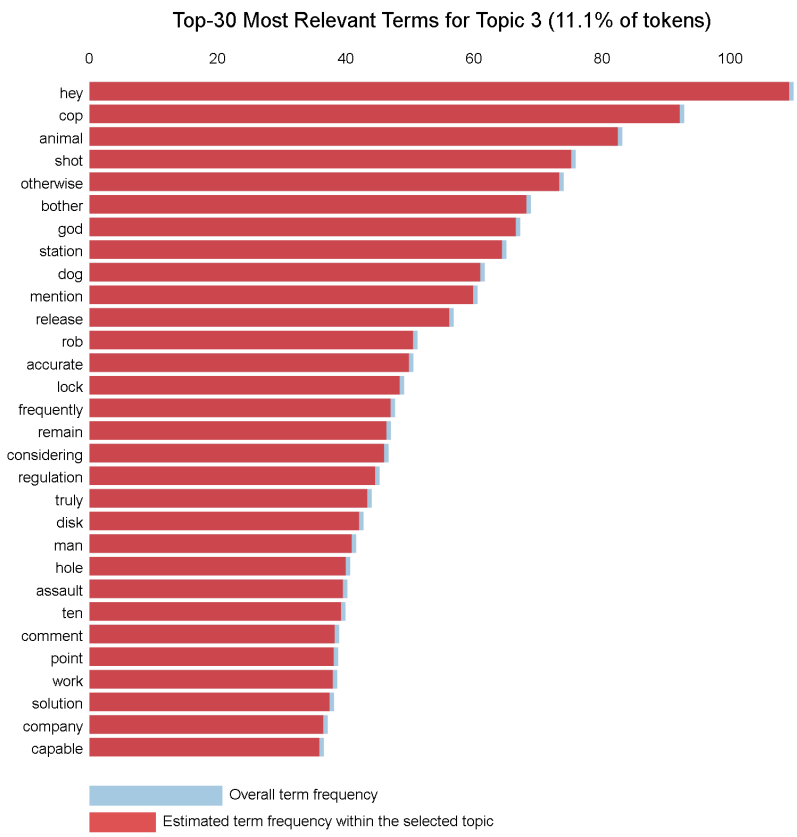

widget
null
mil
navy
mov
failed
request
nothing
shape
major
stream
int
msg
year
serial
encryption
event
add
error
output
sample
received
location
effect
problem
using
parameter
flame
reaction
mirror
reply
pop
okay
fact
may
response
lot

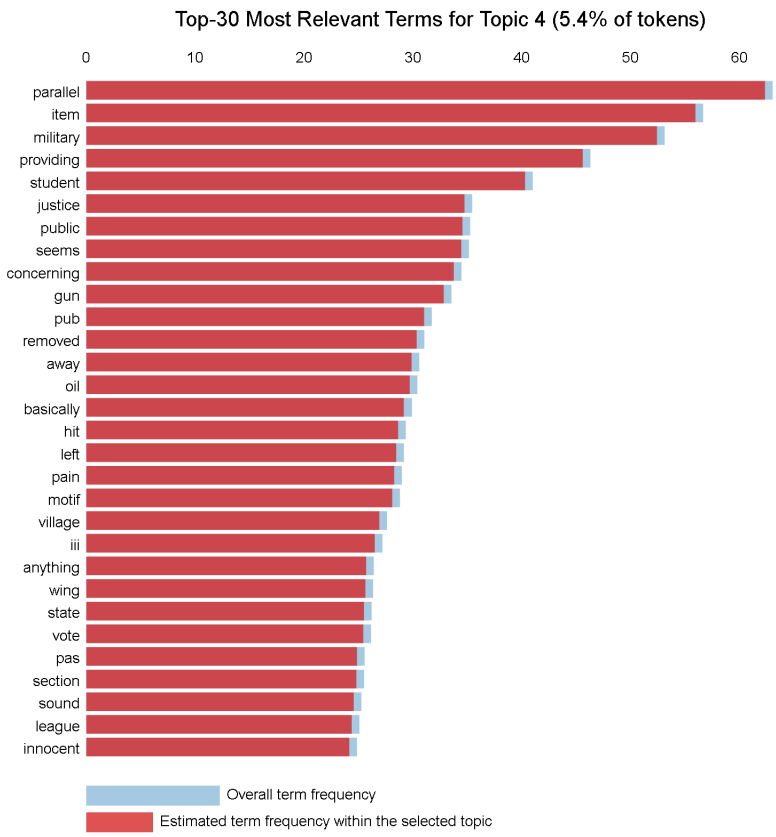

gordon
surrender
pitt
bank
edu
soon
yeah
patient
hospital
effective
blood
pressure
usually
nothing

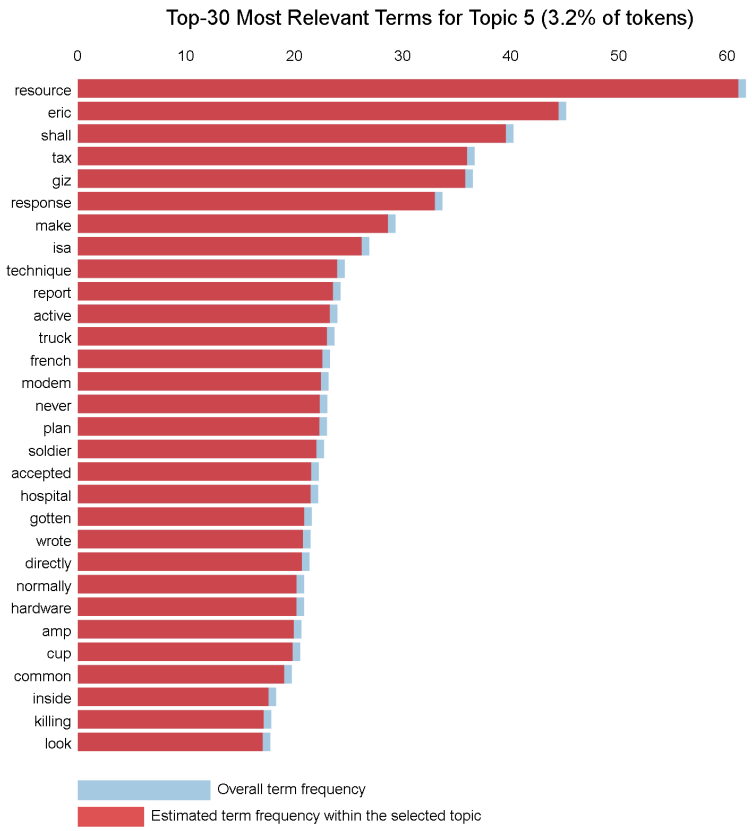

Selected Topic: 0
Previous Topic
Next Topic
Clear Topic
Slide to adjust relevance metric:(2)
λ = 1
0.0
0.2
0.4
0.6
0.8
1.0
Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
Marginal topic distribtion
2%
5%
10%
Top-30 Most Salient Terms(1)
worked
host
hey
resource
analysis
established
cop
parallel
animal
greek
sec
drawing
item
shot
various
eric
otherwise
military
feature
solid
maybe
eternal
toronto
bother
development
operating
shall
god
ordered
kept
Overall term frequency
Estimated term frequency within the selected topic
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

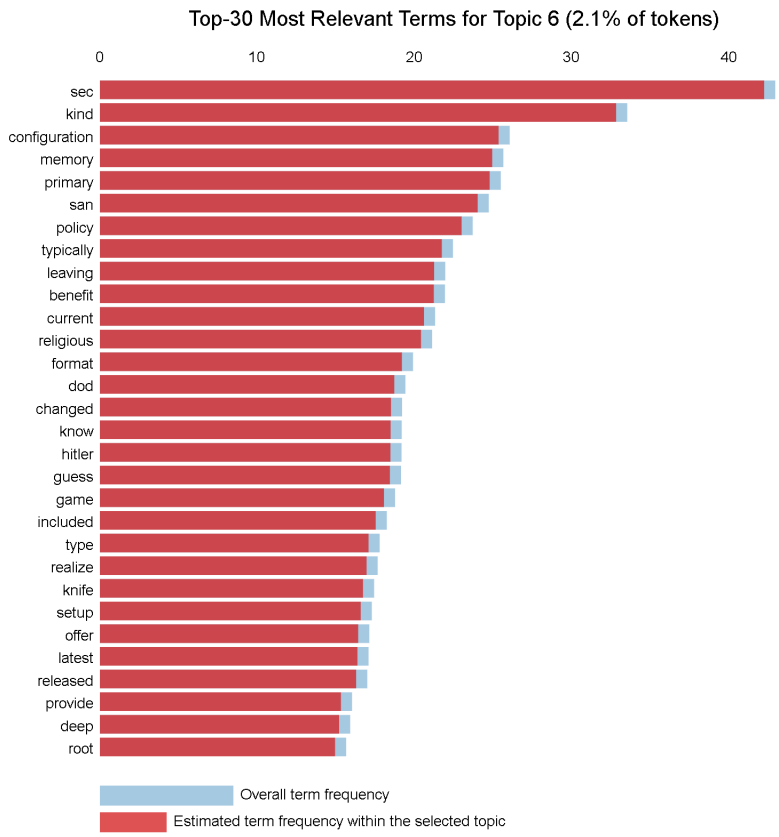

Top-30 Most Relevant Terms for Topic 1 (59.5% of tokens)
drawing
various
kept
hurt
sufficient
sale
seeing
move
send
display
language
federal
spent
gas
lib
process
trip
operation
clean
standing
nation
draw
happens
end
bob
comp
totally
pen
facility
apply
Overall term frequency
Estimated term frequency within the selected topic

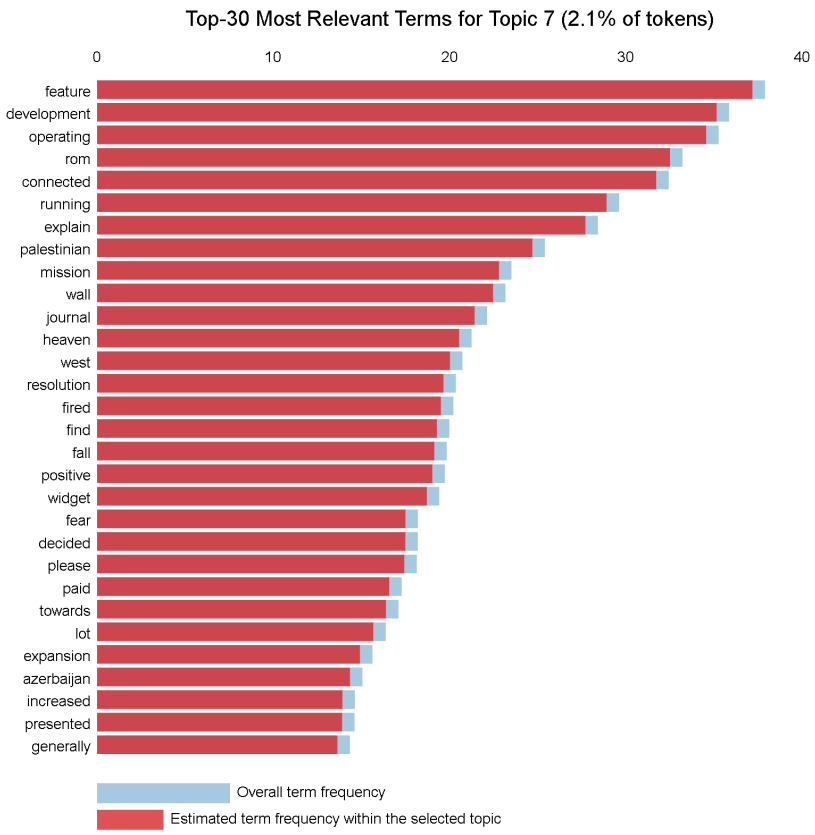

Top-30 Most Relevant Terms for Topic 2 (11.2% of tokens)
0
20
40
60
80
100
120
worked
analysis
established
solid
maybe
eternal
toronto
ordered
defend
recent
legitimate
general
daily
suspect
happening
original
seat
advice
waste
fully
good
edu
showing
receive
plastic
believed
refer
file
object
hint
Overall term frequency
Estimated term frequency within the selected topic

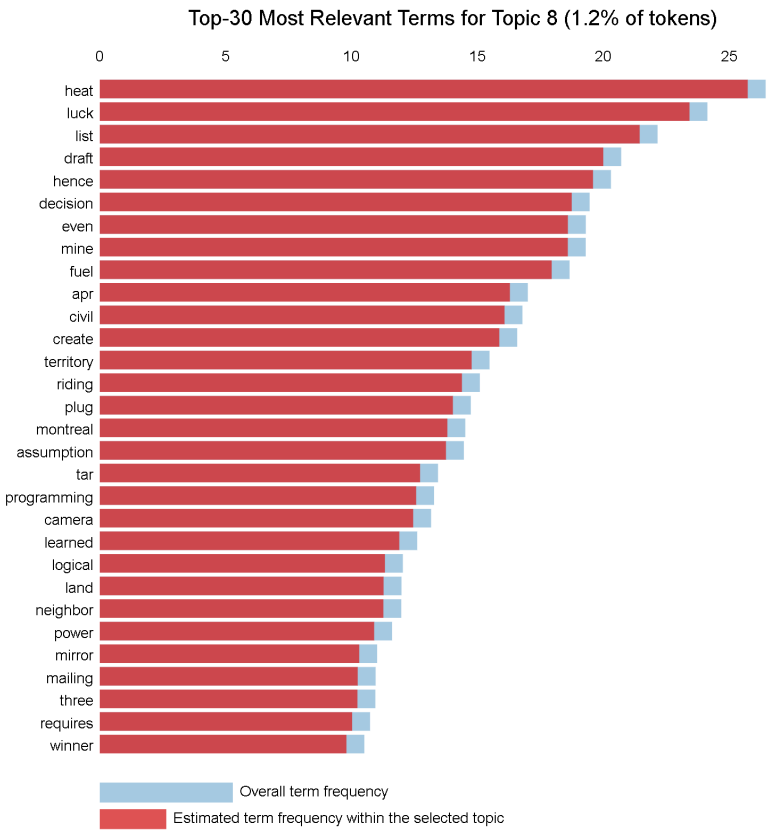

Top-30 Most Relevant Terms for Topic 3 (11.1% of tokens)
0
20
40
60
80
100
hey
cop
animal
shot
otherwise
bother
god
station
dog
mention
release
rob
accurate
lock
frequently
remain
considering
regulation
truly
disk
man
hole
assault
ten
comment
point
work
solution
company
capable
Overall term frequency
Estimated term frequency within the selected topic

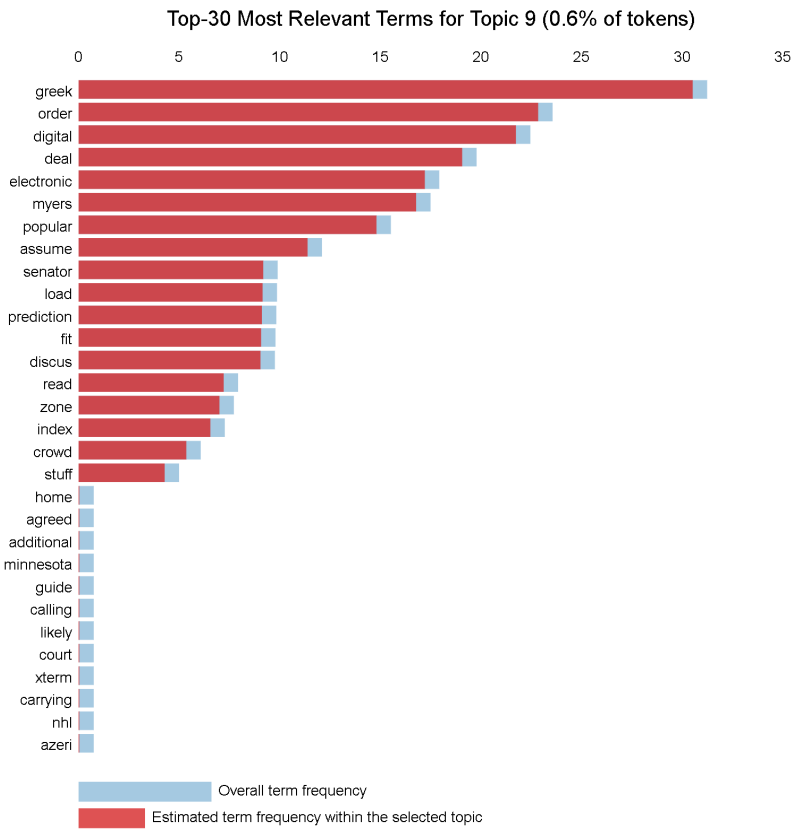

Top-30 Most Relevant Terms for Topic 4 (5.4% of tokens)
0
10
20
30
40
50
60
parallel
item
military
providing
student
justice
public
seems
concerning
gun
pub
removed
away
oil
basically
hit
left
pain
motif
village
iii
anything
wing
state
vote
pas
section
sound
league
innocent
Overall term frequency
Estimated term frequency within the selected topic

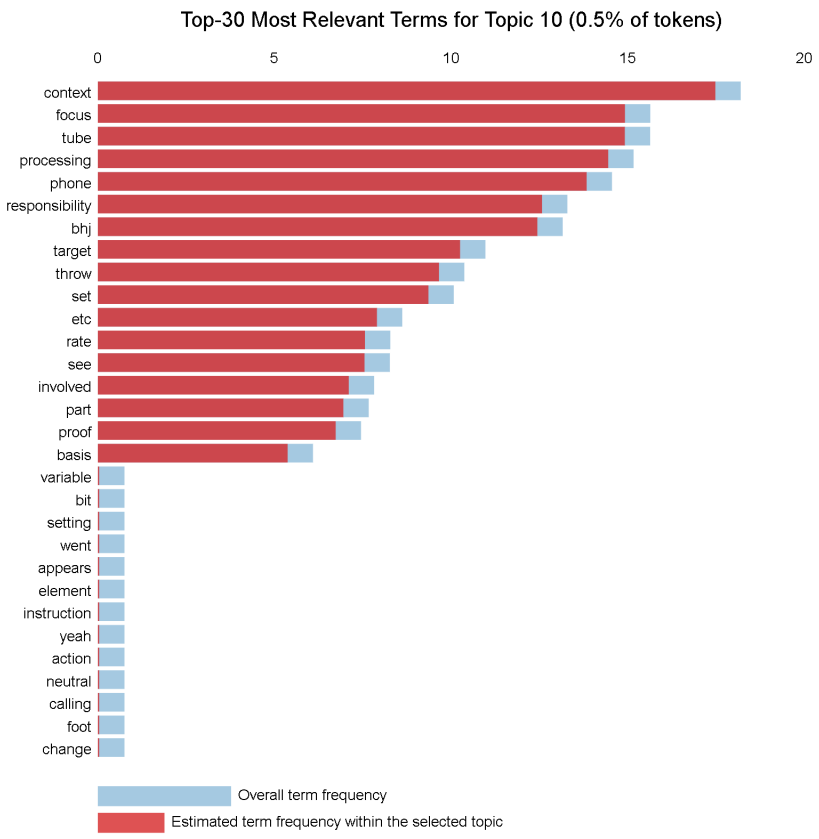

Top-30 Most Relevant Terms for Topic 5 (3.2% of tokens)
0
10
20
30
40
50
60
resource
eric
shall
tax
giz
response
make
isa
technique
report
active
truck
french
modem
never
plan
soldier
accepted
hospital
gotten
wrote
directly
normally
hardware
amp
cup
common
inside
killing
look
Overall term frequency
Estimated term frequency within the selected topic

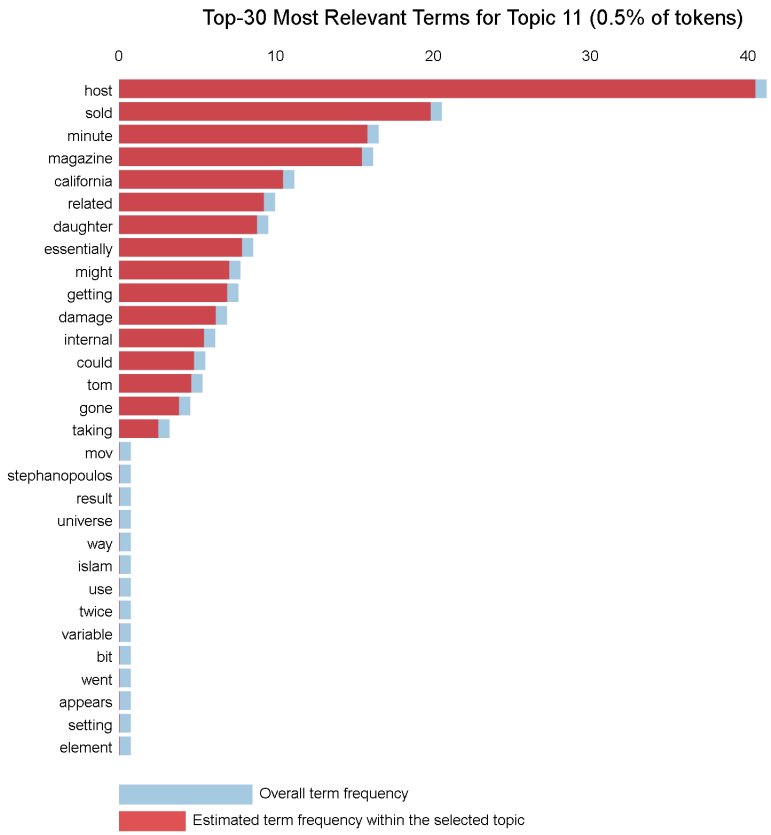

Top-30 Most Relevant Terms for Topic 6 (2.1% of tokens)
0
10
20
30
40
sec
kind
configuration
memory
primary
san
policy
typically
leaving
benefit
current
religious
format
dod
changed
know
hitler
guess
game
included
type
realize
knife
setup
offer
latest
released
provide
deep
root
Overall term frequency
Estimated term frequency within the selected topic

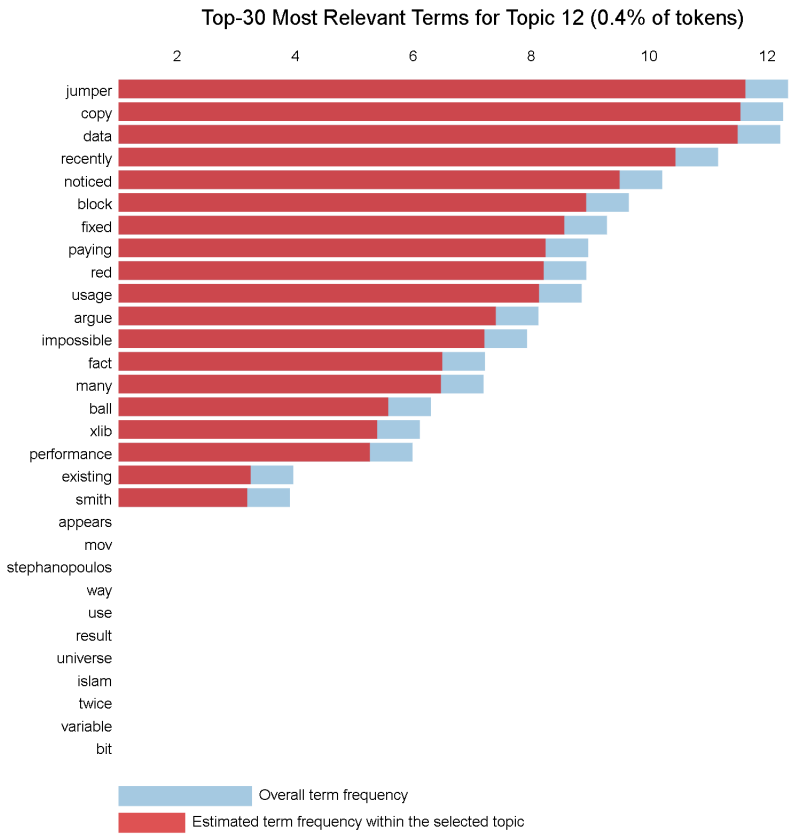

Top-30 Most Relevant Terms for Topic 7 (2.1% of tokens)
0
10
20
30
40
feature
development
operating
rom
connected
running
explain
palestinian
mission
wall
journal
heaven
west
resolution
fired
find
fall
positive
widget
fear
decided
please
paid
towards
lot
expansion
azerbaijan
increased
presented
generally
Overall term frequency
Estimated term frequency within the selected topic

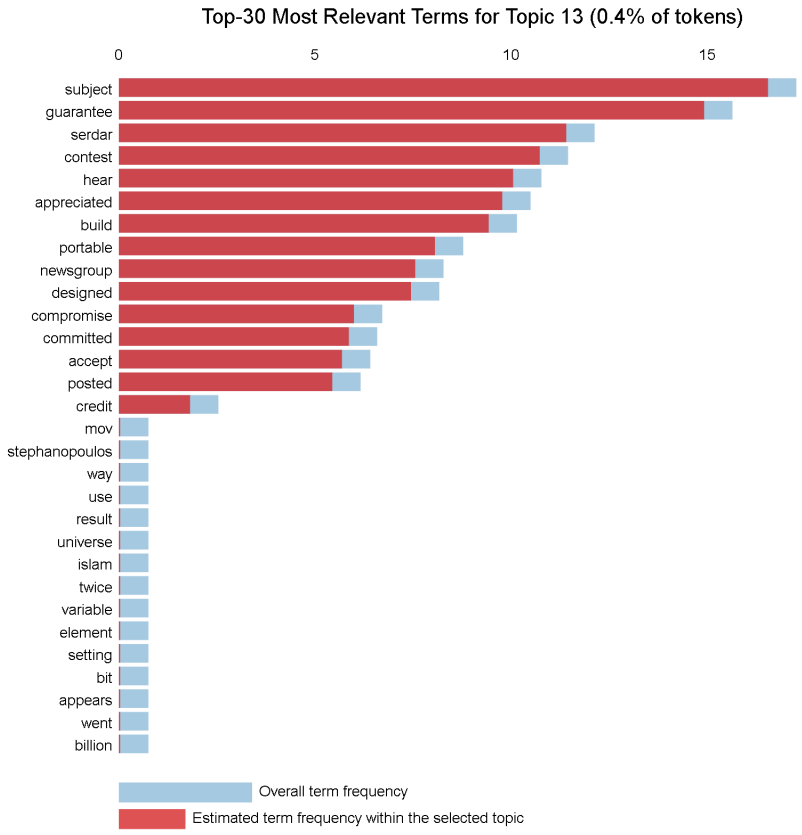

Top-30 Most Relevant Terms for Topic 8 (1.2% of tokens)
0
5
10
15
20
25
heat
luck
list
draft
hence
decision
even
mine
fuel
apr
civil
create
territory
riding
plug
montreal
assumption
tar
programming
camera
learned
logical
land
neighbor
power
mirror
mailing
three
requires
winner
Overall term frequency
Estimated term frequency within the selected topic

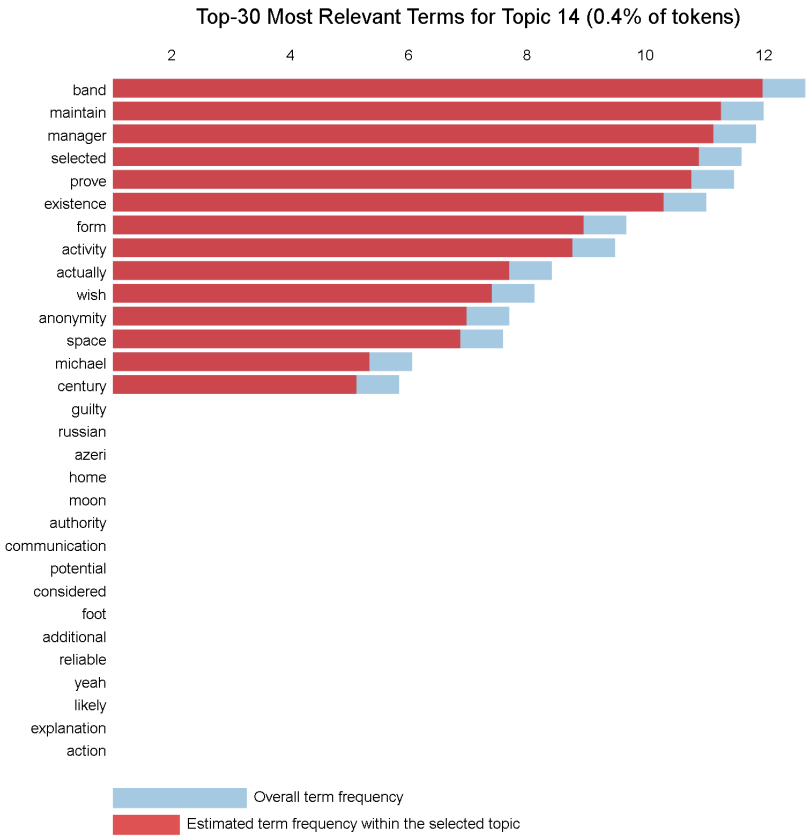

Top-30 Most Relevant Terms for Topic 9 (0.6% of tokens)
0
5
10
15
20
25
30
35
greek
order
digital
deal
electronic
myers
popular
assume
senator
load
prediction
fit
discus
read
zone
index
crowd
stuff
home
agreed
additional
minnesota
guide
calling
likely
court
xterm
carrying
nhl
azeri
Overall term frequency
Estimated term frequency within the selected topic

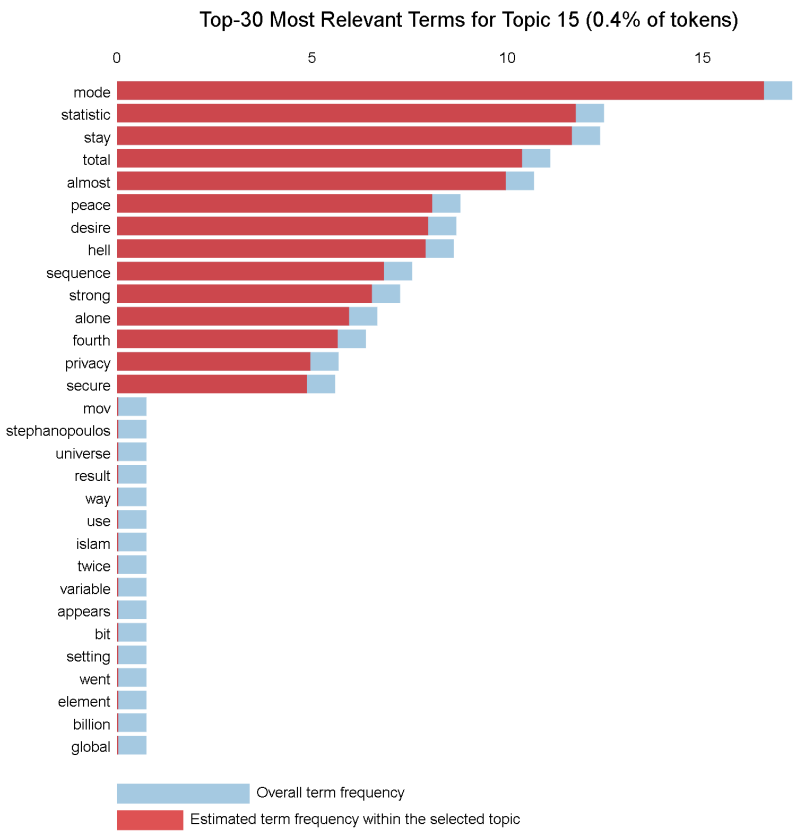

Top-30 Most Relevant Terms for Topic 10 (0.5% of tokens)
0
5
10
15
20
context
focus
tube
processing
phone
responsibility
bhj
target
throw
set
etc
rate
see
involved
part
proof
basis
variable
bit
setting
went
appears
element
instruction
yeah
action
neutral
calling
foot
change
Overall term frequency
Estimated term frequency within the selected topic

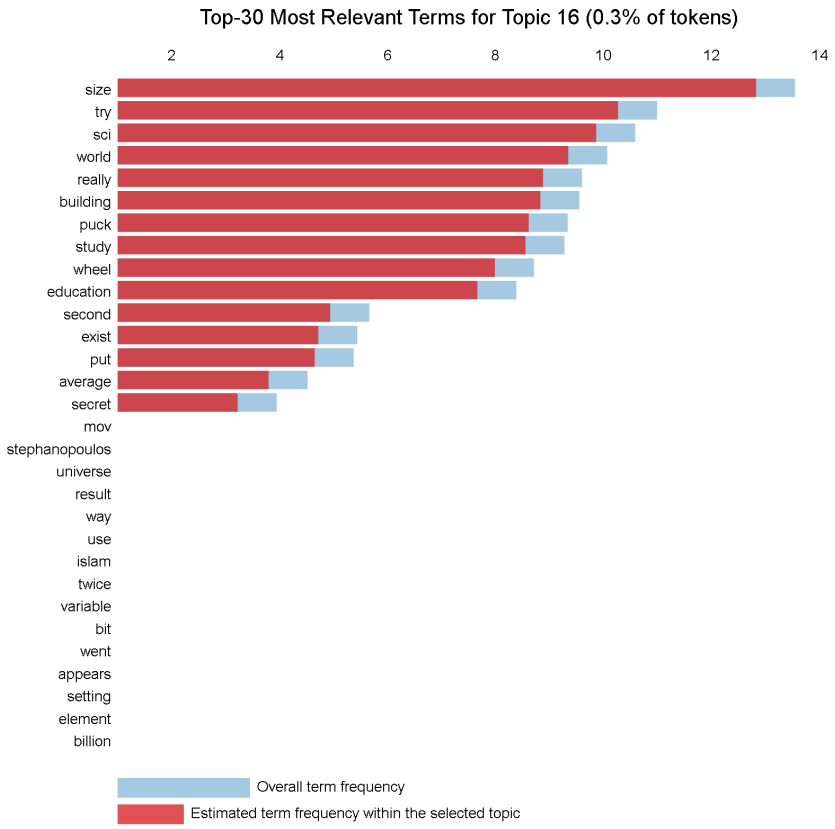

Top-30 Most Relevant Terms for Topic 11 (0.5% of tokens)
0
10
20
30
40
host
sold
minute
magazine
california
related
daughter
essentially
might
getting
damage
internal
could
tom
gone
taking
mov
stephanopoulos
result
universe
way
islam
use
twice
variable
bit
went
appears
setting
element
Overall term frequency
Estimated term frequency within the selected topic

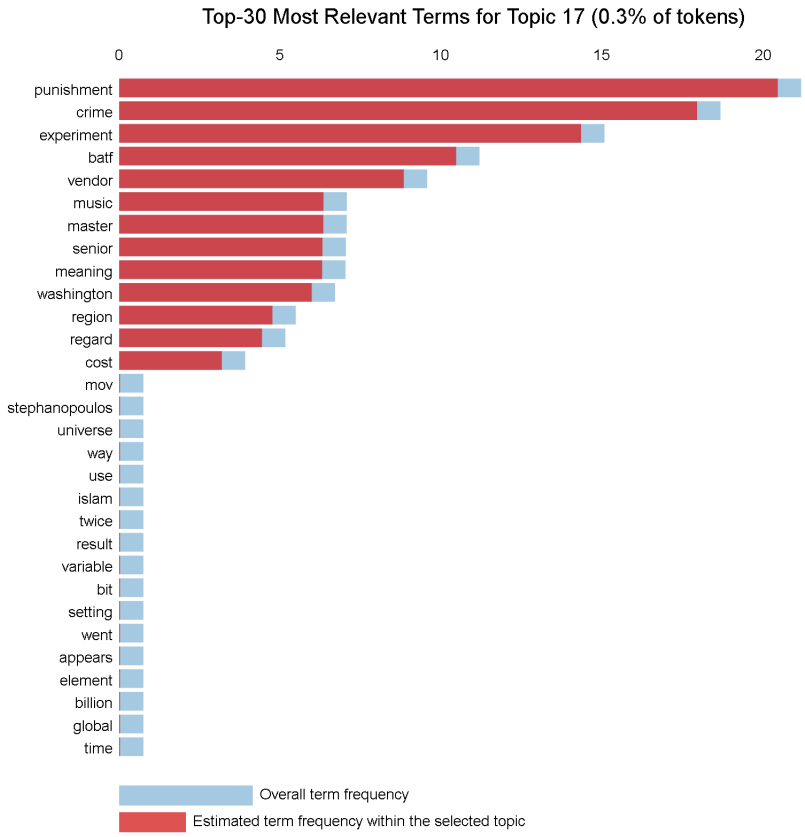

Top-30 Most Relevant Terms for Topic 12 (0.4% of tokens)
2
4
6
8
10
12
jumper
copy
data
recently
noticed
block
fixed
paying
red
usage
argue
impossible
fact
many
ball
xlib
performance
existing
smith
appears
mov
stephanopoulos
way
use
result
universe
islam
twice
variable
bit
Overall term frequency
Estimated term frequency within the selected topic

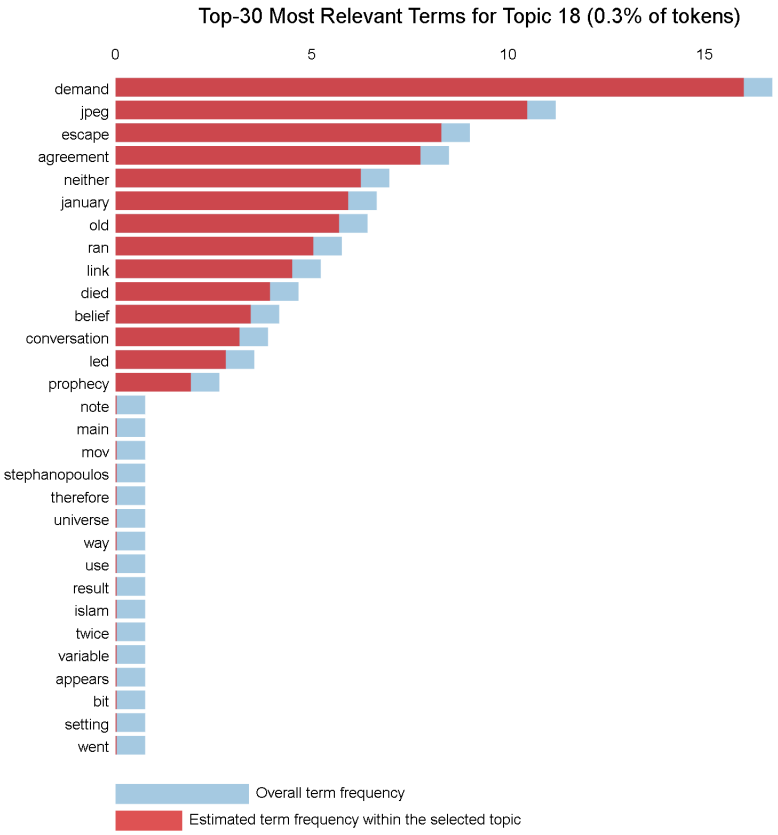

Top-30 Most Relevant Terms for Topic 13 (0.4% of tokens)
0
5
10
15
subject
guarantee
serdar
contest
hear
appreciated
build
portable
newsgroup
designed
compromise
committed
accept
posted
credit
mov
stephanopoulos
way
use
result
universe
islam
twice
variable
element
setting
bit
appears
went
billion
Overall term frequency
Estimated term frequency within the selected topic

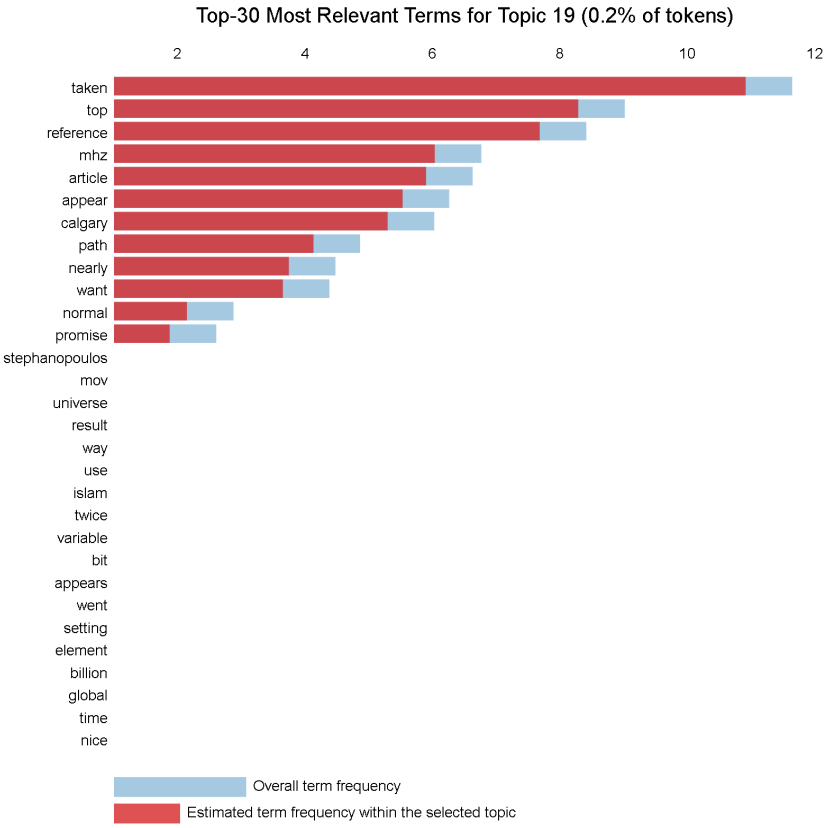

Top-30 Most Relevant Terms for Topic 14 (0.4% of tokens)
2
4
6
8
10
12
band
maintain
manager
selected
prove
existence
form
activity
actually
wish
anonymity
space
michael
century
guilty
russian
azeri
home
moon
authority
communication
potential
considered
foot
additional
reliable
yeah
likely
explanation
action
Overall term frequency
Estimated term frequency within the selected topic

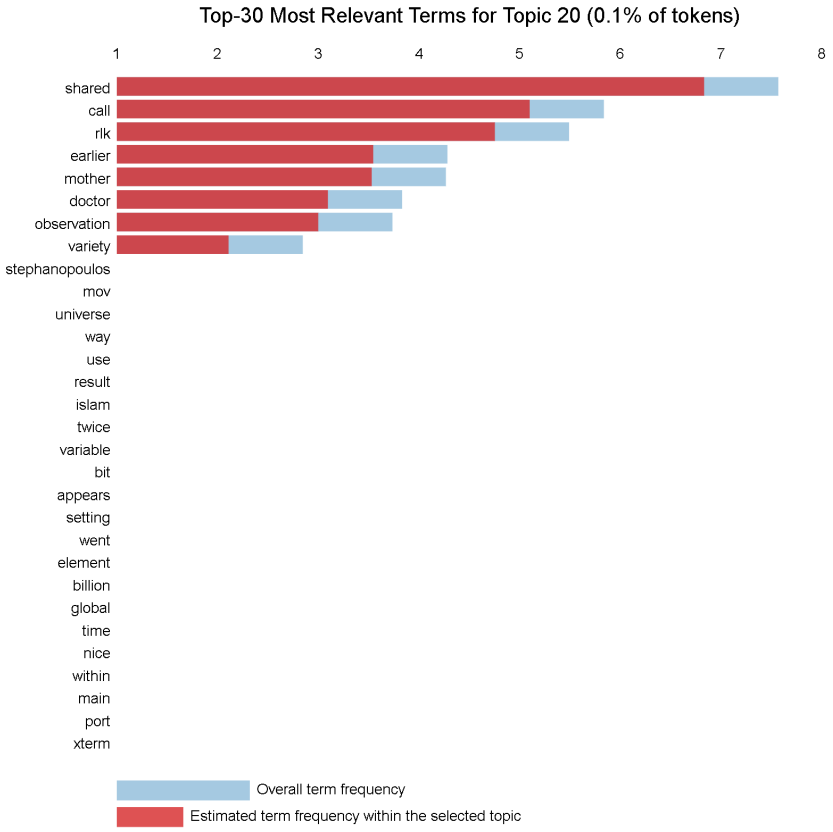

Top-30 Most Relevant Terms for Topic 15 (0.4% of tokens)
0
5
10
15
mode
statistic
stay
total
almost
peace
desire
hell
sequence
strong
alone
fourth
privacy
secure
mov
stephanopoulos
universe
result
way
use
islam
twice
variable
appears
bit
setting
went
element
billion
global
Overall term frequency
Estimated term frequency within the selected topic


Top-30 Most Relevant Terms for Topic 16 (0.3% of tokens)
2
4
6
8
10
12
14
size
try
sci
world
really
building
puck
study
wheel
education
second
exist
put
average
secret
mov
stephanopoulos
universe
result
way
use
islam
twice
variable
bit
went
appears
setting
element
billion
Overall term frequency
Estimated term frequency within the selected topic

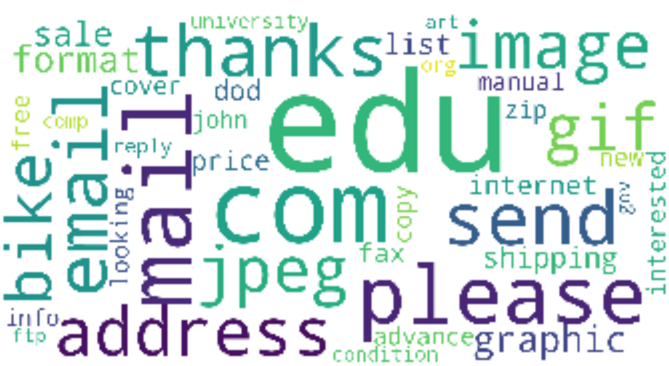
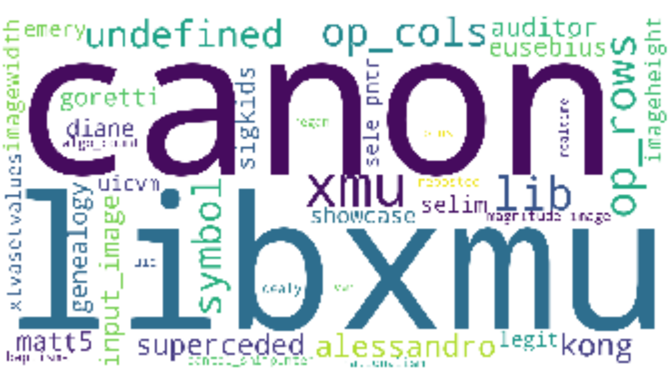

Top-30 Most Relevant Terms for Topic 17 (0.3% of tokens)
0
5
10
15
20
punishment
crime
experiment
batf
vendor
music
master
senior
meaning
washington
region
regard
cost
mov
stephanopoulos
universe
way
use
islam
twice
result
variable
bit
setting
went
appears
element
billion
global
time
Overall term frequency
Estimated term frequency within the selected topic


Top-30 Most Relevant Terms for Topic 18 (0.3% of tokens)
0
5
10
15
demand
jpeg
escape
agreement
neither
january
old
ran
link
died
belief
conversation
led
prophecy
note
main
mov
stephanopoulos
therefore
universe
way
use
result
islam
twice
variable
appears
bit
setting
went
Overall term frequency
Estimated term frequency within the selected topic


Top-30 Most Relevant Terms for Topic 19 (0.2% of tokens)
2
4
6
8
10
12
taken
top
reference
mhz
article
appear
calgary
path
nearly
want
normal
promise
stephanopoulos
mov
universe
result
way
use
islam
twice
variable
bit
appears
went
setting
element
billion
global
time
nice
Overall term frequency
Estimated term frequency within the selected topic


Top-30 Most Relevant Terms for Topic 20 (0.1% of tokens)
1
2
3
4
5
6
7
8
shared
call
rlk
earlier
mother
doctor
observation
variety
stephanopoulos
mov
universe
way
use
result
islam
twice
variable
bit
appears
setting
went
element
billion
global
time
nice
within
main
port
xterm
Overall term frequency
Estimated term frequency within the selected topic

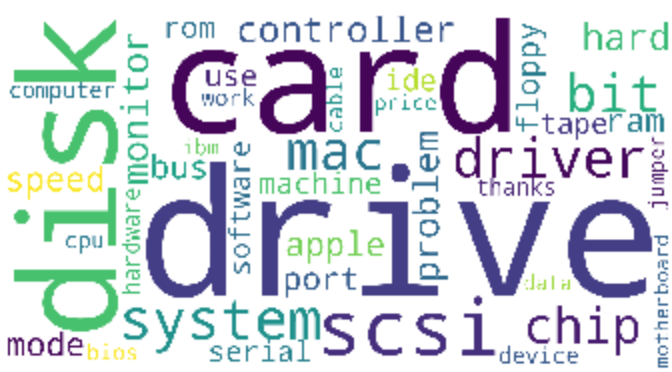

armenian
turkish
jew
greek
muslim
russian
war
turkey
jewish
turk
killed
soldier
village
nazi
armenia
people
history
genocide
azerbaijan
government
population
soviet
territory
million
ottoman
azeri
army
woman
land
massacre
university
argic
iran
town
civilian
republic

max
g9v
b8f
a86
bhj
giz
bxn
qax
chz
okz
b4q
nrhj
r8f
f9d
gizw
nuy
kjz
fyn
rlk
r186
biz
b8g
g9p
qtm
q5o
ghj
u3l
bhjnair
v9fq
q45
mg9v
c8v
p4u
fij
b8e
wm4u
mb8f
yf9
wwiz

edu
thanks
image
com
mail
email
bike
send
please
jpeg
address
gif
format
sale
university
art
list
cover
dod
manual
zip
john
free
reply
price
new
internet
copy
looking
fax
shipping
interested
graphic
advance
condition
info
ftp

canon
libxmu
undefined
op_cols
op_rows
auditor
eusebius
emery
imagewidth
imageheight
goretti
diane
sigkids
xmu
lib
uicvm
showcase
selim
symbol
genealogy
input_image
xlvasctvalues
matt5
superceded
alessandro
legit
kong

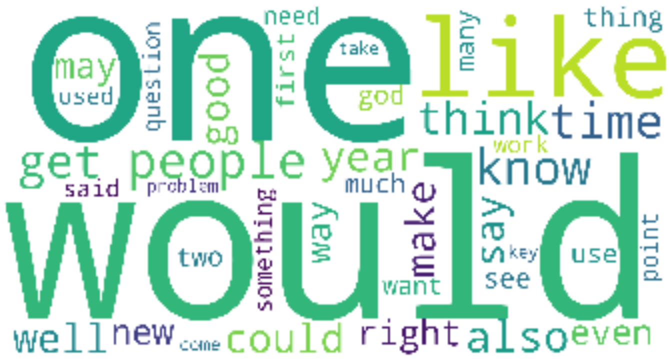

stephanopoulos
navy
naval
override
seminar
frame
lipran
winbench
strap
dee
visualization
afterlife
mil
rit
ulf
w4wg
copper
disney
grafsys
boot

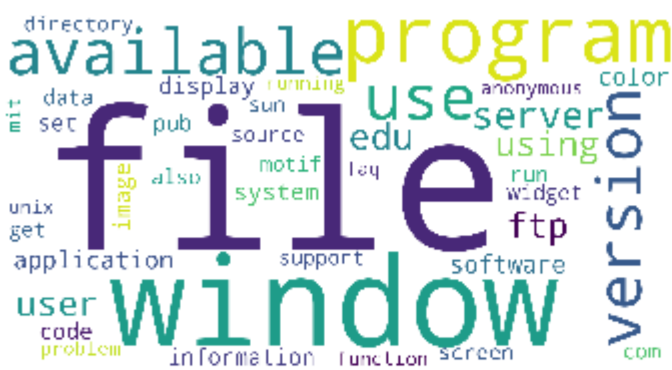

resistor
scope
corp
ingr
catbyte
b30
dtmedin
intergraph
speedo
everywhere
huntsville
uunet
gnd
gd3004
uucp
analog
delegate
internet
blinker
slam
polarity
plot_data
finest
vaccination
namaki
networking
medin
ecc
ssd
cathode
nickelodeon


enviroleague
zoroastrian
swinburne
prometheus
caligiuri
trinitarian
magi
celp
denizen
taoism
mnemonic
critz
ibka
claus
eternality
tul
anania
leviticus
paganism
atheisten
shk
ecclesiastical
rsadsi
subsistence
bunny
atwood
altar
aloysius
wch
unbelief
burzynski
een
mackie
prof
freethought
watchmaker
zoroaster


drive
card
disk
scsi
system
driver
mac
chip
bit
floppy
controller
monitor
hard
rom
use
work
ide
price
cable
computer
tape
ram
problem
bus
software
machine
thanks
speed
jumper
apple
hardware
cpu
port
data
motherboard
mode
bios
serial
device
ibm

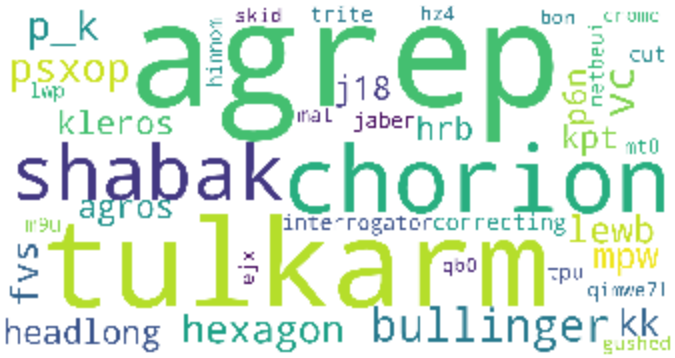

hitter
milwaukee
cooling
team
pit
league
suck
phiwing
bos
red
buf
run
edm
hit
nyr
minor
cup
player
sport
brave
que
cal
murray
mov
ott
pitcher
lost
water
det
season
van
cub
tor
stl
mon
nyi
chi
min
pitch


one
like
need
thing
take
many
first
may
question
used
good
god
think
time
get
people
year
work
know
said
problem
much
would
make
say
something
way
point
two
key
use
see
want
well
new
come
could
right
also
even


directory
available
program
display
running
data
anonymous
color
mit
set
sun
use
server
pub
source
edu
using
file
version
motif
also
run
system
widget
image
unix
get
ftp
application
support
software
window
user
code
problem
information
function
screen
com


g3o
graeme
myhint
psize
trinomials
main_win
gsr
tl
acura
xsizehints
eeg
thx
ldl
y86
wasp
detail_win
scand
integra
stafford
decompress

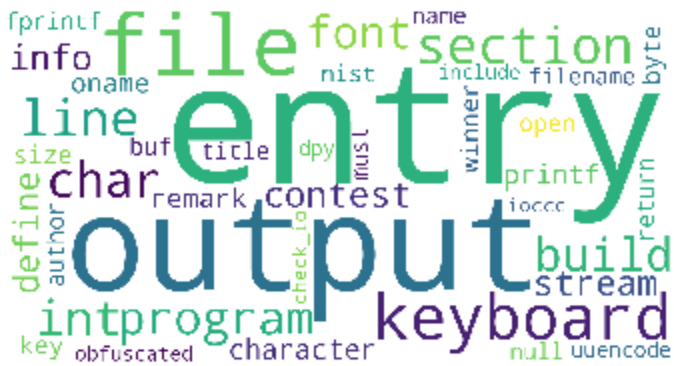

game
team
play
nhl
period
goal
chicago
first
power
ice
puck
calgary
angeles
pittsburgh
boston
king
toronto
montreal
star
ranger
louis
minnesota
leaf
san
detroit
scoring
flyer


agrep
shabak
chorion
tulkarm
p_k
psxop
kleros
hrb
kpt
vc
j18
jaber
fvs
agros
lewb
mpw
headlong
hexagon
bullinger
kk

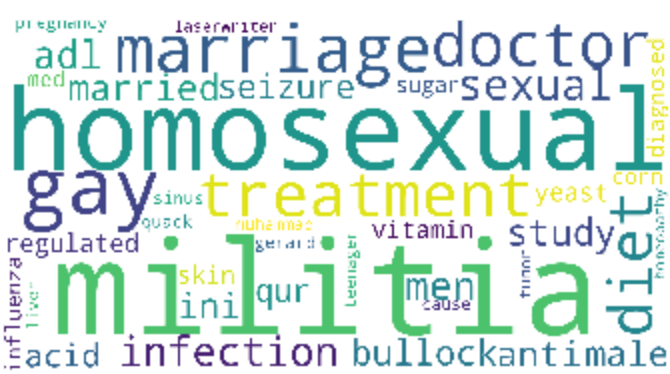

lavrencic
p_c
ihr
kurt
nye
uwec
cjackson
distsq
pps
muskingum
humanist
p_4
nyeda
cheering
clubbing
cnsvax
pads
cns
die
millie
knx
masoretic
funeral
whl
caltrans

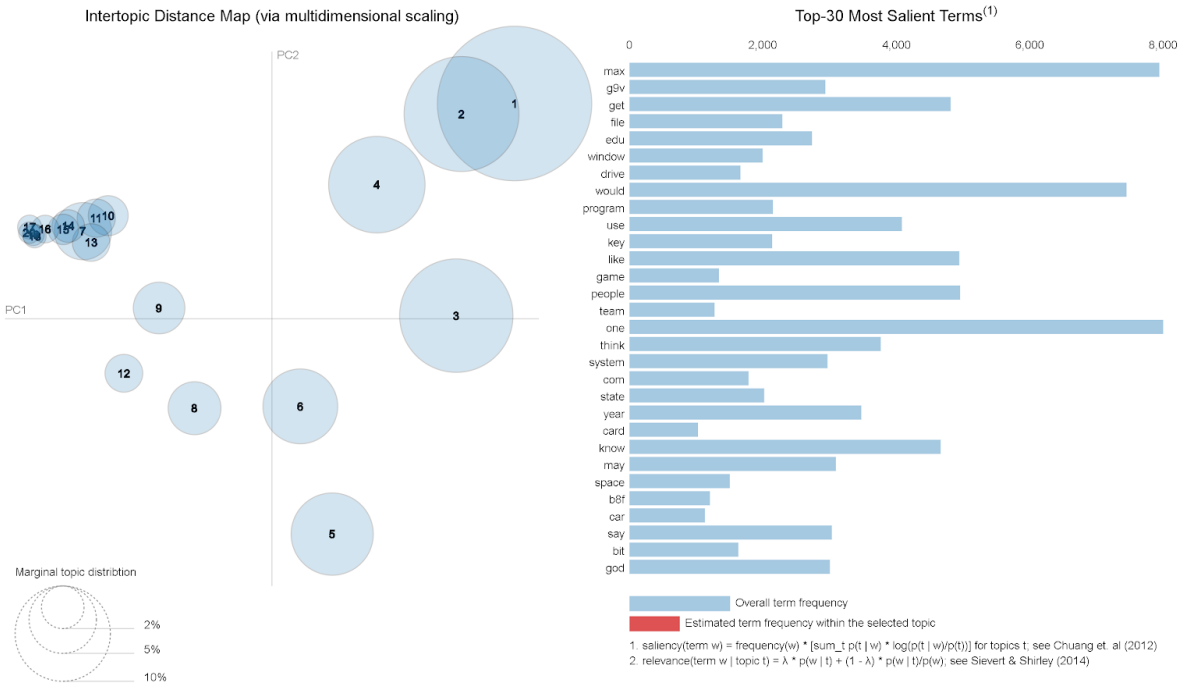

space
mission
program
system
orbit
center
earth
science
data
nasa
moon
satellite
research
launch
shuttle

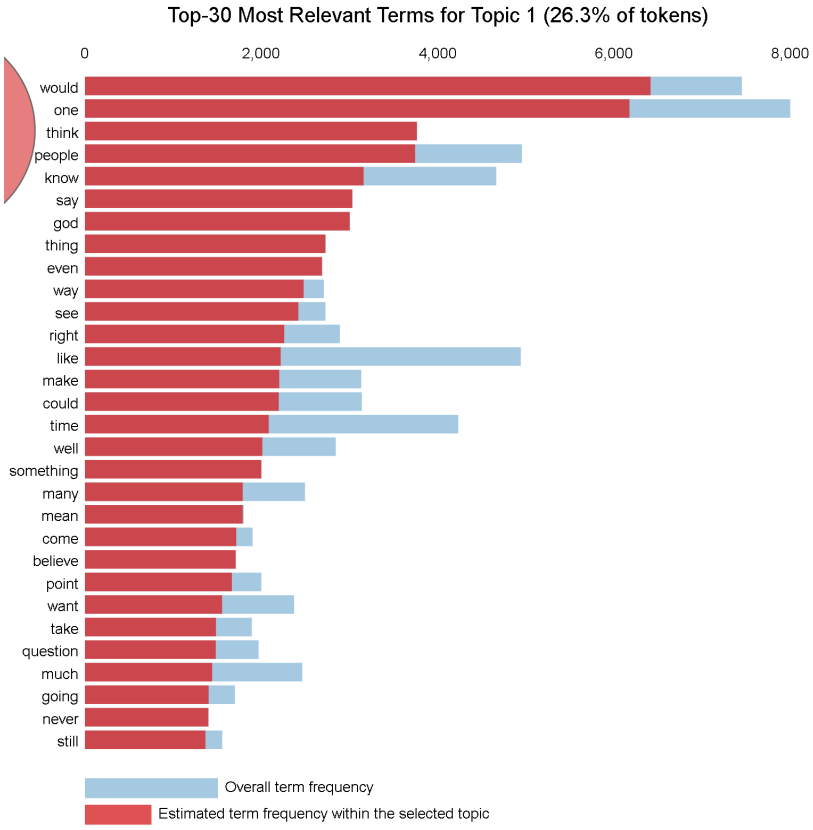

gun
weapon
firearm
crime
control
printer
law
handgun
rate
police
auto
defense
homicide
death
self
assault
violent
rkba
automatic
rifle
state
city
armed
study
carry
vancouver
pistol
limbaugh
shotgun
arm
revolver
seattle
semi
nra
laser
safety
batf
concealed
accident
cycle

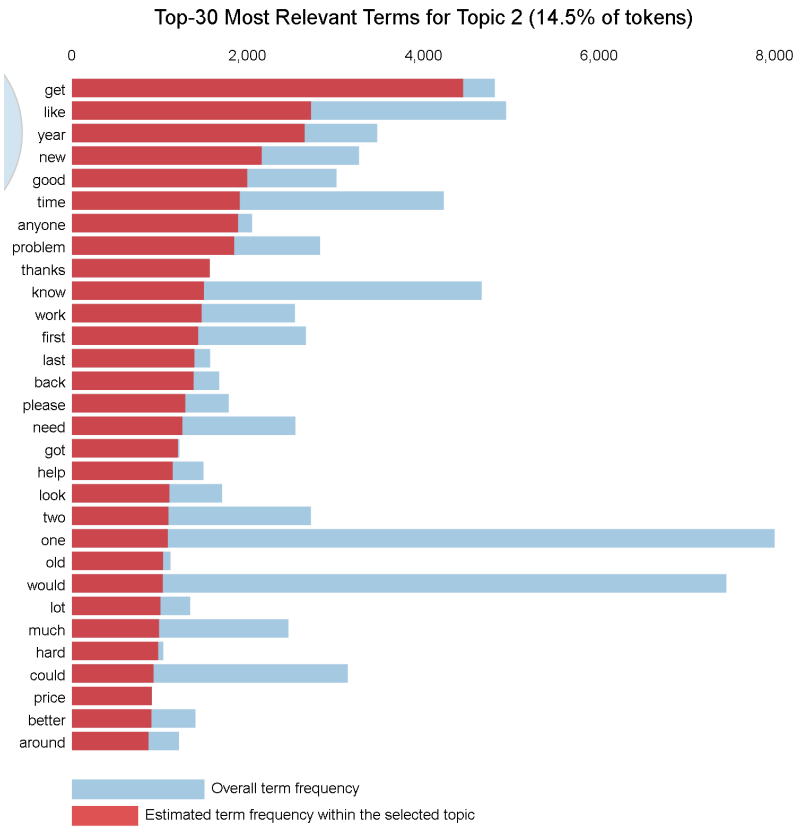

entry
output
file
section
font
keyboard
intprogram
char
line
info
contest
build
stream
character
define
author
remark
title
buf
size
oname
fprintf
name
nist
include
filename
byte
open
printf
winner
must
dpy
return
ioccc
key
obfuscated
null
uuencode

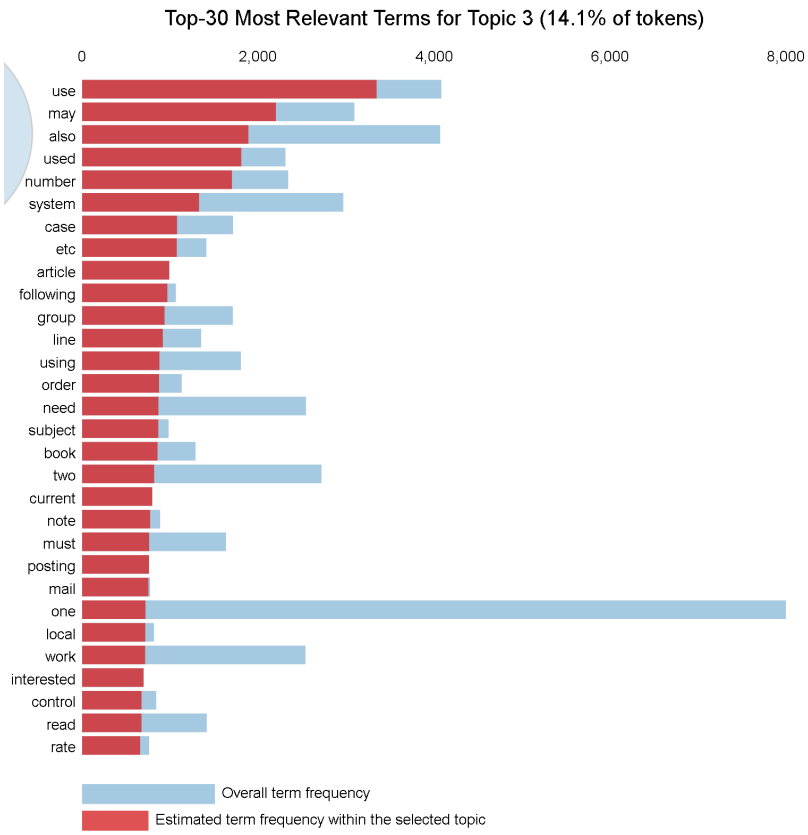

chz
scx
lhz
air
w47
sc
ahf
b9r
ww7
w3s
hzv
mcxsqst3p
w45
cj1
shz
ck8
fdz
w3p
rck
t45
w44
cho
neu
x_scx
sy_
t3q
w3q
uww
xte
t3s
uw0
w1w
mc_
xxi
sj1
cx_scx
gc_

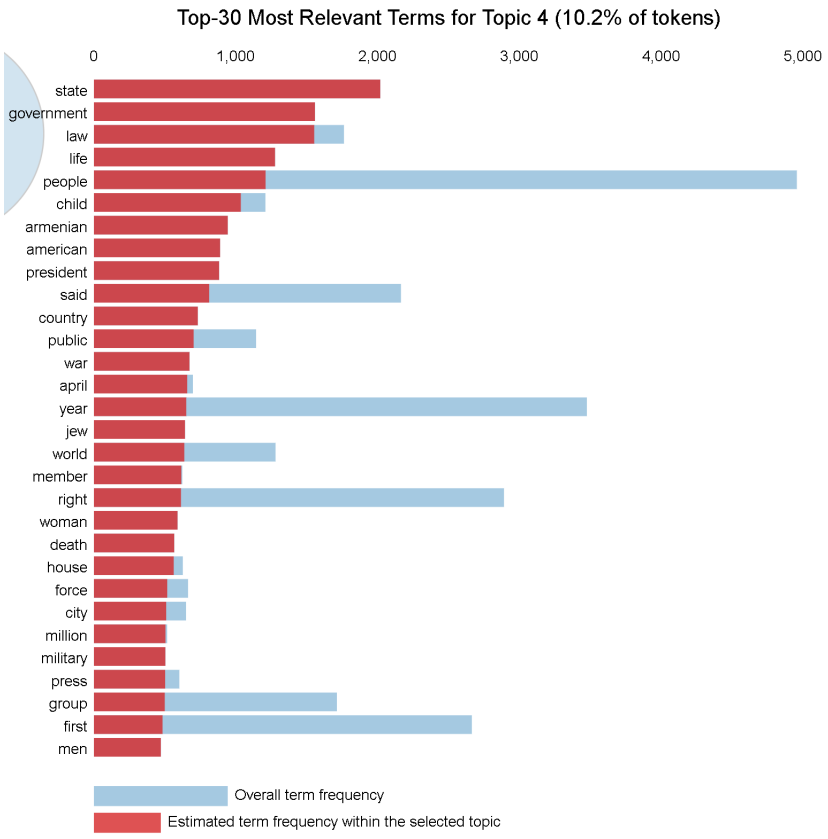

homosexual
militia
treatment
gay
marriagedoctor
married
seizure
sexual
sugar
adl
diet
study
vitamin
yeast
regulated
men
qur
ini
skin
infection
bullockantimale
acid
influenza
diagnosed
sinus
quack
pregnancy
laserwriter
med

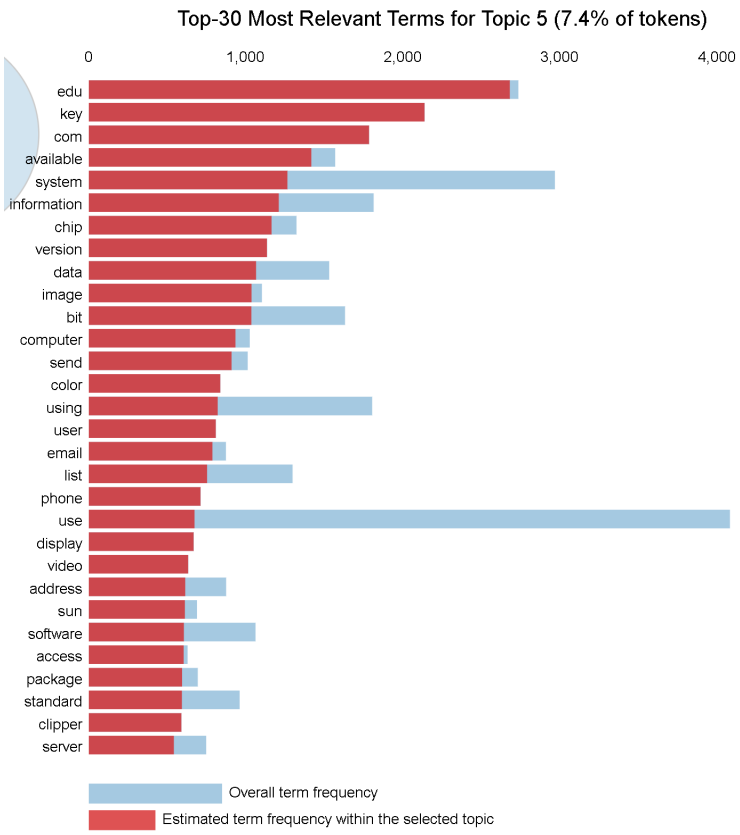

Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
Marginal topic distribtion
2%
5%
10%
Top-30 Most Salient Terms(1)
0
2,000
4,000
6,000
8,000
max
g9v
get
file
edu
window
drive
would
program
use
key
like
game
people
team
one
think
system
com
state
year
card
know
may
space
b8f
car
say
bit
god
Overall term frequency
Estimated term frequency within the selected topic
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

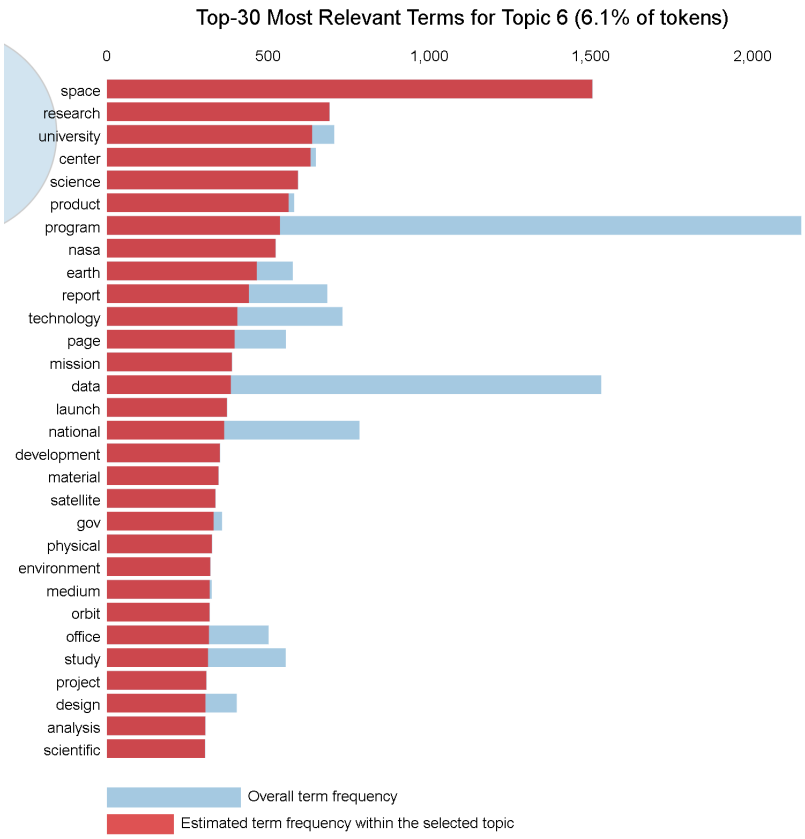

Top-30 Most Relevant Terms for Topic 1 (26.3% of tokens)
0
2,000
4,000
6,000
8,000
would
one
think
people
know
say
god
thing
even
way
see
right
like
make
could
time
well
something
many
mean
come
believe
point
want
take
question
much
going
never
still
Overall term frequency
Estimated term frequency within the selected topic

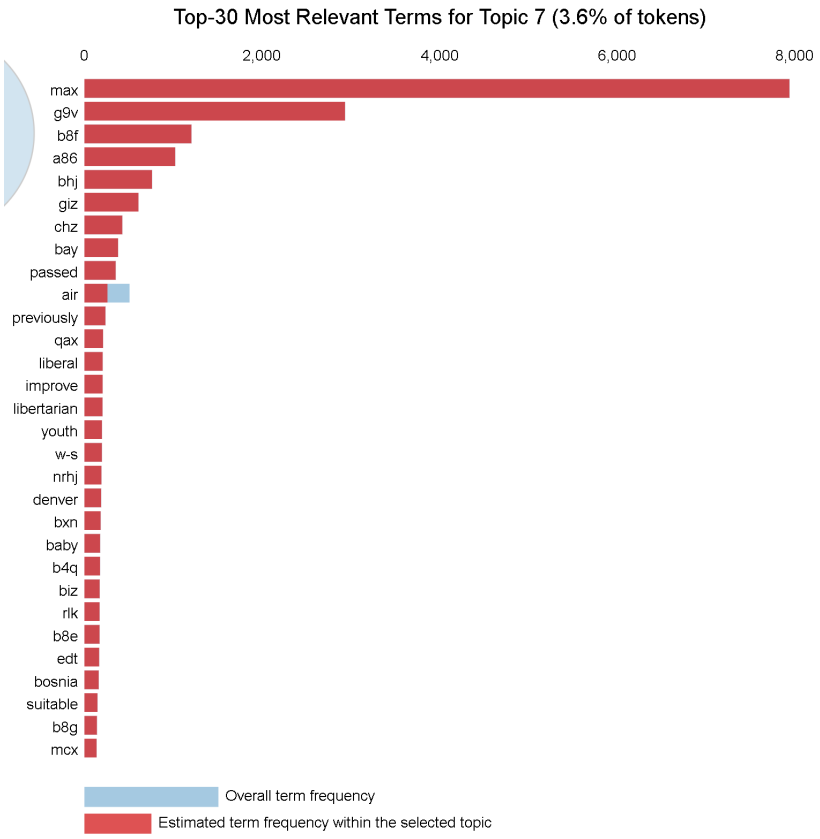

Top-30 Most Relevant Terms for Topic 2 (14.5% of tokens)
0
2,000
4,000
6,000
8,000
get
like
year
new
good
time
anyone
problem
thanks
know
work
first
last
back
please
need
got
help
look
two
one
old
would
lot
much
hard
could
price
better
around
Overall term frequency
Estimated term frequency within the selected topic

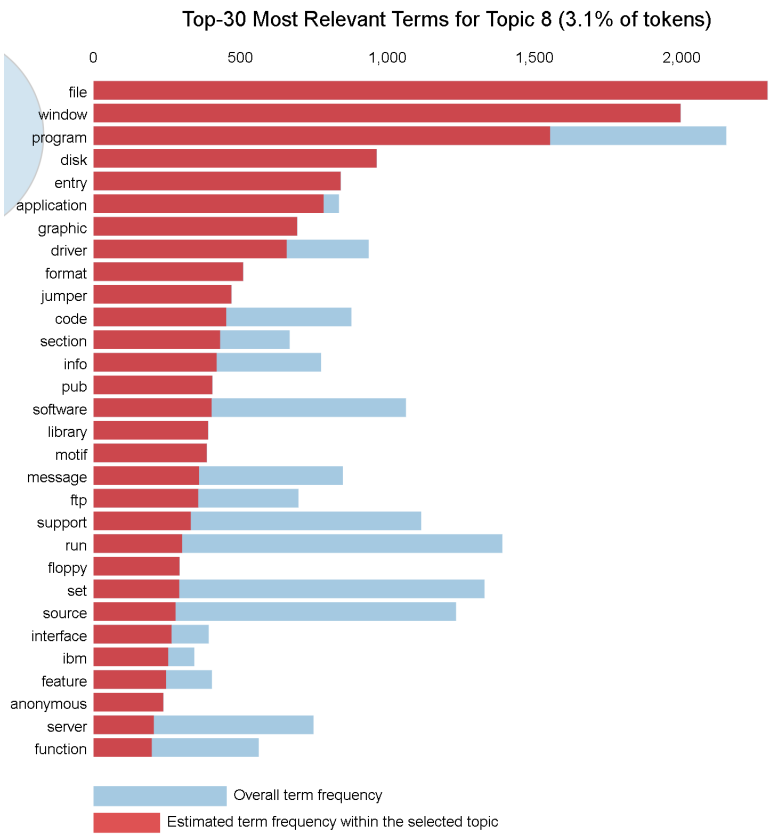

Top-30 Most Relevant Terms for Topic 3 (14.1% of tokens)
0
2,000
4,000
6,000
8,000
use
may
also
used
number
system
case
etc
article
following
group
line
using
order
need
subject
book
two
current
note
must
posting
mail
one
local
work
interested
control
read
rate
Overall term frequency
Estimated term frequency within the selected topic

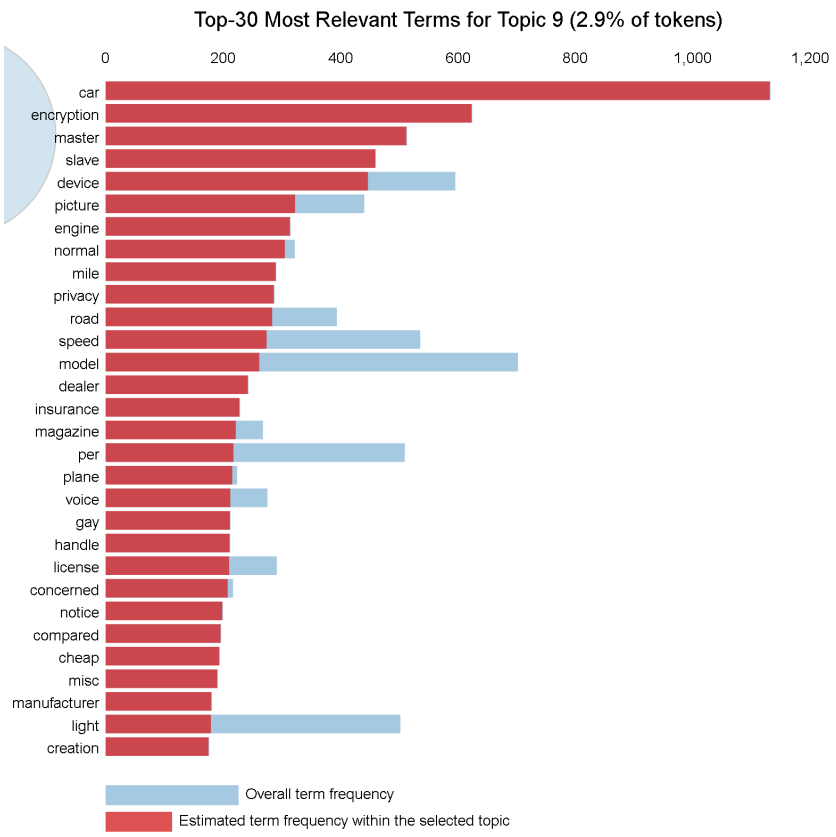

Top-30 Most Relevant Terms for Topic 4 (10.2% of tokens)
0
1,000
2,000
3,000
4,000
5,000
state
government
law
life
people
child
armenian
american
president
said
country
public
war
april
year
jew
world
member
right
woman
death
house
force
city
million
military
press
group
first
men
Overall term frequency
Estimated term frequency within the selected topic

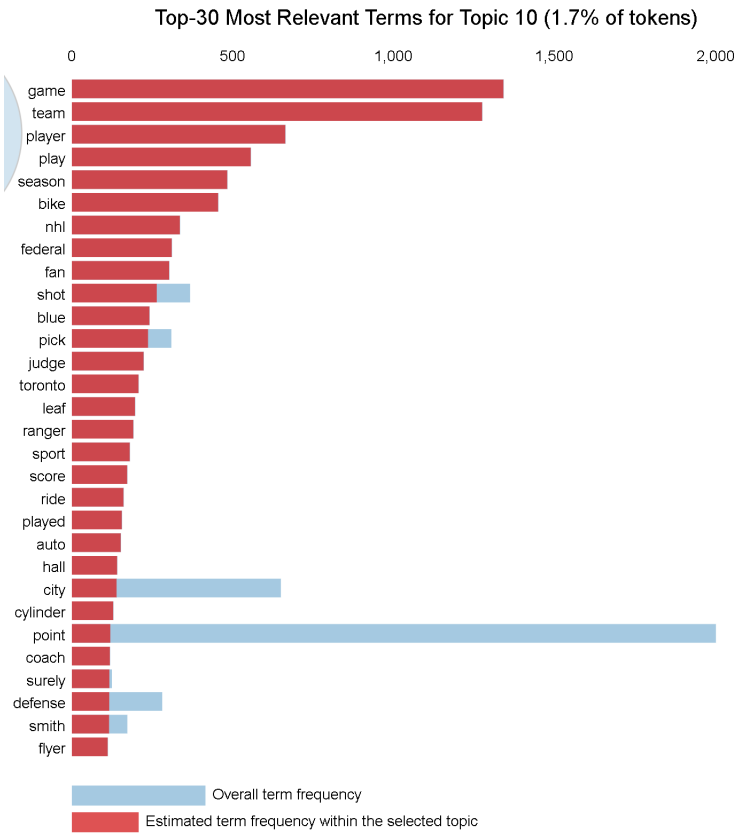

Top-30 Most Relevant Terms for Topic 5 (7.4% of tokens)
0
1,000
2,000
3,000
4,000
edu
key
com
available
system
information
chip
version
data
image
bit
computer
send
color
using
user
email
list
phone
use
display
video
address
sun
software
access
package
standard
clipper
server
Overall term frequency
Estimated term frequency within the selected topic

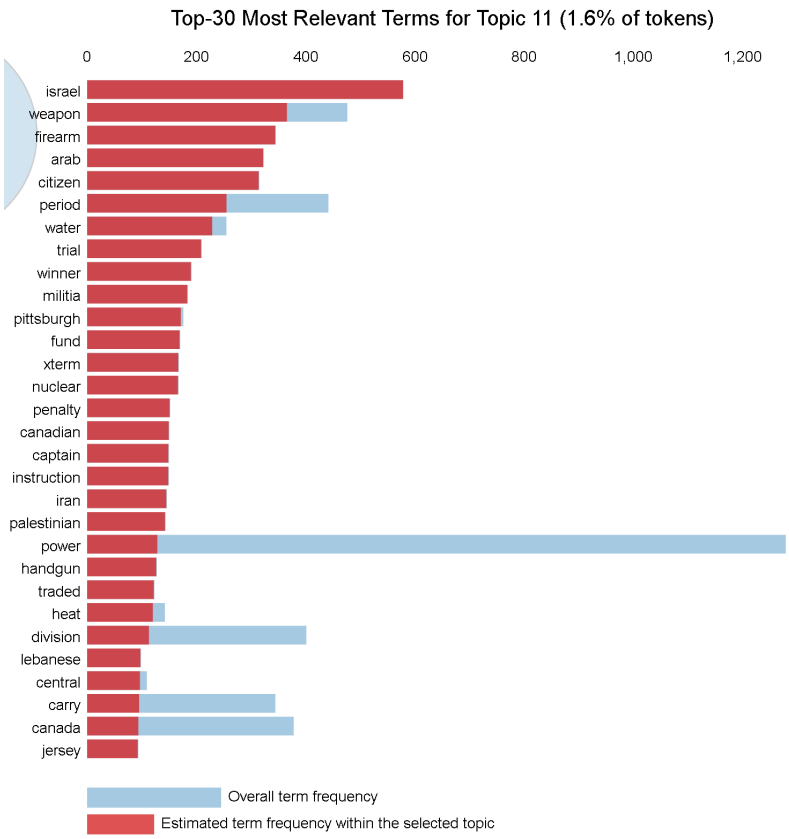

Top-30 Most Relevant Terms for Topic 6 (6.1% of tokens)
0
500
1,000
1,500
2,000
space
research
university
center
science
product
program
nasa
earth
report
technology
page
mission
data
launch
national
development
material
satellite
gov
physical
environment
medium
orbit
office
study
project
design
analysis
scientific
Overall term frequency
Estimated term frequency within the selected topic

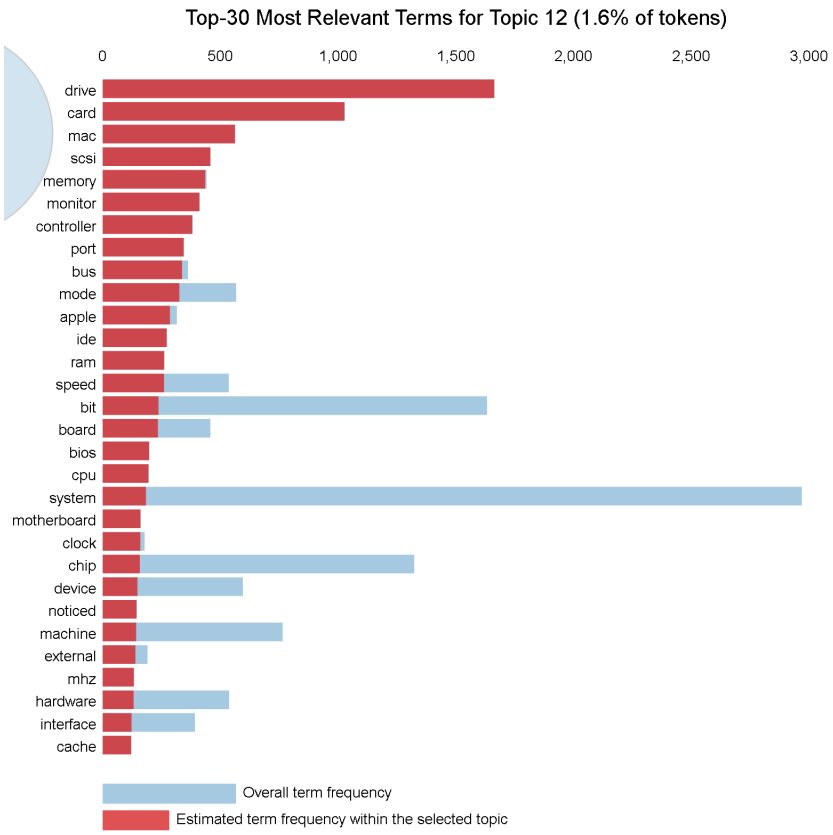

Top-30 Most Relevant Terms for Topic 7 (3.6% of tokens)
0
2,000
4,000
6,000
8,000
max
g9v
b8f
a86
bhj
giz
chz
bay
passed
air
previously
qax
liberal
improve
libertarian
youth
w-s
nrhj
denver
bxn
baby
b4q
biz
rlk
b8e
edt
bosnia
suitable
b8g
mcx
Overall term frequency
Estimated term frequency within the selected topic

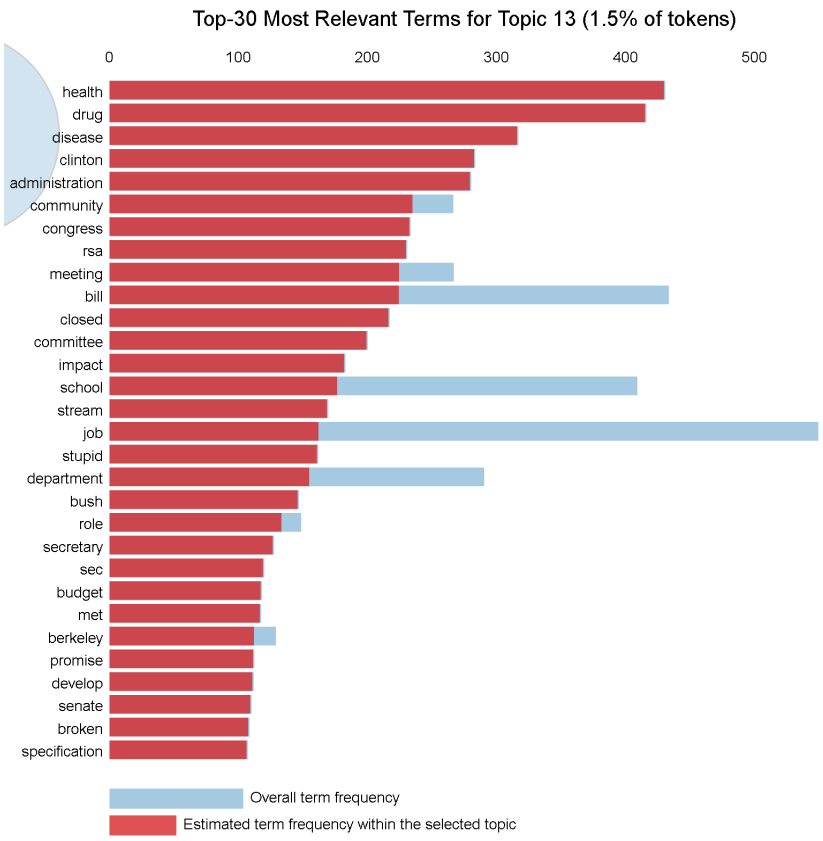

Top-30 Most Relevant Terms for Topic 8 (3.1% of tokens)
0
500
1,000
1,500
2,000
file
window
program
disk
entry
application
graphic
driver
format
jumper
code
section
info
pub
software
library
motif
message
ftp
support
run
floppy
set
source
interface
ibm
feature
anonymous
server
function
Overall term frequency
Estimated term frequency within the selected topic

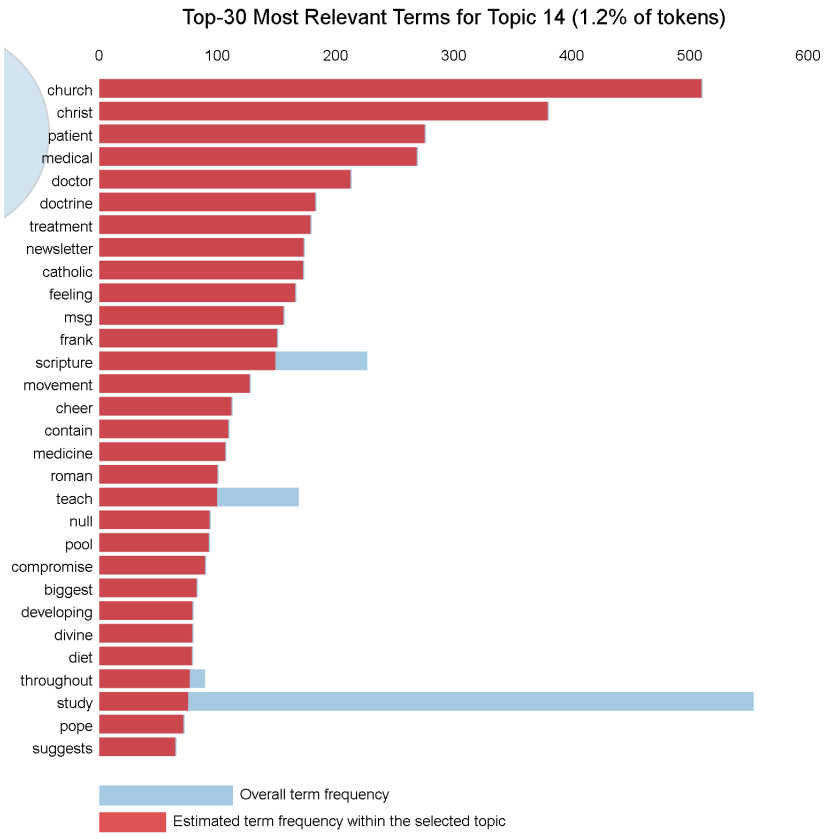

Top-30 Most Relevant Terms for Topic 9 (2.9% of tokens)
0
200
400
600
800
1,000
1,200
car
encryption
master
slave
device
picture
engine
normal
mile
privacy
road
speed
model
dealer
insurance
magazine
per
plane
voice
gay
handle
license
concerned
notice
compared
cheap
misc
manufacturer
light
creation
Overall term frequency
Estimated term frequency within the selected topic

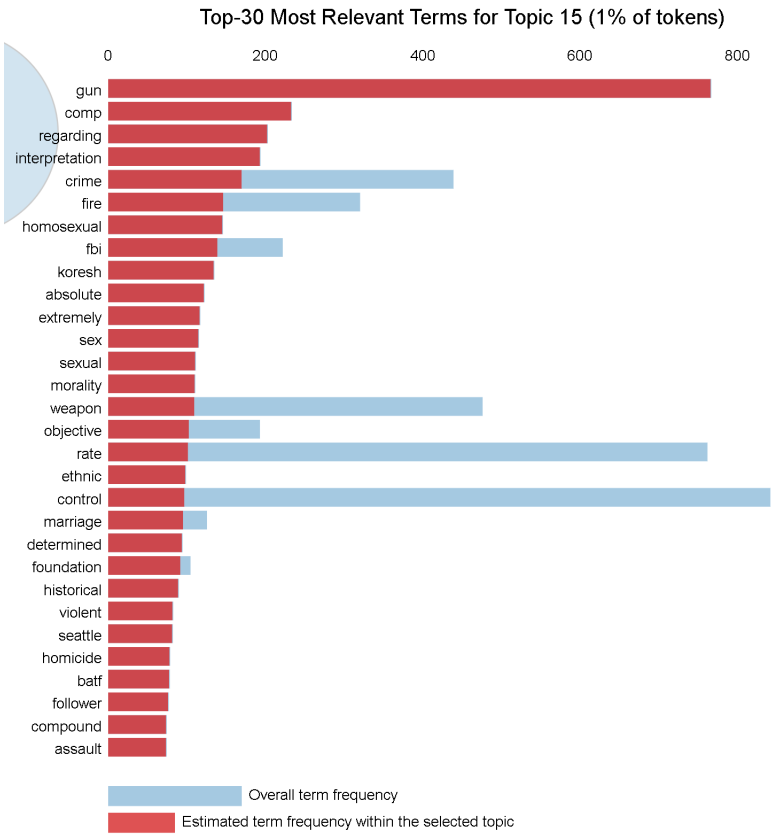

Top-30 Most Relevant Terms for Topic 10 (1.7% of tokens)
0
500
1,000
1,500
2,000
game
team
player
play
season
bike
nhl
federal
fan
shot
blue
pick
judge
toronto
leaf
ranger
sport
score
ride
played
auto
hall
city
cylinder
point
coach
surely
defense
smith
flyer
Overall term frequency
Estimated term frequency within the selected topic

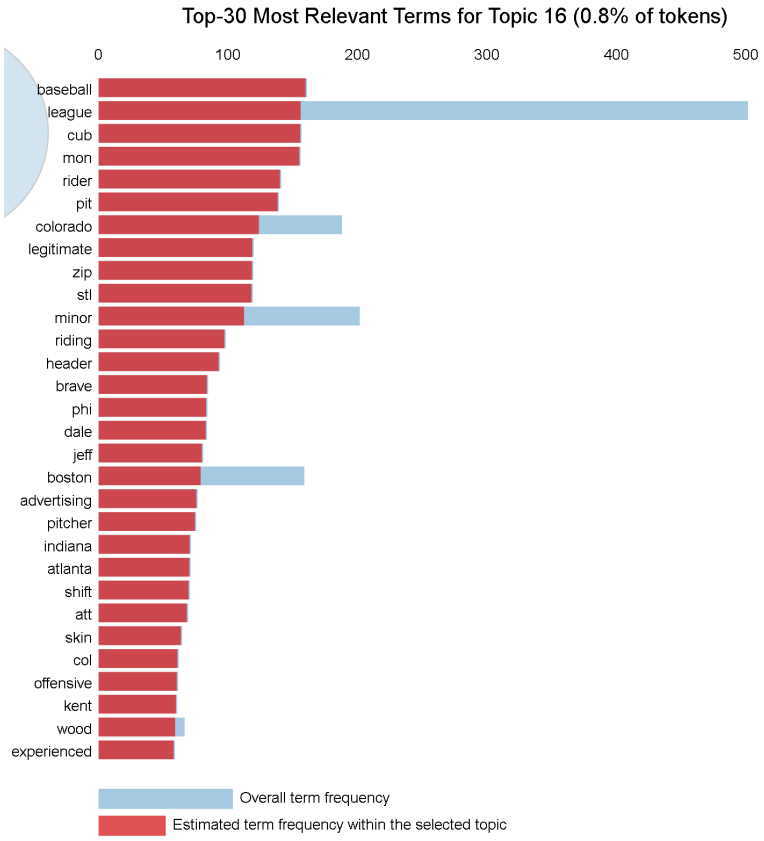

Top-30 Most Relevant Terms for Topic 11 (1.6% of tokens)
0
200
400
600
800
1,000
1,200
israel
weapon
firearm
arab
citizen
period
water
trial
winner
militia
pittsburgh
fund
xterm
nuclear
penalty
canadian
captain
instruction
iran
palestinian
power
handgun
traded
heat
division
lebanese
central
carry
canada
jersey
Overall term frequency
Estimated term frequency within the selected topic

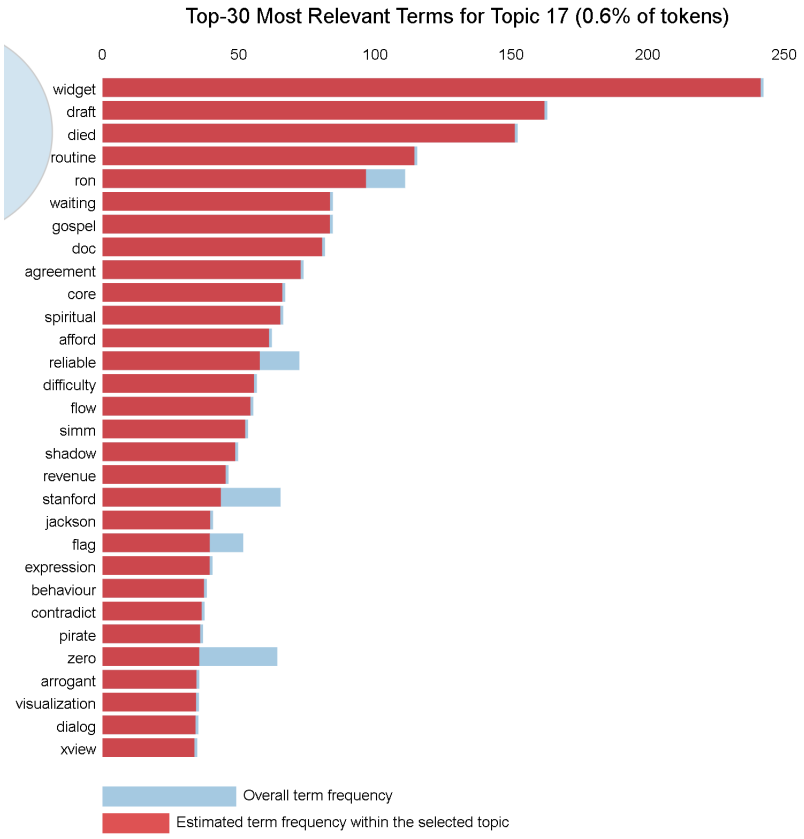

Top-30 Most Relevant Terms for Topic 12 (1.6% of tokens)
0
500
1,000
1,500
2,000
2,500
3,000
drive
card
mac
scsi
memory
monitor
controller
port
bus
mode
apple
ide
ram
speed
bit
board
bios
cpu
system
motherboard
clock
chip
device
noticed
machine
external
mhz
hardware
interface
cache
Overall term frequency
Estimated term frequency within the selected topic

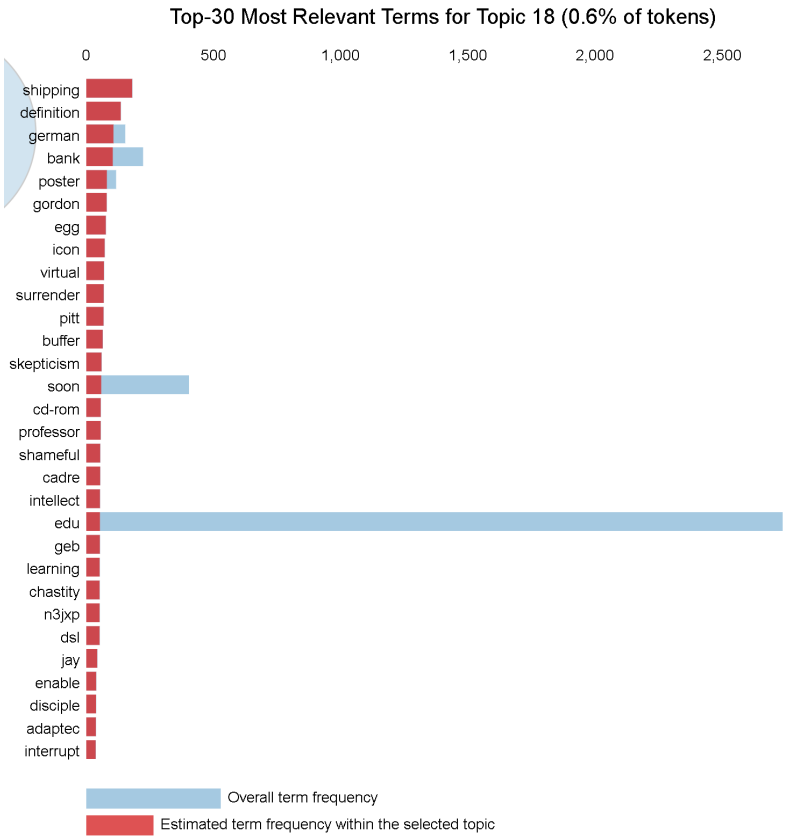

Top-30 Most Relevant Terms for Topic 13 (1.5% of tokens)
0
100
200
300
400
500
health
drug
disease
clinton
administration
community
congress
rsa
meeting
bill
closed
committee
impact
school
stream
job
stupid
department
bush
role
secretary
sec
budget
met
berkeley
promise
develop
senate
broken
specification
Overall term frequency
Estimated term frequency within the selected topic

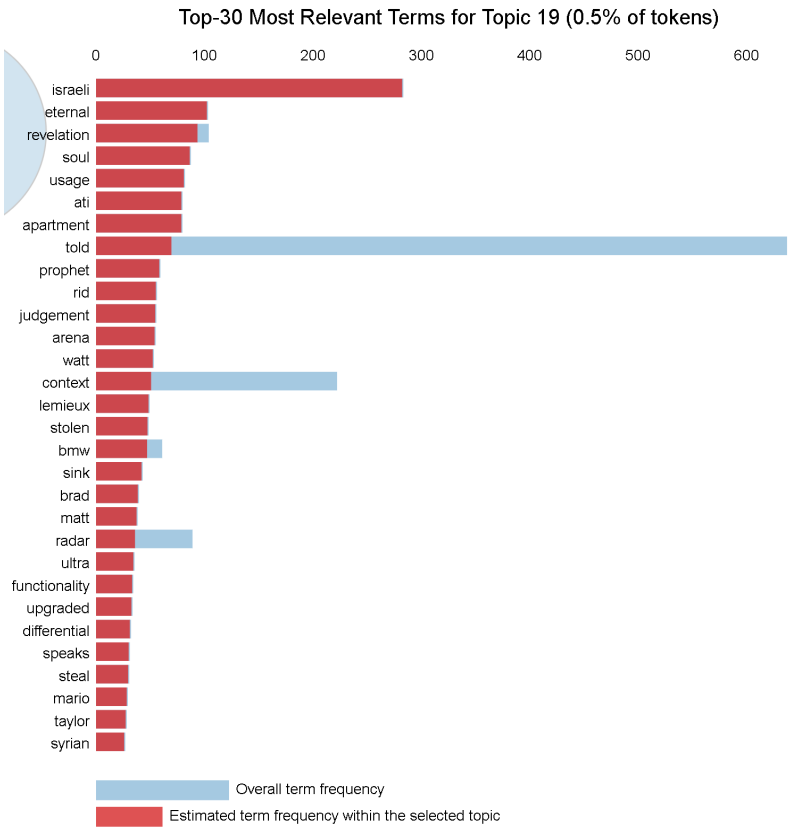

Top-30 Most Relevant Terms for Topic 14 (1.2% of tokens)
0
100
200
300
400
500
600
church
christ
patient
medical
doctor
doctrine
treatment
newsletter
catholic
feeling
msg
frank
scripture
movement
cheer
contain
medicine
roman
teach
null
pool
compromise
biggest
developing
divine
diet
throughout
study
pope
suggests
Overall term frequency
Estimated term frequency within the selected topic

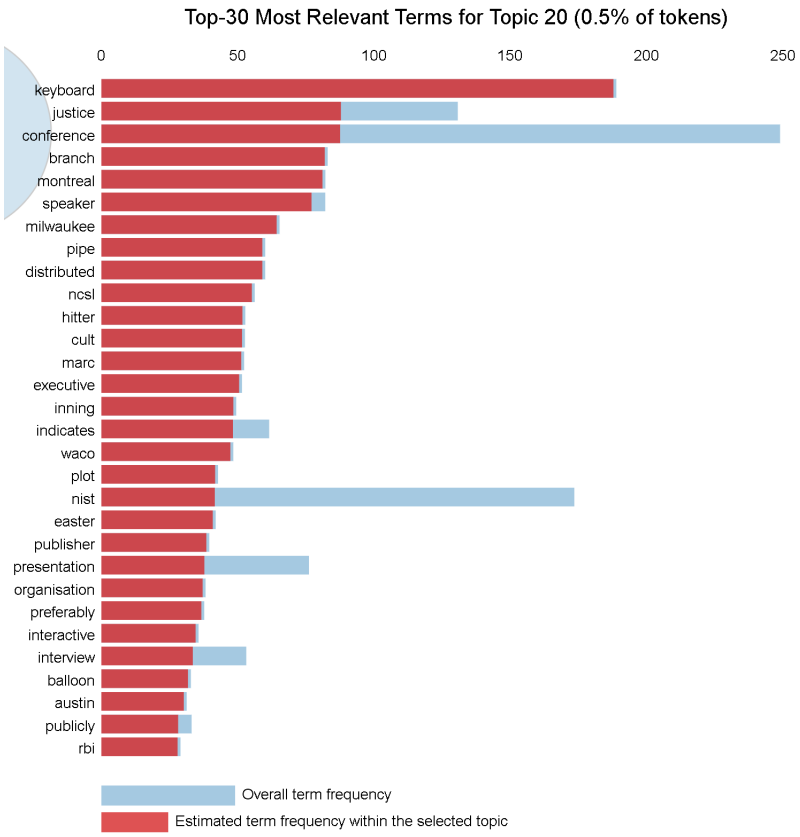

Top-30 Most Relevant Terms for Topic 15 (1% of tokens)
0
200
400
600
800
gun
comp
regarding
interpretation
crime
fire
homosexual
fbi
koresh
absolute
extremely
sex
sexual
morality
weapon
objective
rate
ethnic
control
marriage
determined
foundation
historical
violent
seattle
homicide
batf
follower
compound
assault
Overall term frequency
Estimated term frequency within the selected topic

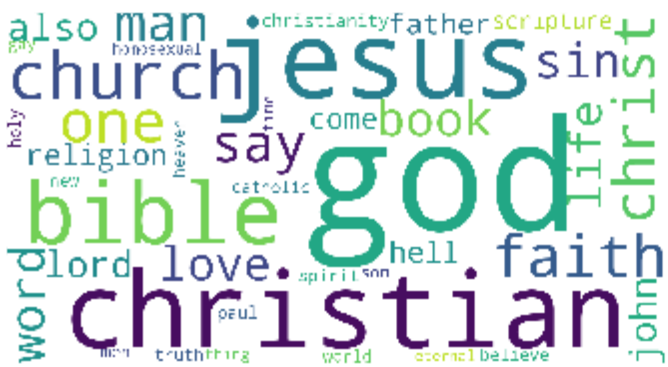

Top-30 Most Relevant Terms for Topic 16 (0.8% of tokens)
0
100
200
300
400
500
baseball
league
cub
mon
rider
pit
colorado
legitimate
zip
stl
minor
riding
header
brave
phi
dale
jeff
boston
advertising
pitcher
indiana
atlanta
shift
att
skin
col
offensive
kent
wood
experienced
Overall term frequency
Estimated term frequency within the selected topic


Top-30 Most Relevant Terms for Topic 17 (0.6% of tokens)
0
50
100
150
200
250
widget
draft
died
routine
ron
waiting
gospel
doc
agreement
core
spiritual
afford
reliable
difficulty
flow
simm
shadow
revenue
stanford
jackson
flag
expression
behaviour
contradict
pirate
zero
arrogant
visualization
dialog
xview
Overall term frequency
Estimated term frequency within the selected topic

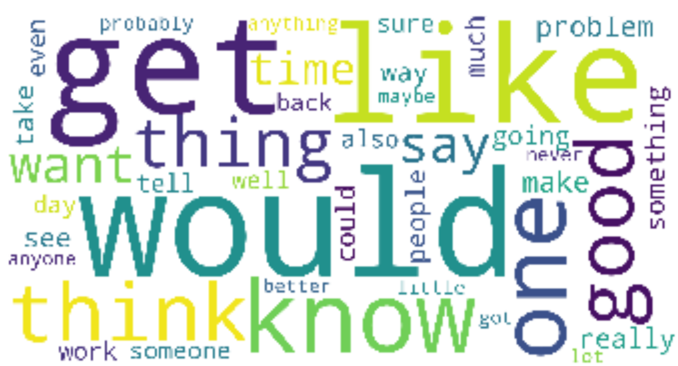

Top-30 Most Relevant Terms for Topic 18 (0.6% of tokens)
0
500
1,000
1,500
2,000
2,500
shipping
definition
german
bank
poster
gordon
egg
icon
virtual
surrender
pitt
buffer
skepticism
soon
cd-rom
professor
shameful
cadre
intellect
edu
geb
learning
chastity
n3jxp
dsl
jay
enable
disciple
adaptec
interrupt
Overall term frequency
Estimated term frequency within the selected topic

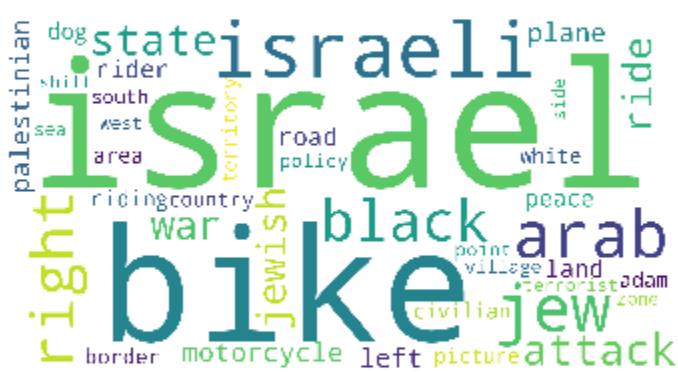

Top-30 Most Relevant Terms for Topic 19 (0.5% of tokens)
0
100
200
300
400
500
600
israeli
eternal
revelation
soul
usage
ati
apartment
told
prophet
rid
judgement
arena
watt
context
lemieux
stolen
bmw
sink
brad
matt
radar
ultra
functionality
upgraded
differential
speaks
steal
mario
taylor
syrian
Overall term frequency
Estimated term frequency within the selected topic


Top-30 Most Relevant Terms for Topic 20 (0.5% of tokens)
0
50
100
150
200
250
keyboard
justice
conference
branch
montreal
speaker
milwaukee
pipe
distributed
ncsl
hitter
cult
marc
executive
inning
indicates
waco
plot
nist
easter
publisher
presentation
organisation
preferably
interactive
interview
balloon
austin
publicly
rbi
Overall term frequency
Estimated term frequency within the selected topic

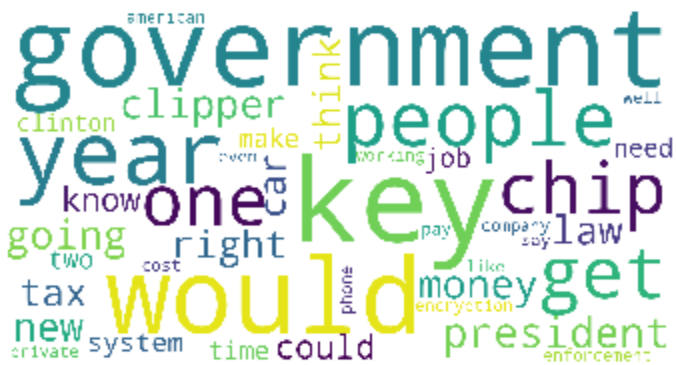

god
christian
jesus
bible
church
christ
faith
love
lord
word
one
say
book
life
sin
man
also
religion
come
father
scripture
christianity
homosexual
hell
john
catholic
paul
truth
thing
world
believe
spirit
son
holy
new


max
g9v
b8f
a86
bhj
giz
bxn
chz
okz
rlk
b4q
fij
b9r
biz
nrhj
b8e
qax
air
part
end

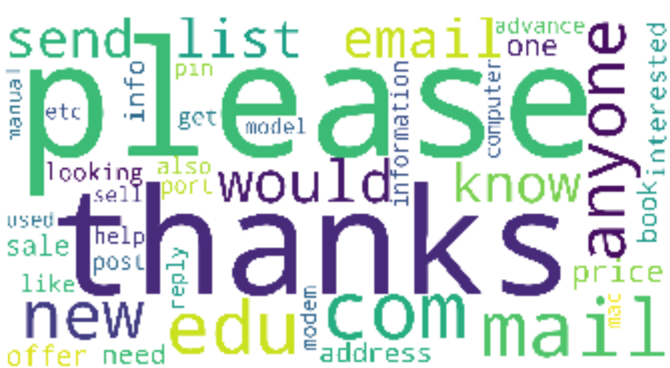

would
like
get
know
think
one
good
thing
say
want
time
probably
anything
sure
problem
even
much
way
maybe
back
take
also
going
never
something
tell
well
make
day
could
people
see
anyone
better
little
got
really
work
someone
lot

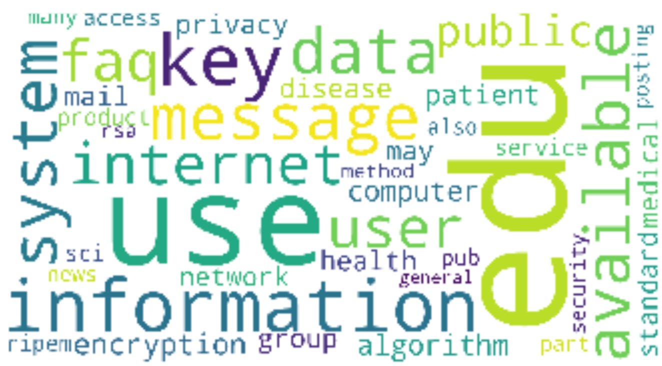

israel
bike
israeli
state
black
arab
right
jewish
jew
attack
ride
war
dog
plane
rider
palestinian
south
west
sea
area
road
policy
white
territory
riding
country
peace
point
village
land
adam
terrorist
zone
civilian
border
motorcycle
left
picture
side

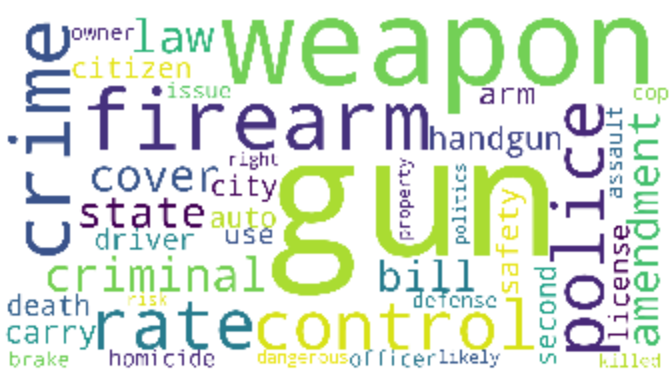

site
ftp
object
anonymous
food
comment
aid
steve
host
poster
myers
archive
effect
reaction
msg
family
cause
zip
math
escape
recent
mirror
shell
anonymity
ray
number
abuse
found
contain
good
others
might
missing
announced
description
problem
sex
shareware
collection


government
key
would
year
people
chip
one
get
clipper
think
car
know
going
right
law
money
president
tax
new
system
could
time
two
make
job
need
clinton
american
well
pay
company
cost
like
phone
encryption
private
enforcement
working

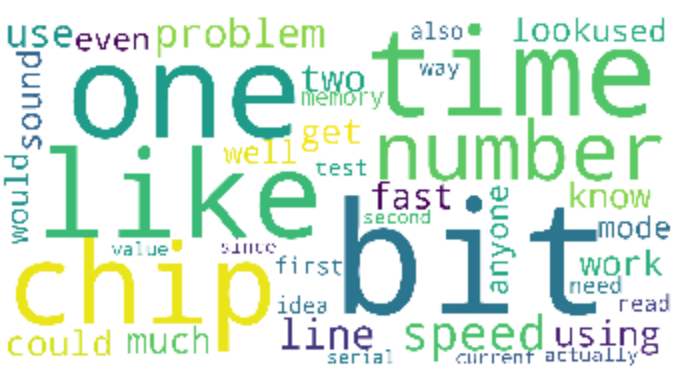

armenian
people
said
turkish
went
government
woman
greek
child
one
year
war
nazi
two
first
told
turk
came
muslim
turkey
home
killed
world
men
city
armenia
dead
saw
soldier
russian
jew
time
history
town
country
german
army
started
building
azerbaijani


please
thanks
send
list
email
anyone
would
know
mail
com
edu
new
info
information
computer
interested
advance
one
pin
etc
get
model
manual
looking
also
port
sell
used
sale
help
post
like
reply
modem
address
price
book
mac
offer
need

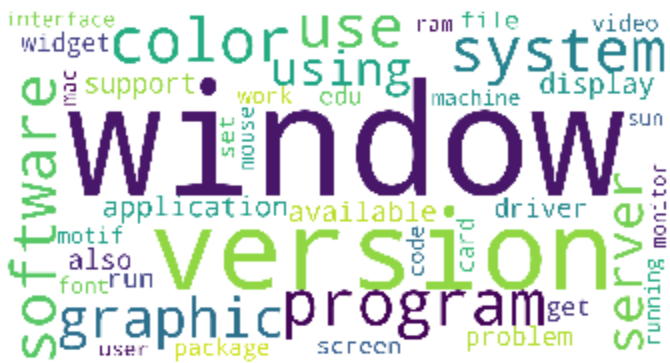

edu
use
information
message
key
faq
data
public
system
internet
available
user
computer
patient
disease
privacy
access
mail
health
network
encryption
algorithm
group
standard
medical
security
posting
service
also
may
method
pub
general
sci
news
rsa
ripem
part

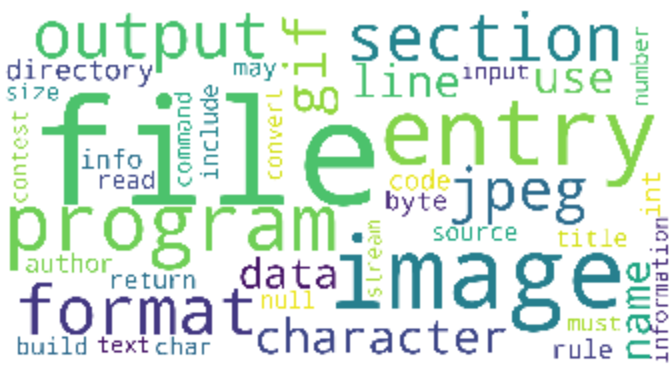

gun
weapon
firearm
crime
police
control
rate
criminal
state
cover
law
handgun
amendment
bill
safety
license
second
city
auto
driver
use
death
carry
homicide
brake
officer
likely
dangerous
killed
assault
politics
property
right
issue
citizen
owner
arm
cop
defense
risk

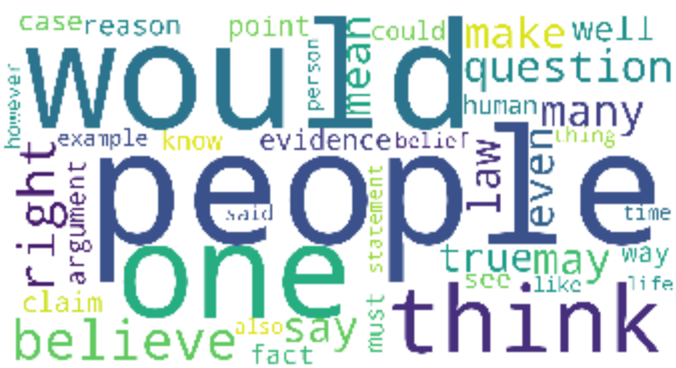

drive
disk
scsi
card
controller
hard
system
floppy
ide
bios
bus
tape
master
rom
device
motherboard
jumper
boot
transfer
slot
mcg
board
power
problem
isa
slave
two
switch
head
backup
connector
internal
upgrade
driver
adapter
data
use
militia
cd
external


one
time
like
bit
chip
number
problem
use
even
two
memory
get
well
test
fast
know
mode
work
speed
using
line
could
much
anyone
would
sound
look
used
also
way
second
value
since
first
idea
need
read
serial
current
actually


game
team
year
player
hockey
last
good
play
point
run
think
would
hit
well
time
first
nhl
fan
season
win
one
league
two
baseball
home
stephanopoulos
pick
played
next
back
great
goal
best
score
get
better
got
made
wing
right

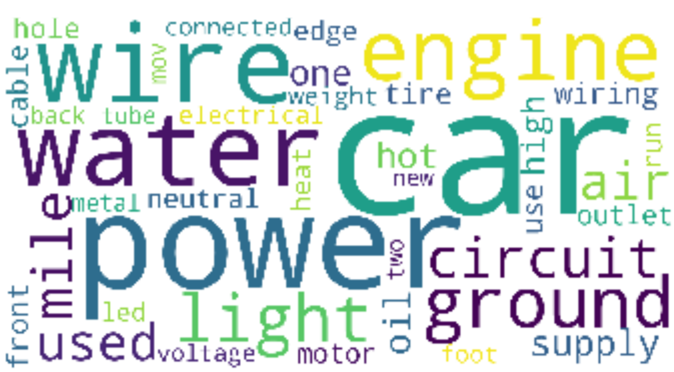

window
version
program
graphic
software
server
system
color
use
using
widget
interface
support
display
video
file
ram
machine
edu
work
mac
set
mouse
sun
application
available
driver
motif
also
run
font
code
card
running
monitor
get
problem
user
package
screen

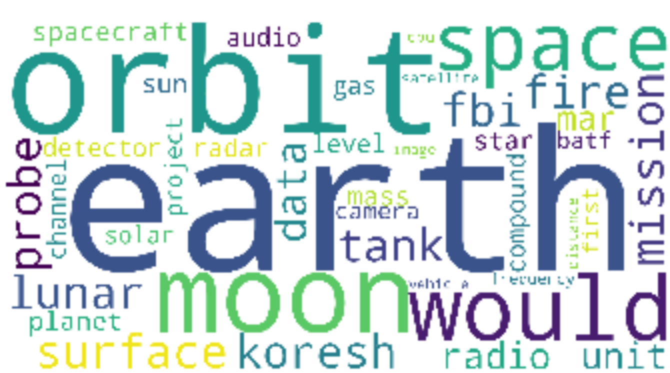

file
entry
program
image
format
output
section
gif
line
use
directory
may
input
number
size
contest
info
read
command
include
convert
code
jpeg
byte
int
source
title
author
return
data
stream
null
name
information
character
must
rule
build
text
char

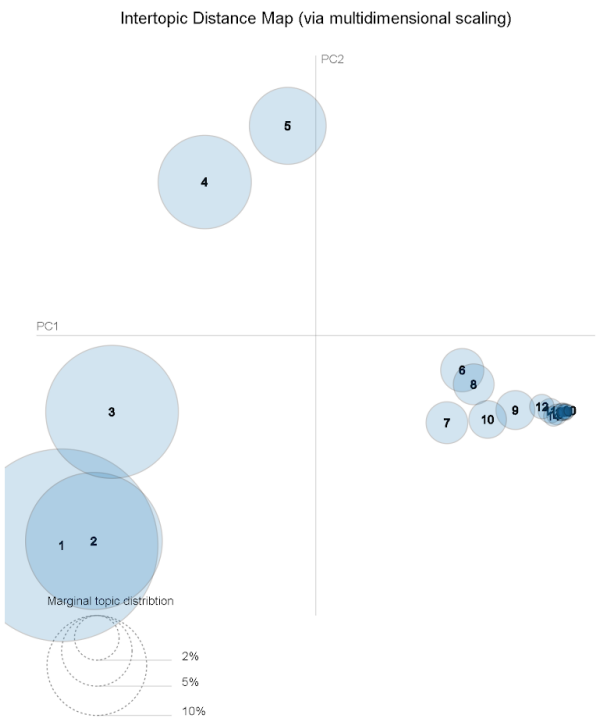

would
people
one
think
case
reason
point
could
make
well
question
mean
person
human
many
however
example
know
evidence
belief
thing
right
argument
law
even
said
statement
time
true
may
way
see
like
life
claim
believe
also
say
must
fact

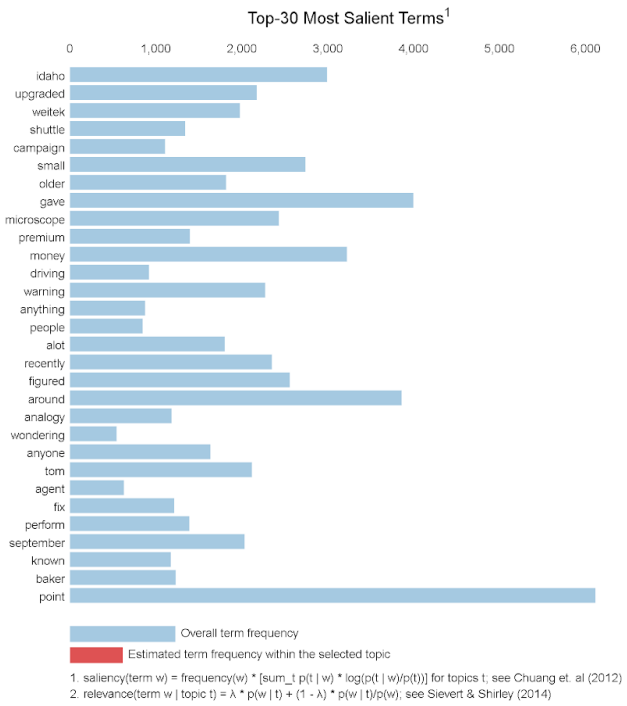

system
new
president
national
research
space
satellite
shuttle
general
office
center
also
science
united
state
launch
program
april
study
flight
technology
nasa
press
washington
american
development
available
university
year
report
group
information
organization

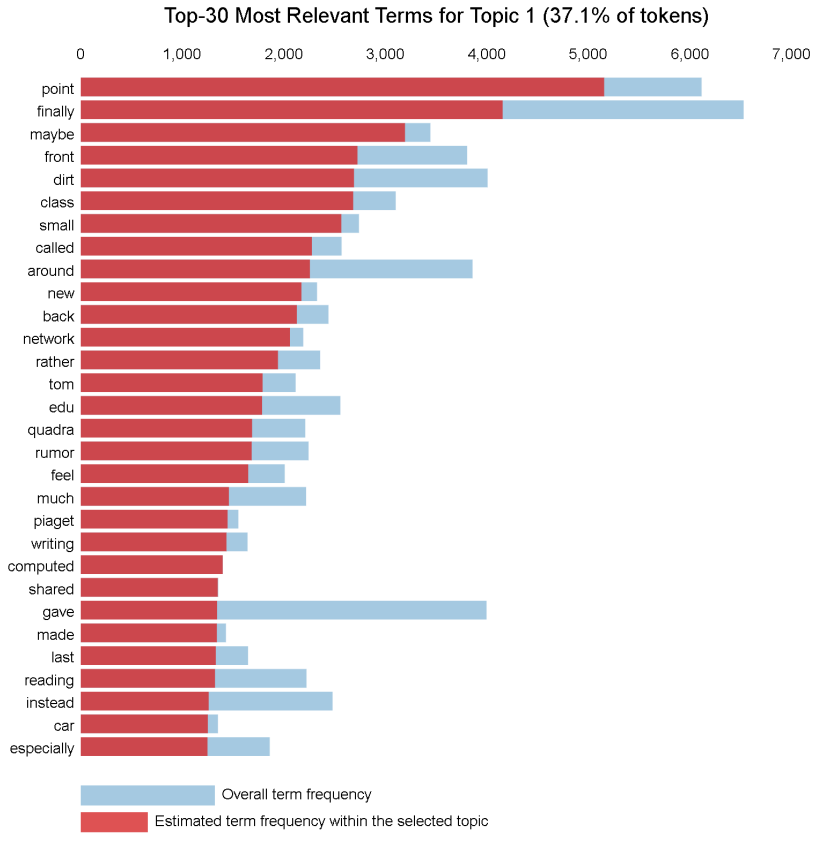

buffalo
san
boston
francisco
king
van
mon
new
edu
pittsburgh
det
pit
bos
que
andrew frank
calgary
philadelphia
period
cub
cal
toronto
louis
play
chicago
angeles
los
second
win
min
stl
power
bay
tor
league
detroit
jose
chi

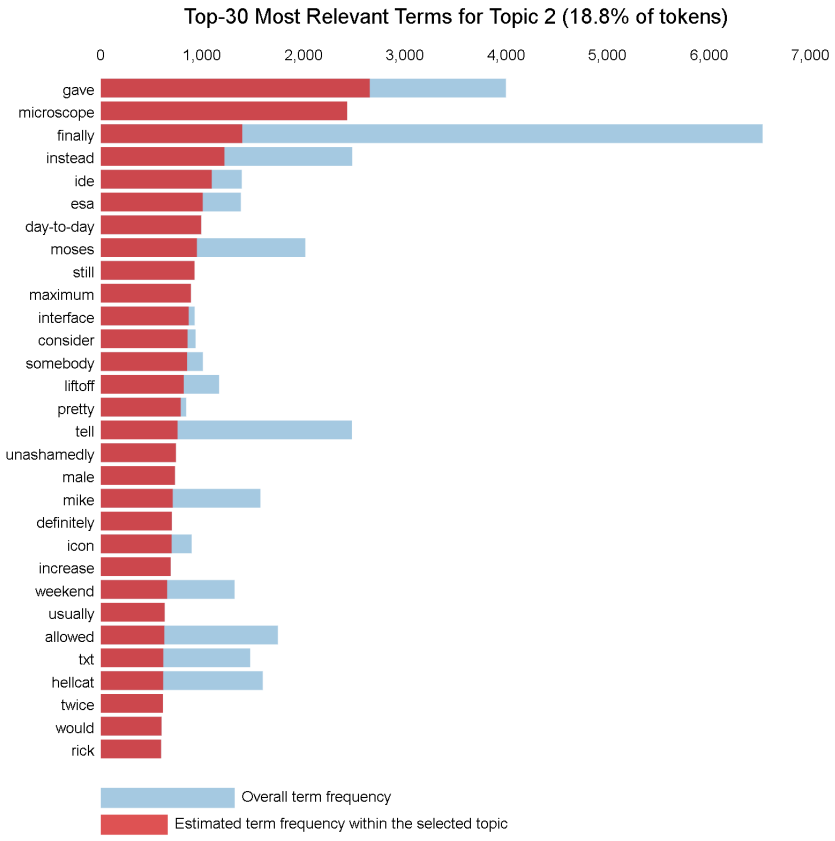

hole
connected
edge
wire
engine
cable
mov
one
weight
tire
wiring
back
tube
electrical
water
car
high
hot
run
heat
new
air
use
metal
neutral
outlet
mile
power
two
circuit
front
led
light
ground
oil
used
voltage
motor
foot
supply

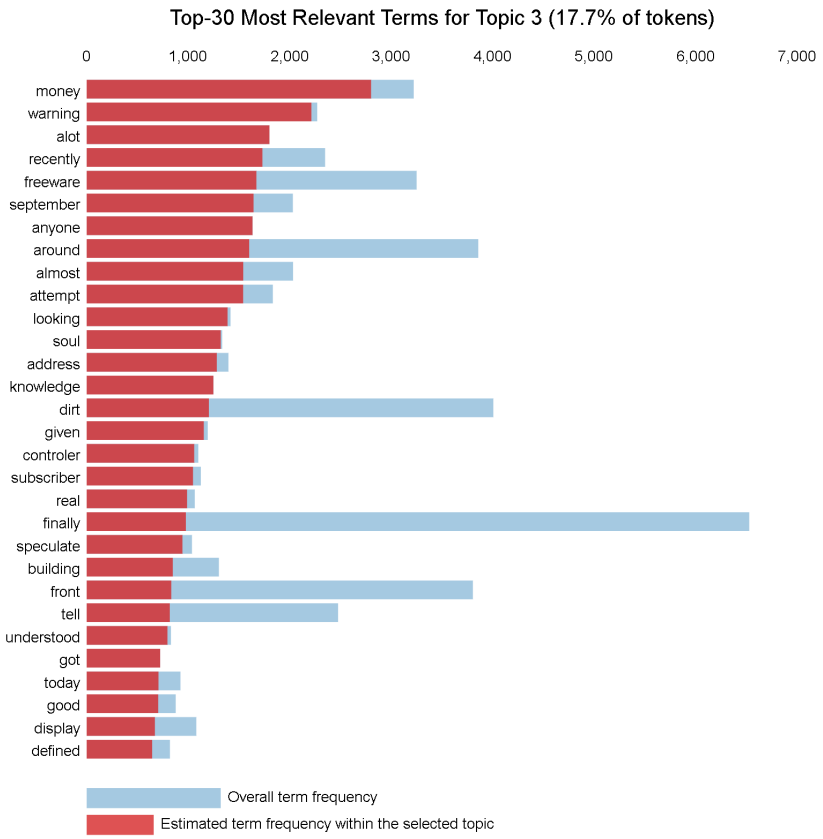

spacecraft
audio
orbit
space
sun
gas
fire
fbi
mar
mission
probe
detector
radar
level
star
batf
project
data
earth
channel
mass
camera
solar
compound
first
tank
lunar
moon
would
planet
surface
koresh
radio
unit

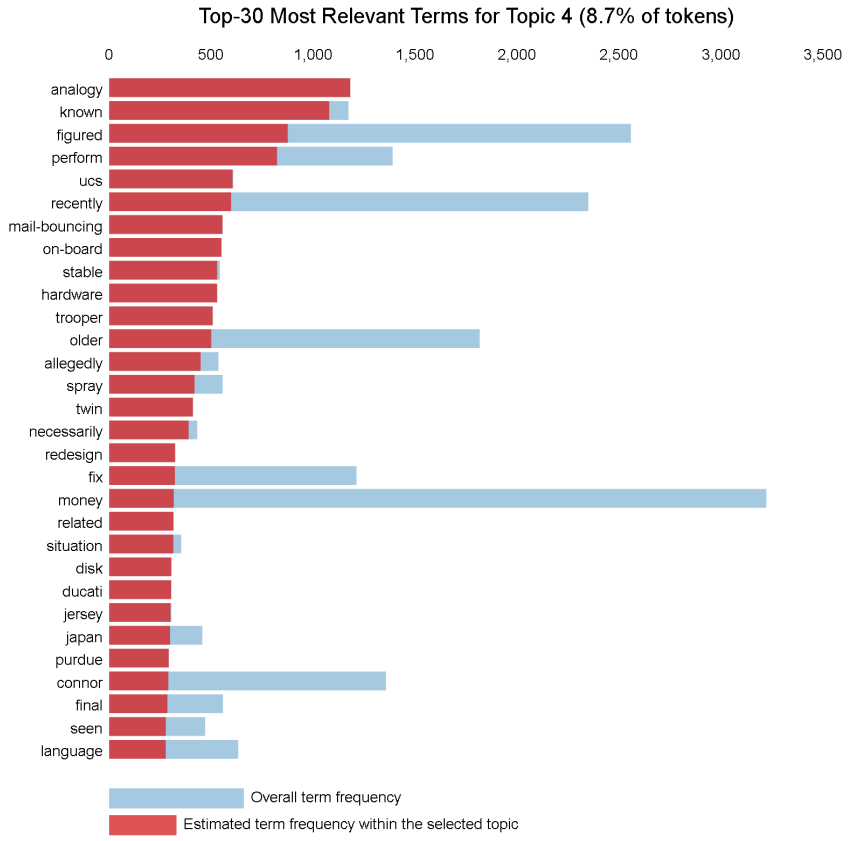

Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
1
2
3
4
5
6
7
8
9
10
12
Marginal topic distribtion
2%
5%
10%

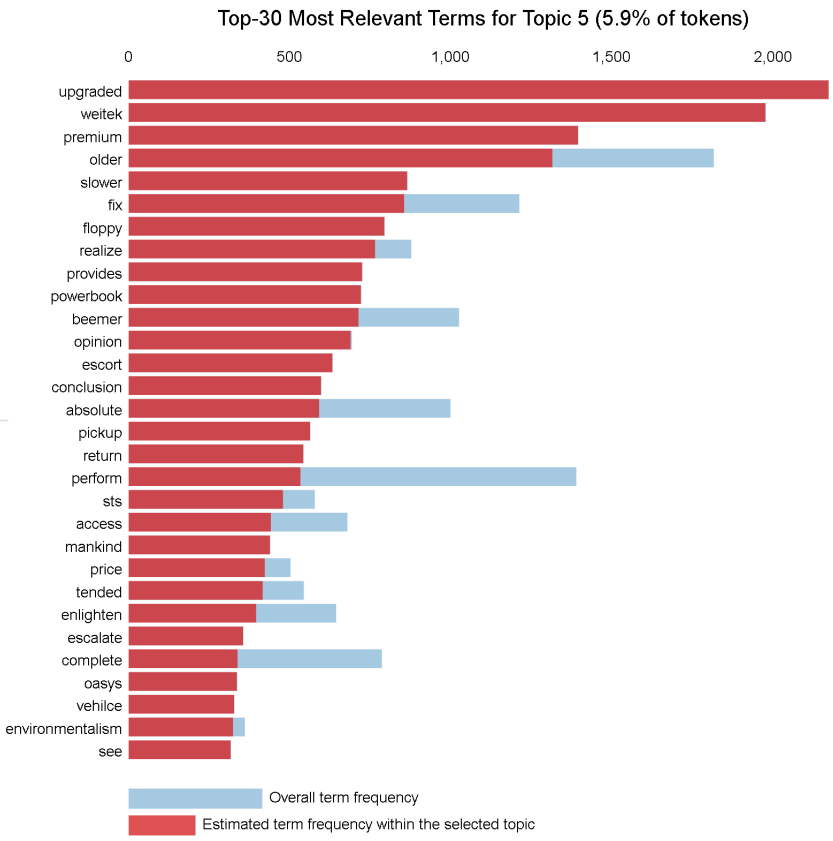

Top-30 Most Salient Terms[1]
0
1,000
2,000
3,000
4,000
5,000
6,000
idaho
upgraded
weitek
shuttle
campaign
small
older
gave
microscope
premium
money
driving
warning
anything
people
alot
recently
figured
around
analogy
wondering
anyone
tom
agent
fix
perform
september
known
baker
point
Overall term frequency
Estimated term frequency within the selected topic
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

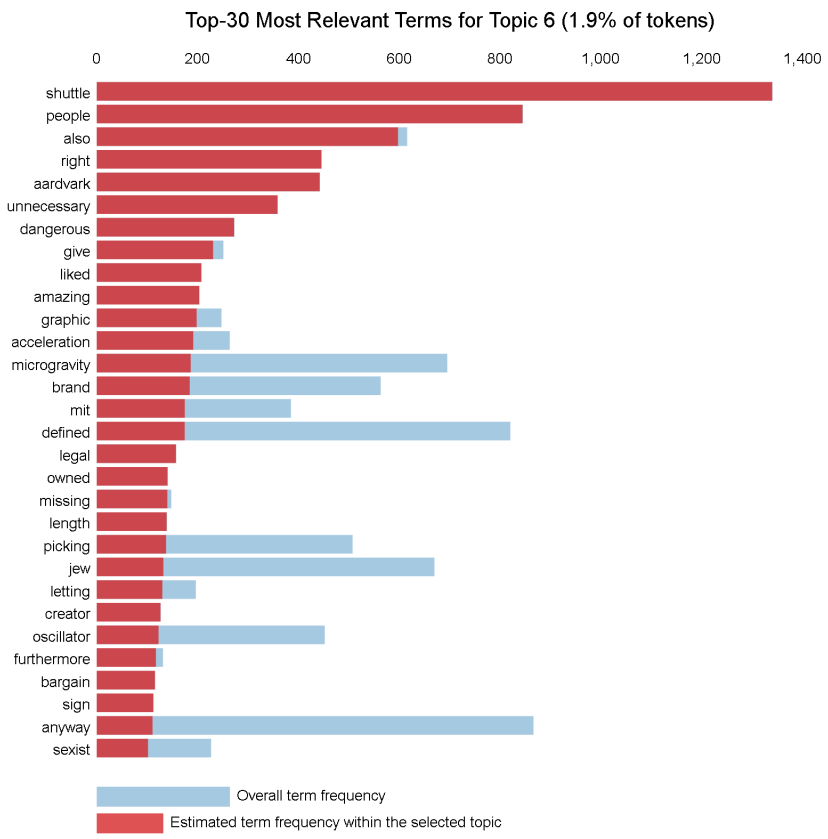

Top-30 Most Relevant Terms for Topic 1 (37.1% of tokens)
0
1,000
2,000
3,000
4,000
5,000
6,000
7,000
point
finally
maybe
front
dirt
class
small
called
around
new
back
network
rather
tom
edu
quadra
rumor
feel
much
piaget
writing
computed
shared
gave
made
last
reading
instead
car
especially
Overall term frequency
Estimated term frequency within the selected topic

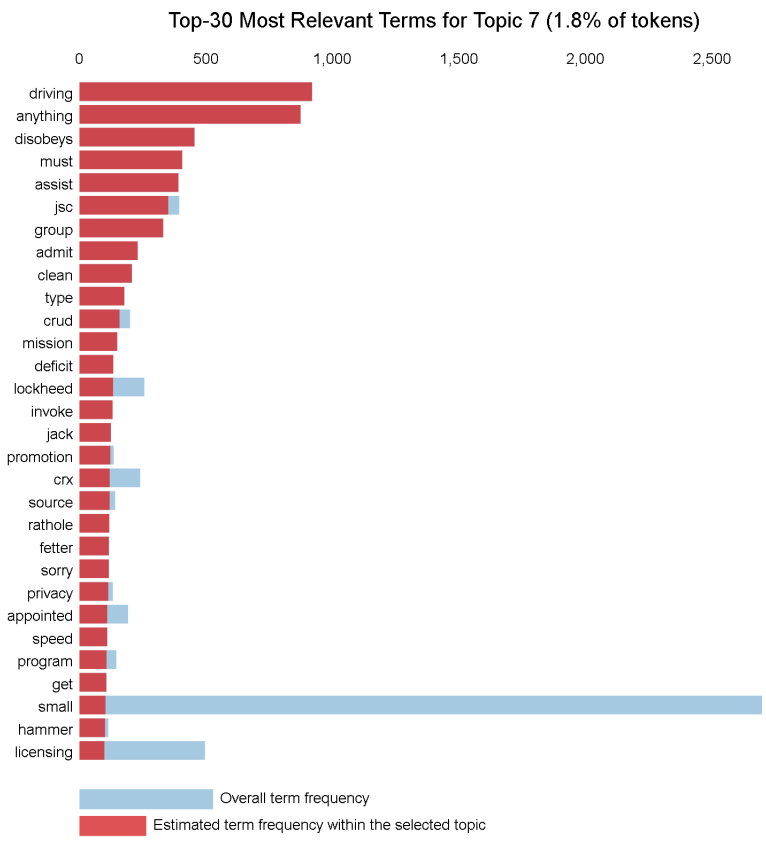

Top-30 Most Relevant Terms for Topic 2 (18.8% of tokens)
0
1,000
2,000
3,000
4,000
5,000
6,000
7,000
gave
microscope
finally
instead
ide
esa
day-to-day
moses
still
maximum
interface
consider
somebody
liftoff
pretty
tell
unashamedly
male
mike
definitely
icon
increase
weekend
usually
allowed
txt
hellcat
twice
would
rick
Overall term frequency
Estimated term frequency within the selected topic

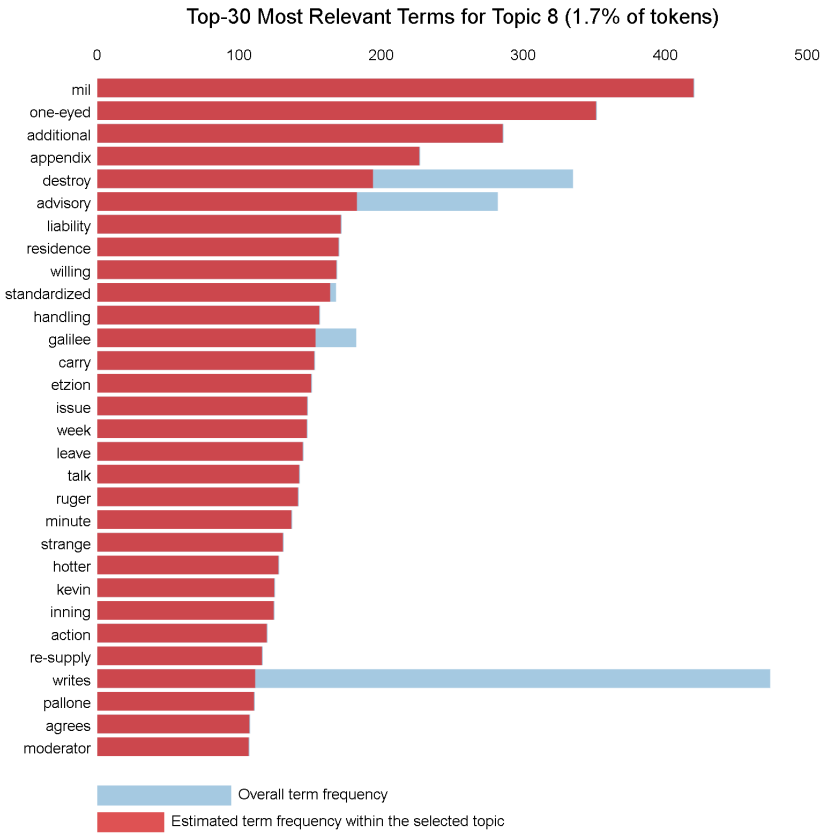

Top-30 Most Relevant Terms for Topic 3 (17.7% of tokens)
0
1,000
2,000
3,000
4,000
5,000
6,000
7,000
money
warning
alot
recently
freeware
september
anyone
around
almost
attempt
looking
soul
address
knowledge
dirt
given
controler
subscriber
real
finally
speculate
building
front
tell
understood
got
today
good
display
defined
Overall term frequency
Estimated term frequency within the selected topic

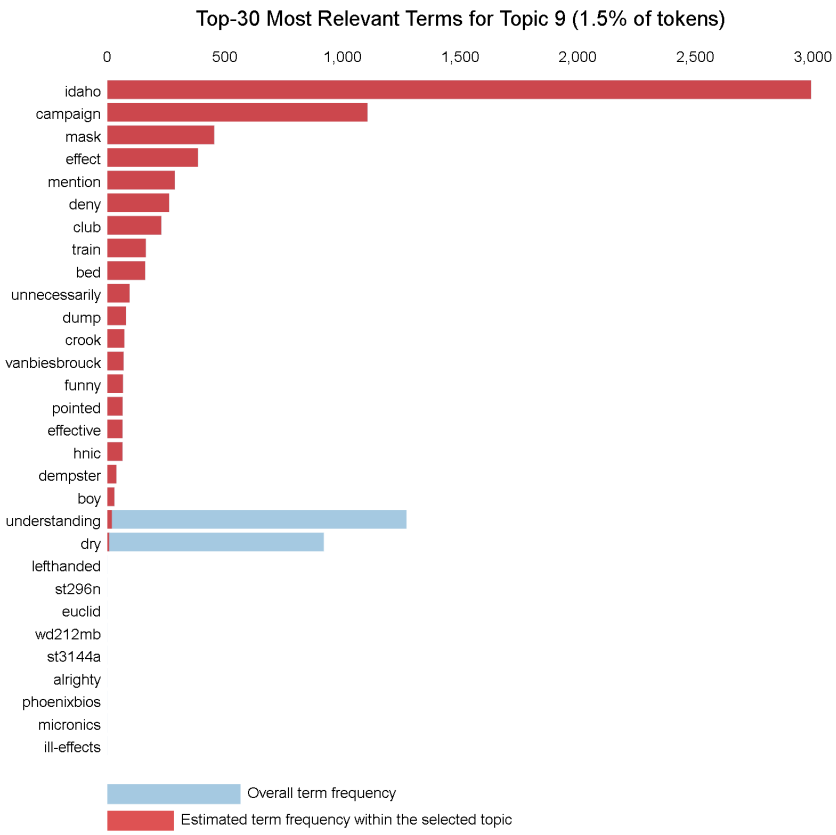

Top-30 Most Relevant Terms for Topic 4 (8.7% of tokens)
0
500
1,000
1,500
2,000
2,500
3,000
3,500
analogy
known
figured
perform
ucs
recently
mail-bouncing
on-board
stable
hardware
trooper
older
allegedly
spray
twin
necessarily
redesign
fix
money
related
situation
disk
ducati
jersey
japan
purdue
connor
final
seen
language
Overall term frequency
Estimated term frequency within the selected topic

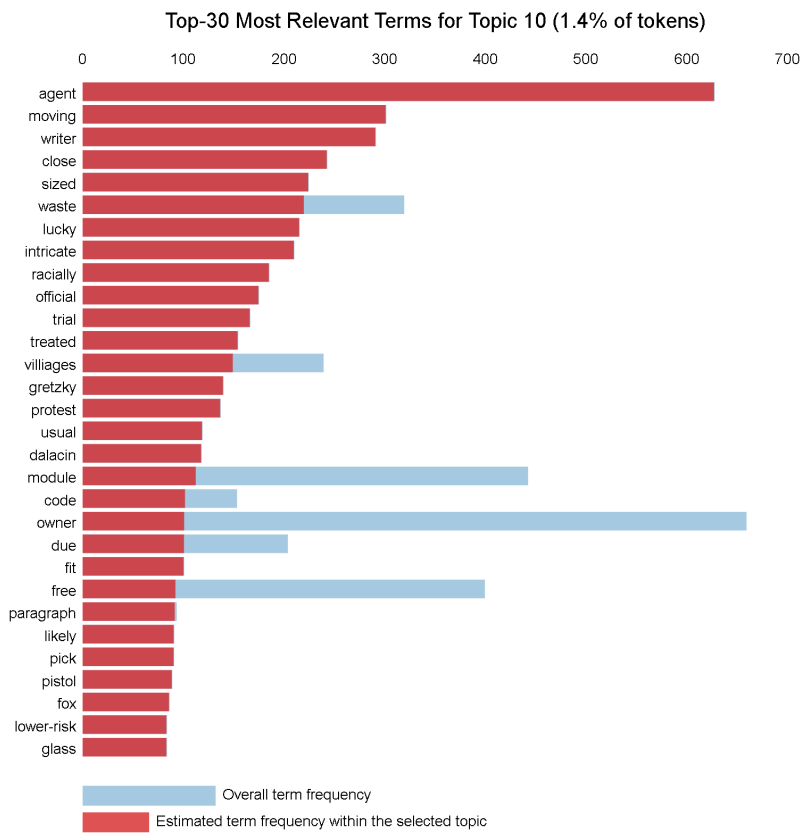

Top-30 Most Relevant Terms for Topic 5 (5.9% of tokens)
0
500
1,000
1,500
2,000
upgraded
weitek
premium
older
slower
fix
floppy
realize
provides
powerbook
beemer
opinion
escort
conclusion
absolute
pickup
return
perform
sts
access
mankind
price
tended
enlighten
escalate
complete
oasys
vehilce
environmentalism
see
Overall term frequency
Estimated term frequency within the selected topic

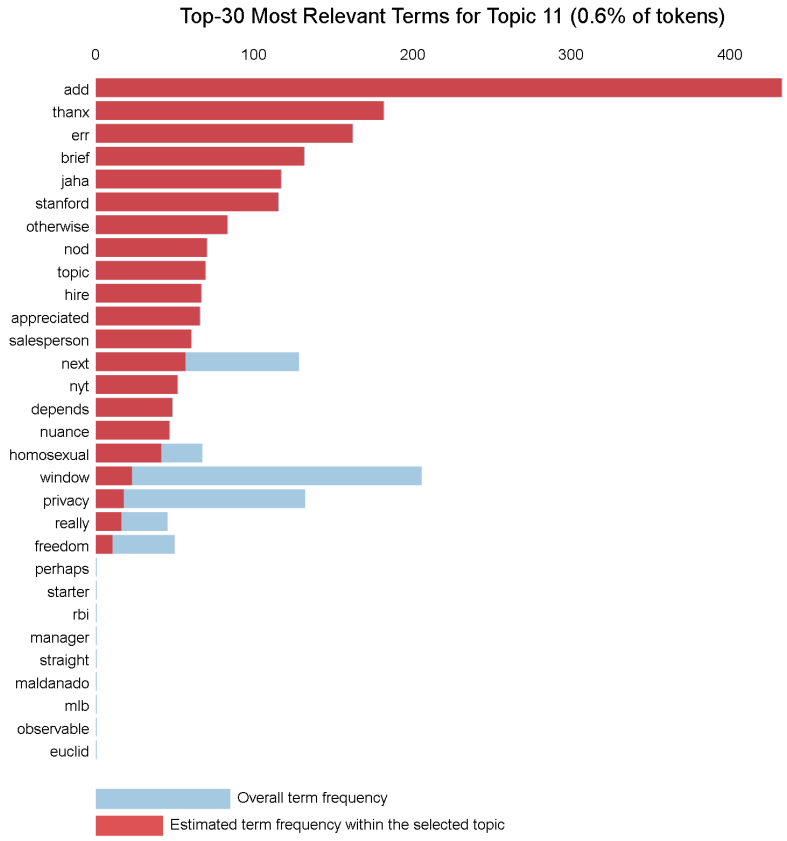

Top-30 Most Relevant Terms for Topic 6 (1.9% of tokens)
0
200
400
600
800
1,000
1,200
1,400
shuttle
people
also
right
aardvark
unnecessary
dangerous
give
liked
amazing
graphic
acceleration
microgravity
brand
mit
defined
legal
owned
missing
length
picking
jew
letting
creator
oscillator
furthermore
bargain
sign
anyway
sexist
Overall term frequency
Estimated term frequency within the selected topic

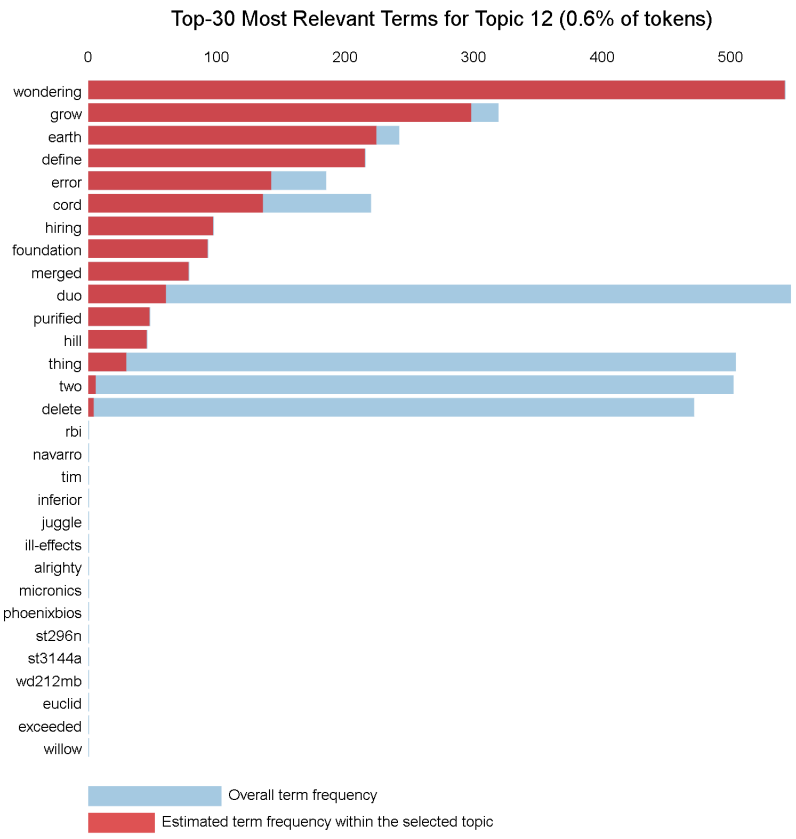

Top-30 Most Relevant Terms for Topic 7 (1.8% of tokens)
0
500
1,000
1,500
2,000
2,500
driving
anything
disobeys
must
assist
jsc
group
admit
clean
type
crud
mission
deficit
lockheed
invoke
jack
promotion
crx
source
rathole
fetter
sorry
privacy
appointed
speed
program
get
small
hammer
licensing
Overall term frequency
Estimated term frequency within the selected topic

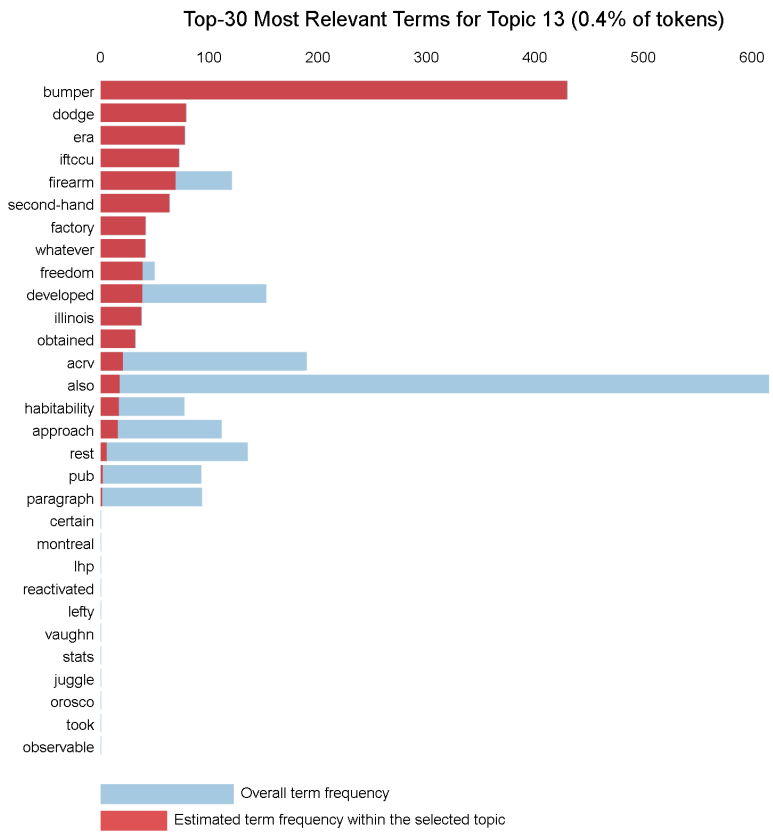

Top-30 Most Relevant Terms for Topic 8 (1.7% of tokens)
0
100
200
300
400
500
mil
one-eyed
additional
appendix
destroy
advisory
liability
residence
willing
standardized
handling
galilee
carry
etzion
issue
week
leave
talk
ruger
minute
strange
hotter
kevin
inning
action
re-supply
writes
pallone
agrees
moderator
Overall term frequency
Estimated term frequency within the selected topic

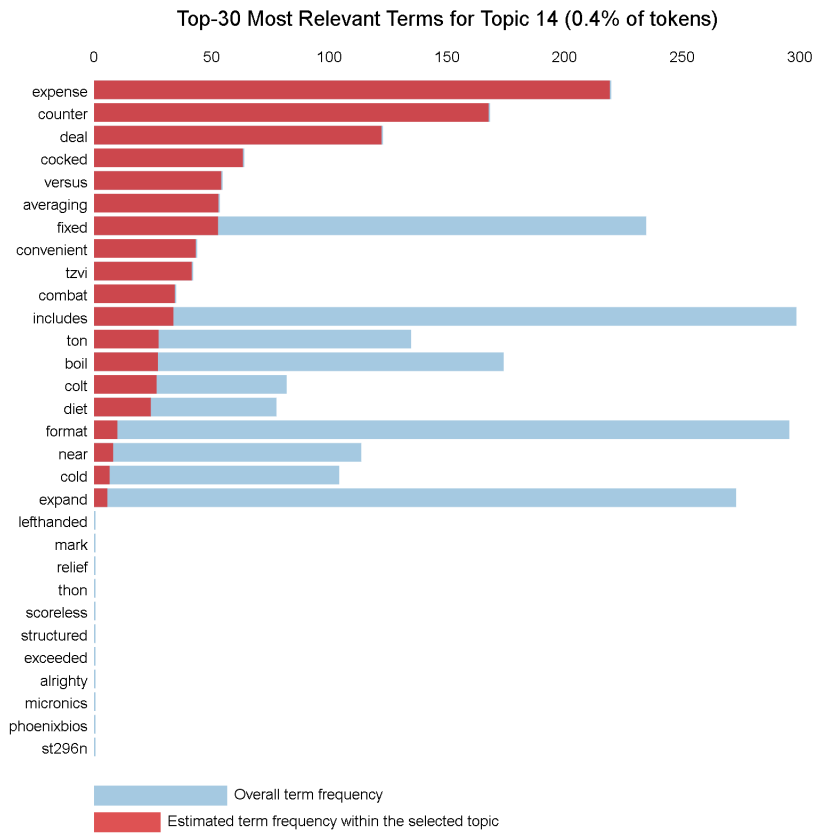

Top-30 Most Relevant Terms for Topic 9 (1.5% of tokens)
0
500
1,000
1,500
2,000
2,500
3,000
idaho
campaign
mask
effect
mention
deny
club
train
bed
unnecessarily
dump
crook
vanbiesbrouck
funny
pointed
effective
hnic
dempster
boy
understanding
dry
lefthanded
st296n
euclid
wd212mb
st3144a
alrighty
phoenixbios
micronics
ill-effects
Overall term frequency
Estimated term frequency within the selected topic

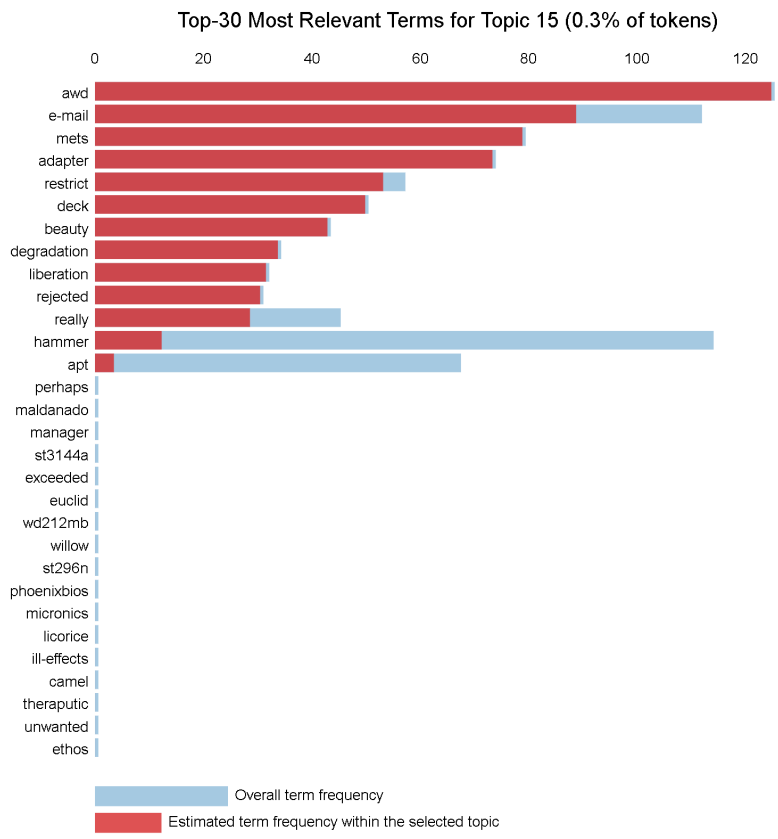

Top-30 Most Relevant Terms for Topic 10 (1.4% of tokens)
0
100
200
300
400
500
600
700
agent
moving
writer
close
sized
waste
lucky
intricate
racially
official
trial
treated
villiages
gretzky
protest
usual
dalacin
module
code
owner
due
fit
free
paragraph
likely
pick
pistol
fox
lower-risk
glass
Overall term frequency
Estimated term frequency within the selected topic

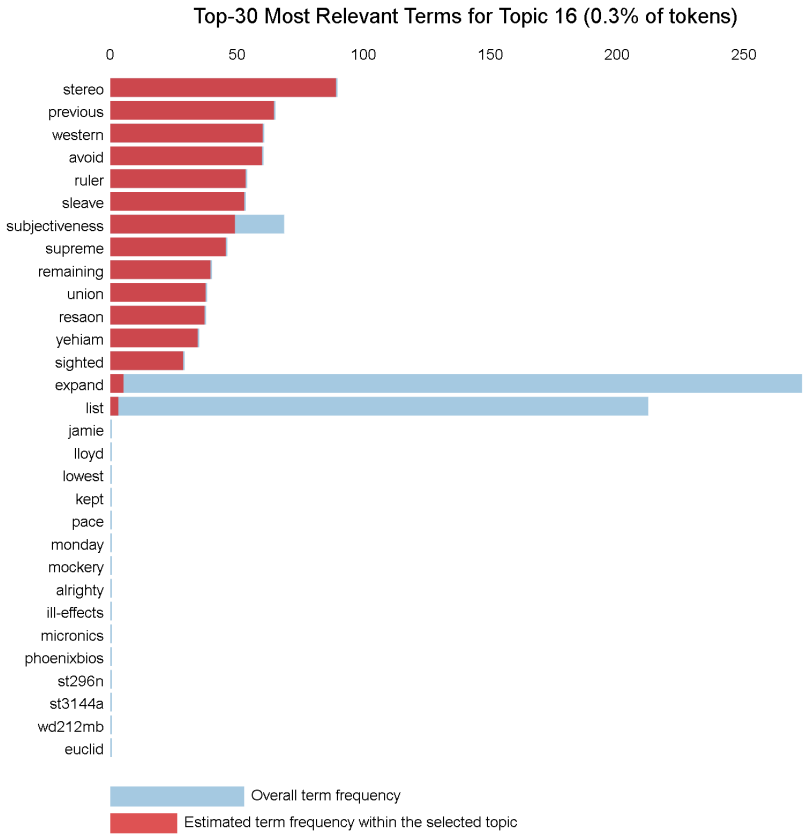

Top-30 Most Relevant Terms for Topic 11 (0.6% of tokens)
0
100
200
300
400
add
thanx
err
brief
jaha
stanford
otherwise
nod
topic
hire
appreciated
salesperson
next
nyt
depends
nuance
homosexual
window
privacy
really
freedom
perhaps
starter
rbi
manager
straight
maldanado
mlb
observable
euclid
Overall term frequency
Estimated term frequency within the selected topic

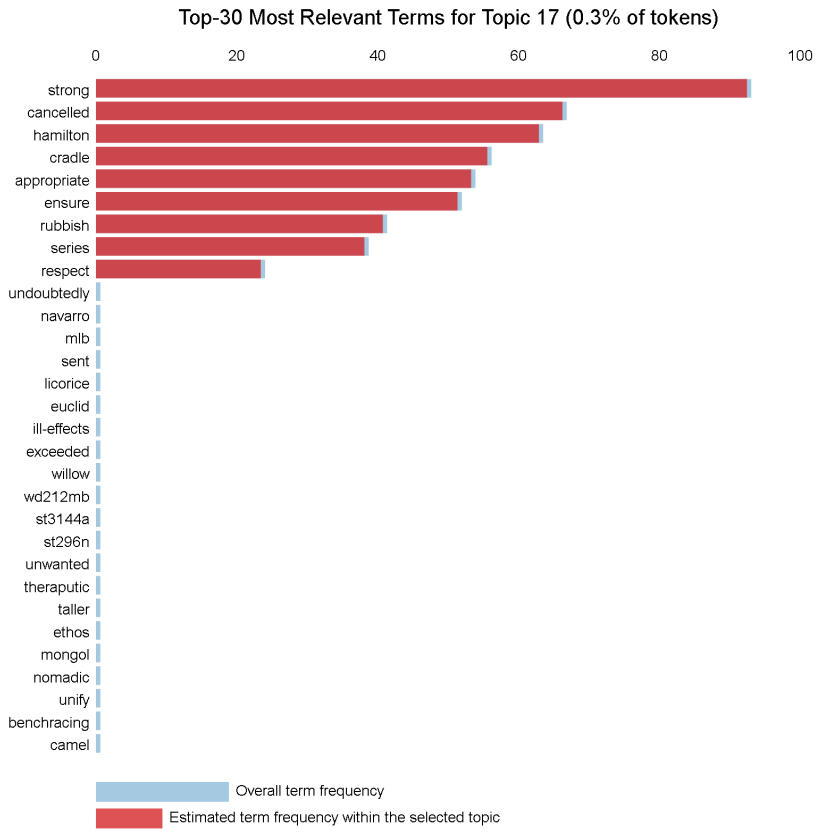

Top-30 Most Relevant Terms for Topic 12 (0.6% of tokens)
0
100
200
300
400
500
wondering
grow
earth
define
error
cord
hiring
foundation
merged
duo
purified
hill
thing
two
delete
rbi
navarro
tim
inferior
juggle
ill-effects
alrighty
micronics
phoenixbios
st296n
st3144a
wd212mb
euclid
exceeded
willow
Overall term frequency
Estimated term frequency within the selected topic

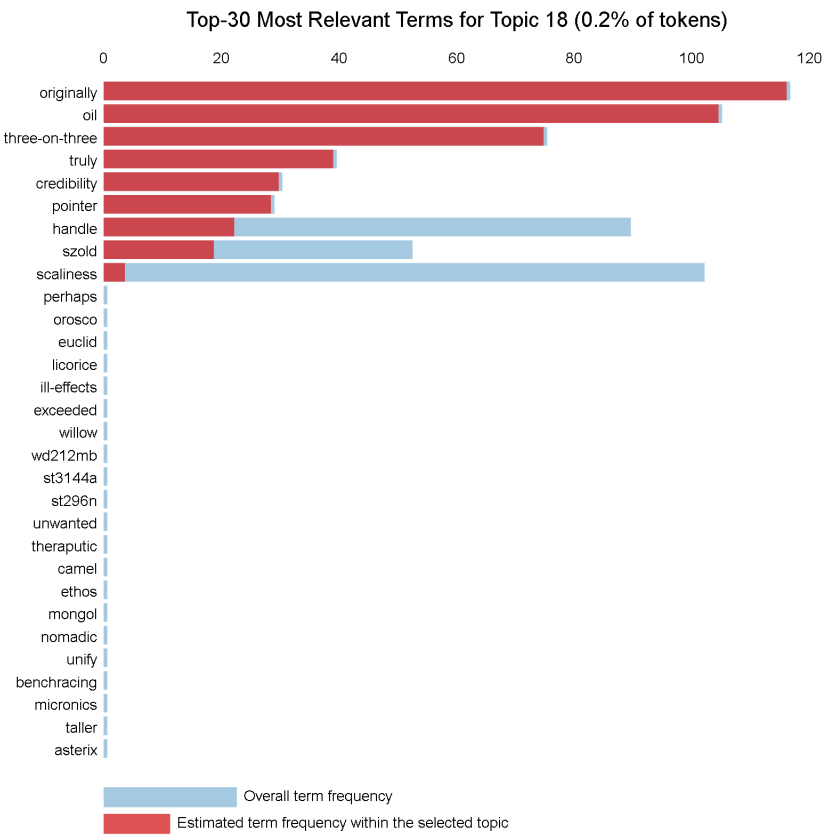

Top-30 Most Relevant Terms for Topic 13 (0.4% of tokens)
0
100
200
300
400
500
600
bumper
dodge
era
iftccu
firearm
second-hand
factory
whatever
freedom
developed
illinois
obtained
acrv
also
habitability
approach
rest
pub
paragraph
certain
montreal
lhp
reactivated
lefty
vaughn
stats
juggle
orosco
took
observable
Overall term frequency
Estimated term frequency within the selected topic

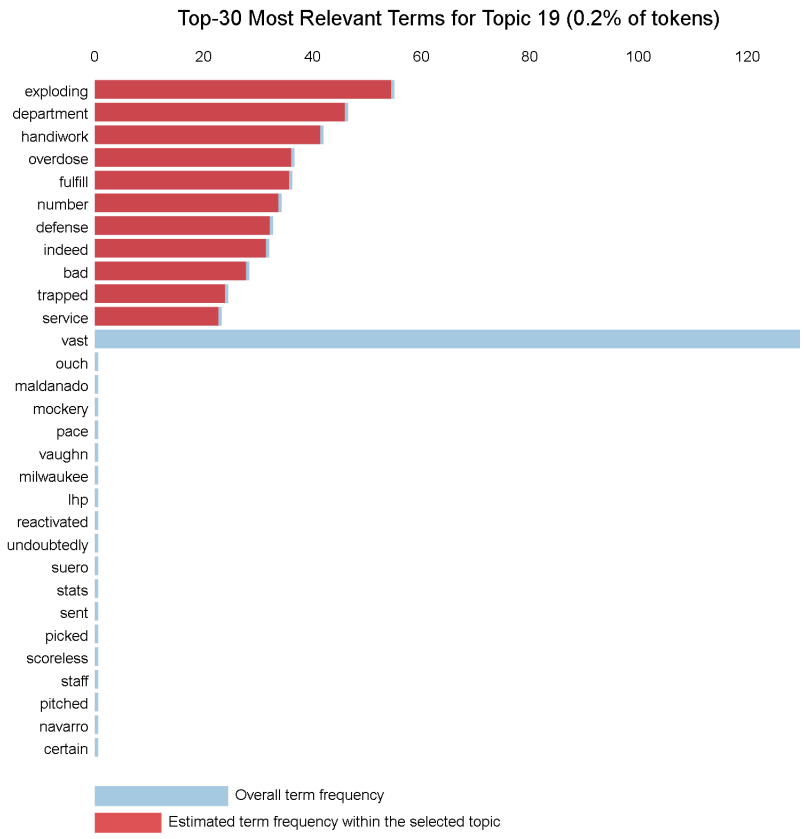

Top-30 Most Relevant Terms for Topic 14 (0.4% of tokens)
0
50
100
150
200
250
300
expense
counter
deal
cocked
versus
averaging
fixed
convenient
tzvi
combat
includes
ton
boil
colt
diet
format
near
cold
expand
lefthanded
mark
relief
thon
scoreless
structured
exceeded
alrighty
micronics
phoenixbios
st296n
Overall term frequency
Estimated term frequency within the selected topic

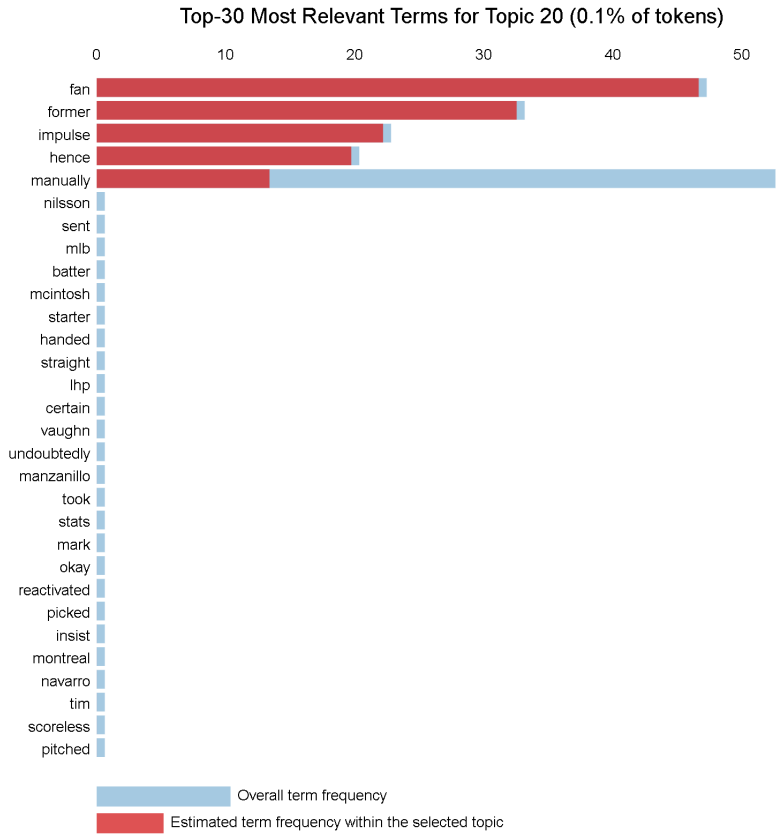

Top-30 Most Relevant Terms for Topic 15 (0.3% of tokens)
0
20
40
60
80
100
120
awd
e-mail
mets
adapter
restrict
deck
beauty
degradation
liberation
rejected
really
hammer
apt
perhaps
maldanado
manager
st3144a
exceeded
euclid
wd212mb
willow
st296n
phoenixbios
micronics
licorice
ill-effects
camel
theraputic
unwanted
ethos
Overall term frequency
Estimated term frequency within the selected topic


Top-30 Most Relevant Terms for Topic 16 (0.3% of tokens)
0
50
100
150
200
250
stereo
previous
western
avoid
ruler
sleave
subjectiveness
supreme
remaining
union
resaon
yehiam
sighted
expand
list
jamie
lloyd
lowest
kept
pace
monday
mockery
alrighty
ill-effects
micronics
phoenixbios
st296n
st3144a
wd212mb
euclid
Overall term frequency
Estimated term frequency within the selected topic


Top-30 Most Relevant Terms for Topic 17 (0.3% of tokens)
0
20
40
60
80
100
strong
cancelled
hamilton
cradle
appropriate
ensure
rubbish
series
respect
undoubtedly
navarro
mlb
sent
licorice
euclid
ill-effects
exceeded
willow
wd212mb
st3144a
st296n
unwanted
theraputic
taller
ethos
mongol
nomadic
unify
benchracing
camel
Overall term frequency
Estimated term frequency within the selected topic

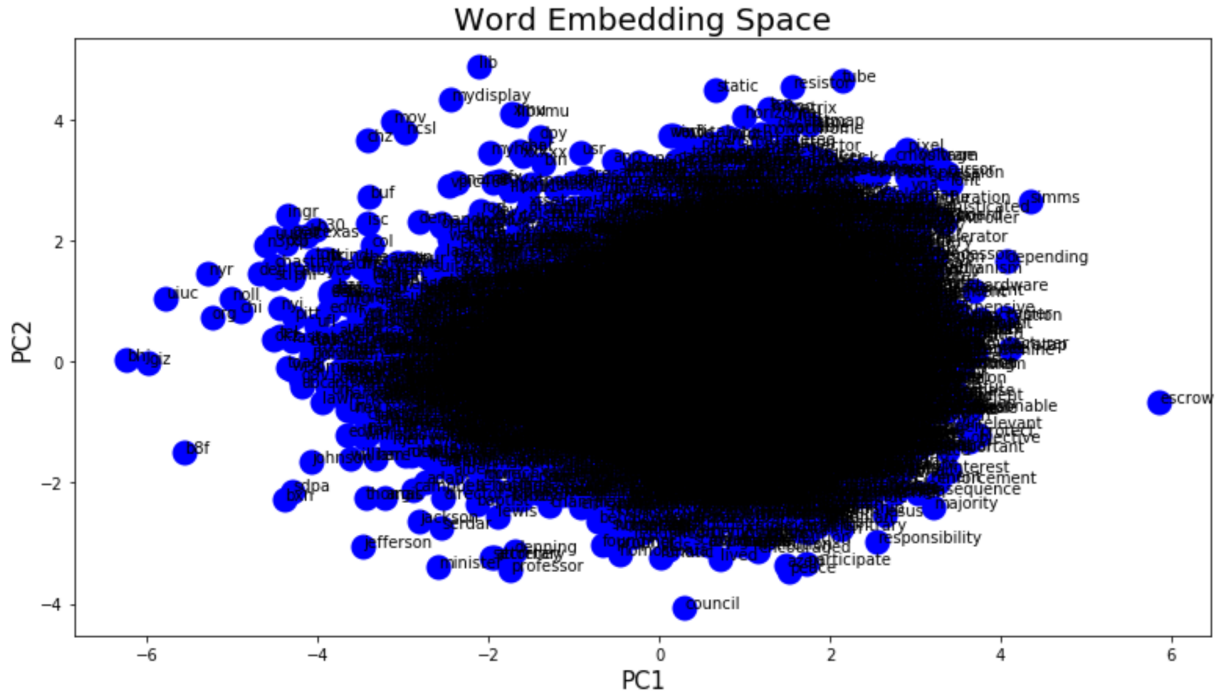

Top-30 Most Relevant Terms for Topic 18 (0.2% of tokens)
0
20
40
60
80
100
120
originally
oil
three-on-three
truly
credibility
pointer
handle
szold
scaliness
perhaps
orosco
euclid
licorice
ill-effects
exceeded
willow
wd212mb
st3144a
st296n
unwanted
theraputic
camel
ethos
mongol
nomadic
unify
benchracing
micronics
taller
asterix
Overall term frequency
Estimated term frequency within the selected topic


Top-30 Most Relevant Terms for Topic 19 (0.2% of tokens)
0
20
40
60
80
100
120
exploding
department
handiwork
overdose
fulfill
number
defense
indeed
bad
trapped
service
vast
ouch
maldanado
mockery
pace
vaughn
milwaukee
lhp
reactivated
undoubtedly
suero
stats
sent
picked
scoreless
staff
pitched
navarro
certain
Overall term frequency
Estimated term frequency within the selected topic

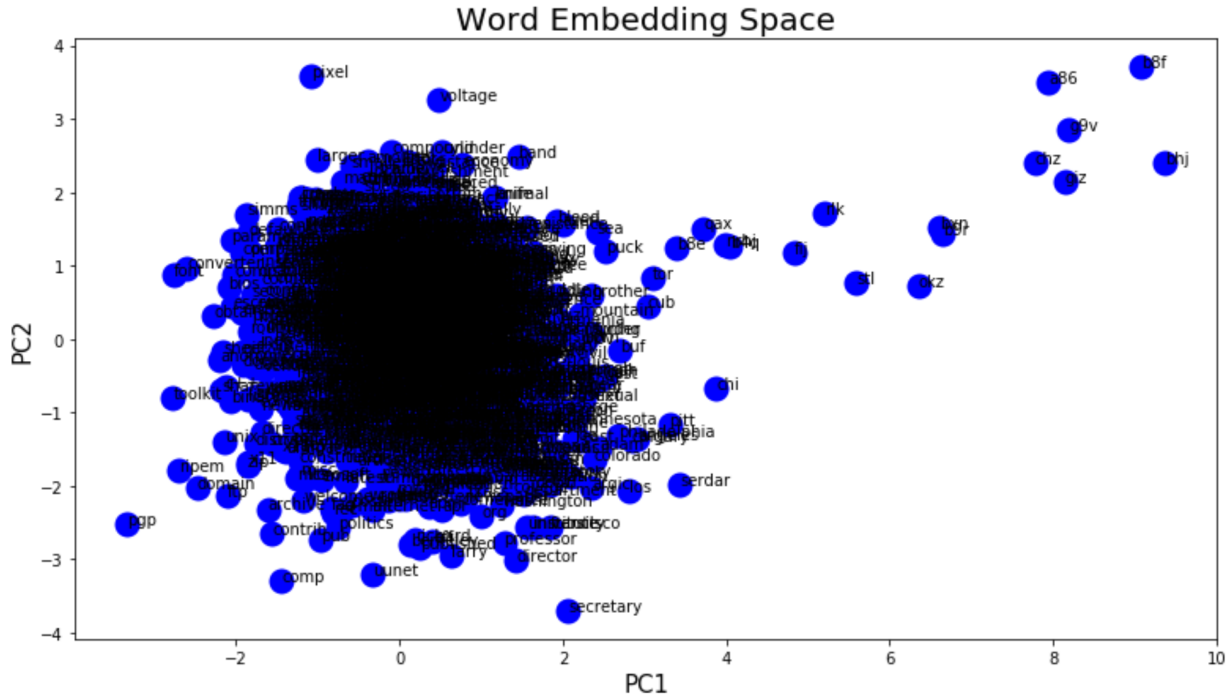

Top-30 Most Relevant Terms for Topic 20 (0.1% of tokens)
0
10
20
30
40
50
fan
former
impulse
hence
manually
nilsson
sent
mlb
batter
mcintosh
starter
handed
straight
lhp
certain
vaughn
undoubtedly
manzanillo
took
stats
mark
okay
reactivated
picked
insist
montreal
navarro
tim
scoreless
pitched
Overall term frequency
Estimated term frequency within the selected topic

Appendix F: Word2Vec Visualizations

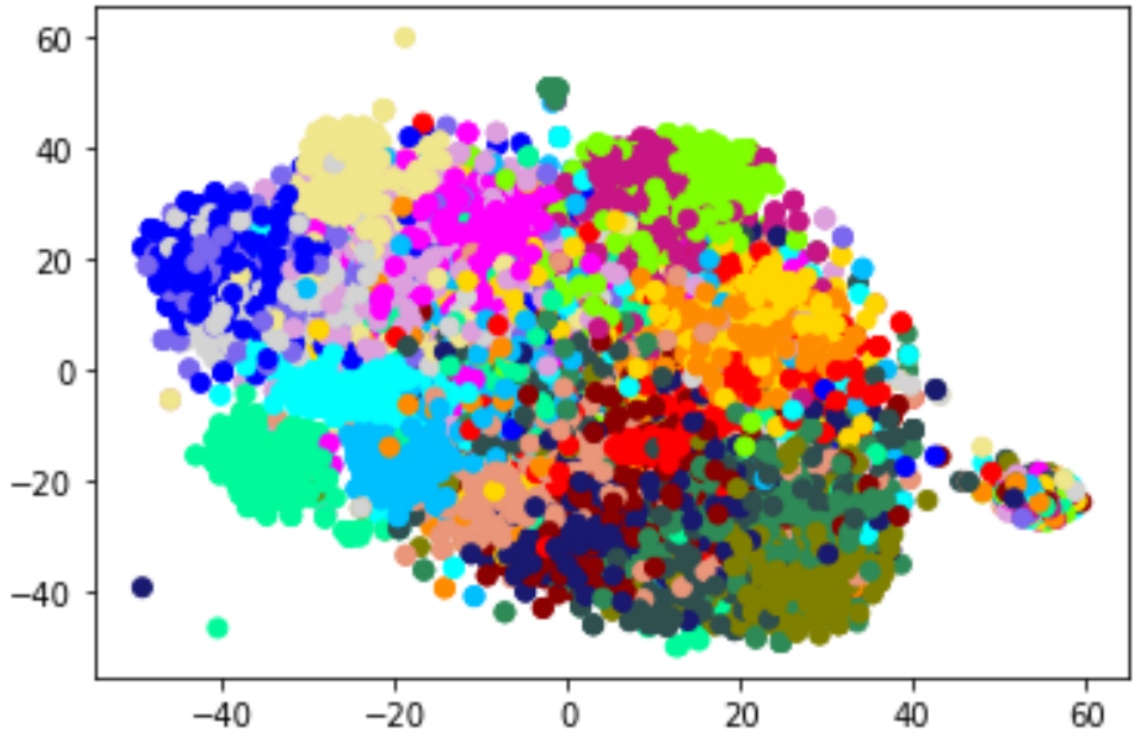
Complete vocabulary:



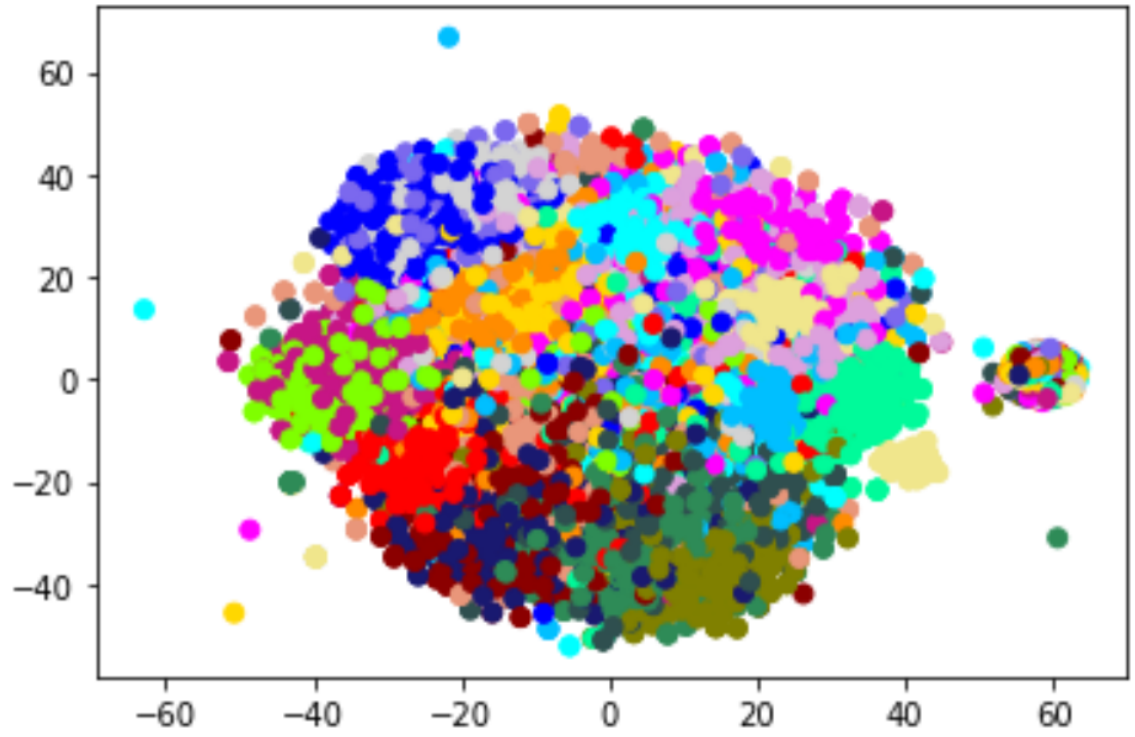
Top 2k most frequent words:



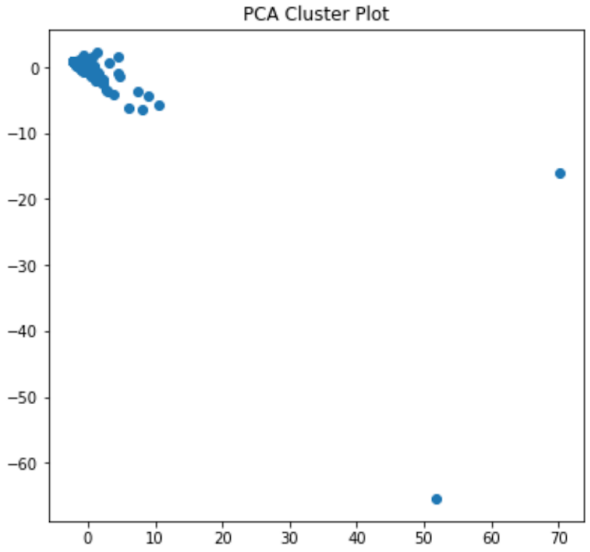
Doc2Vec for the complete dataset

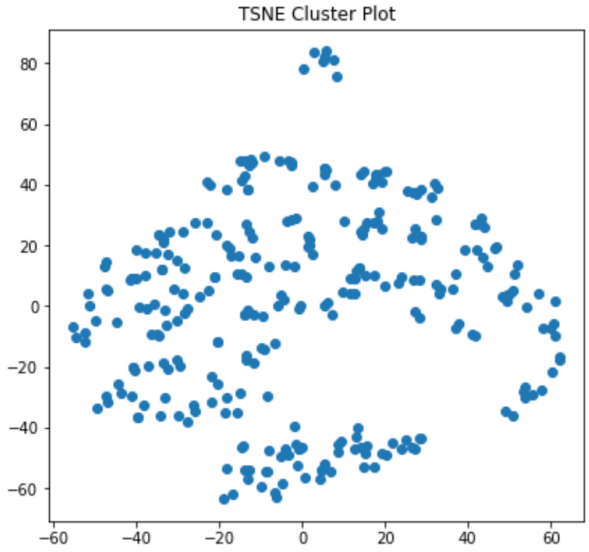


Doc2Vec for the 2k most frequent words

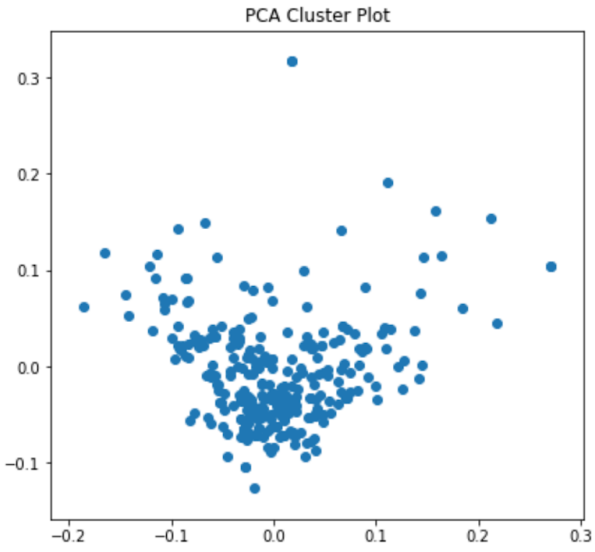


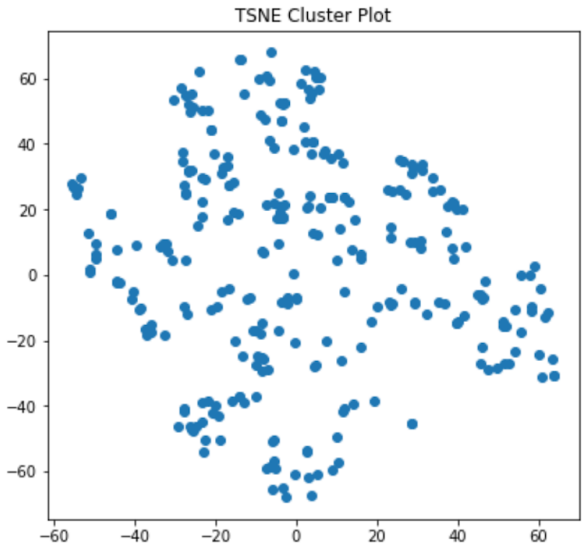
Bag-of-Words:



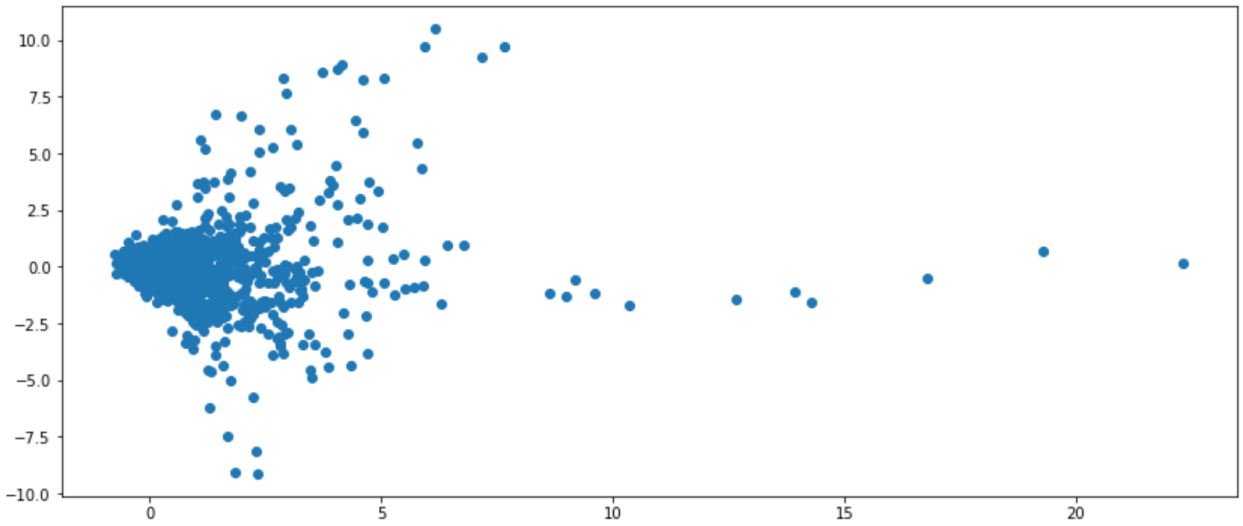


TF-IDF:

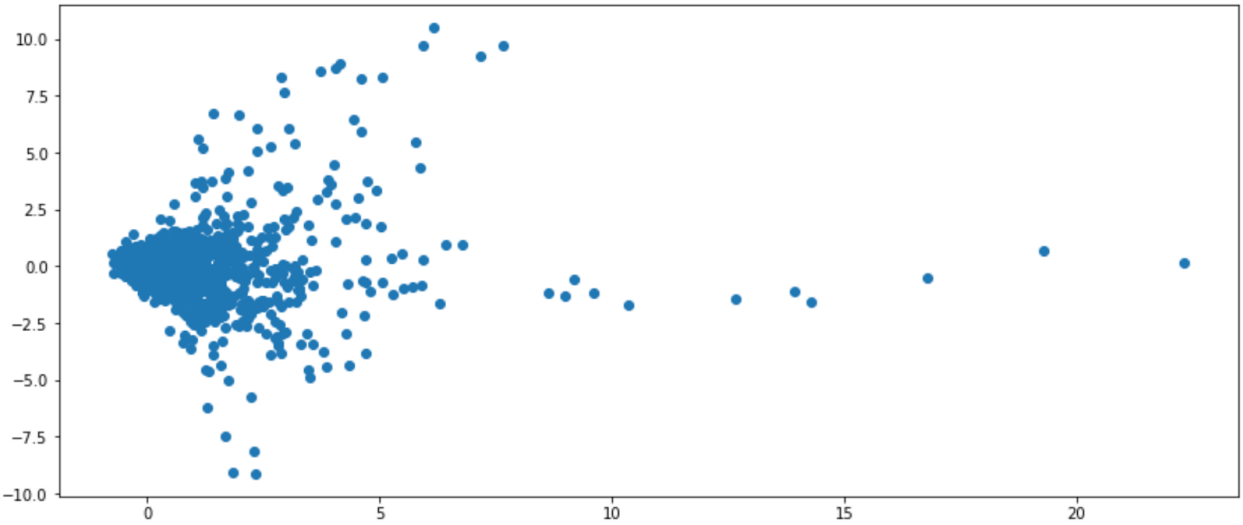




Topic Distribution:



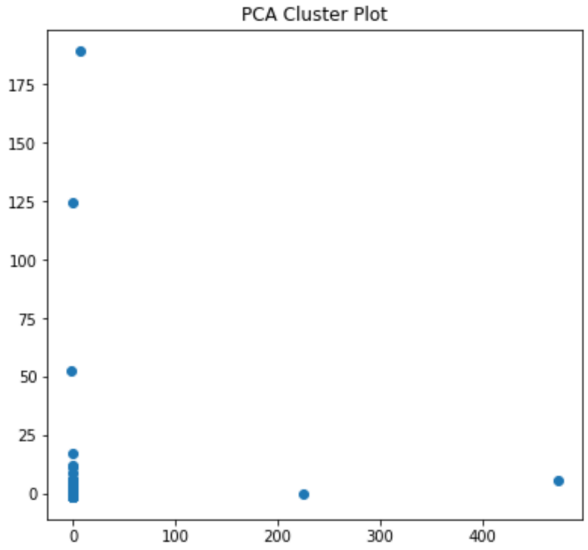
Doc2Vec:

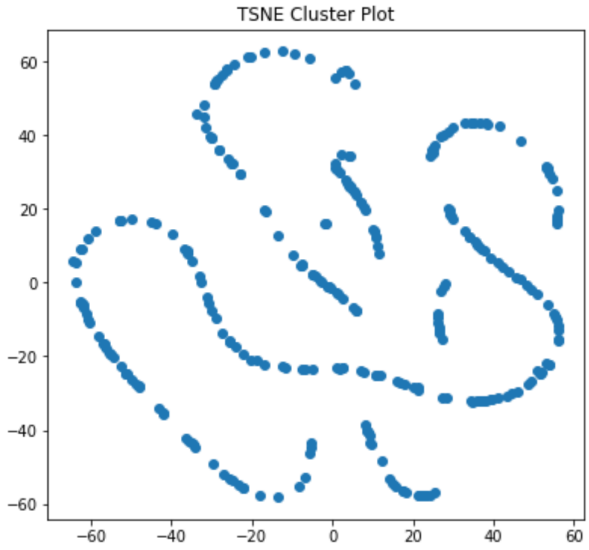


NMI Comparison:

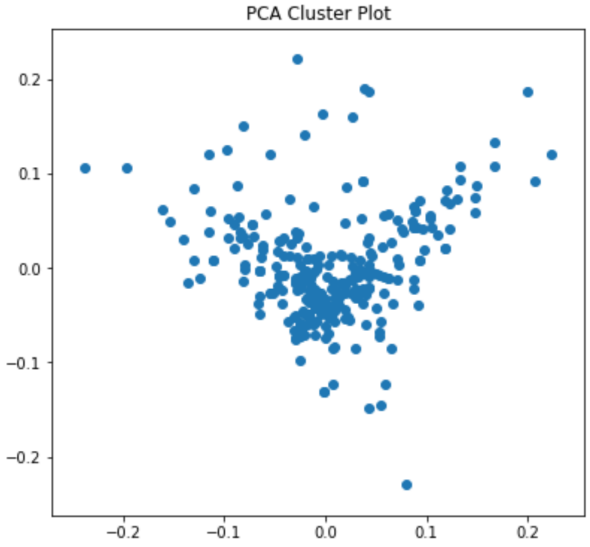
Bag Of Words	TFIDF	Topic Distribution	Doc2Vec
0.00872521360354	0.1624854029375	0.0257184904549024	0.1182999314946

Bag-of-words with 2k vocabulary:

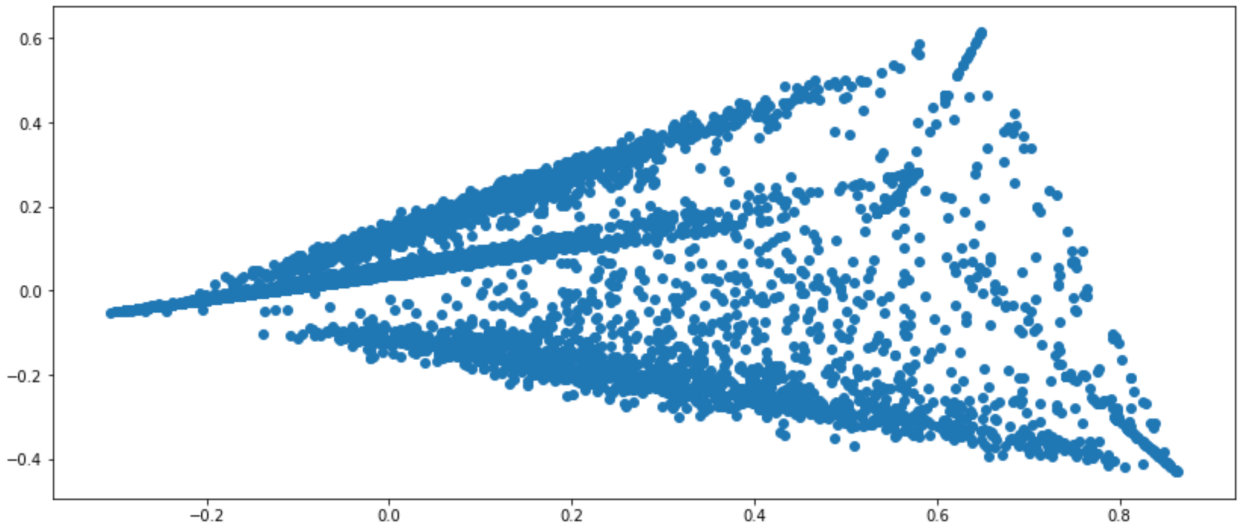




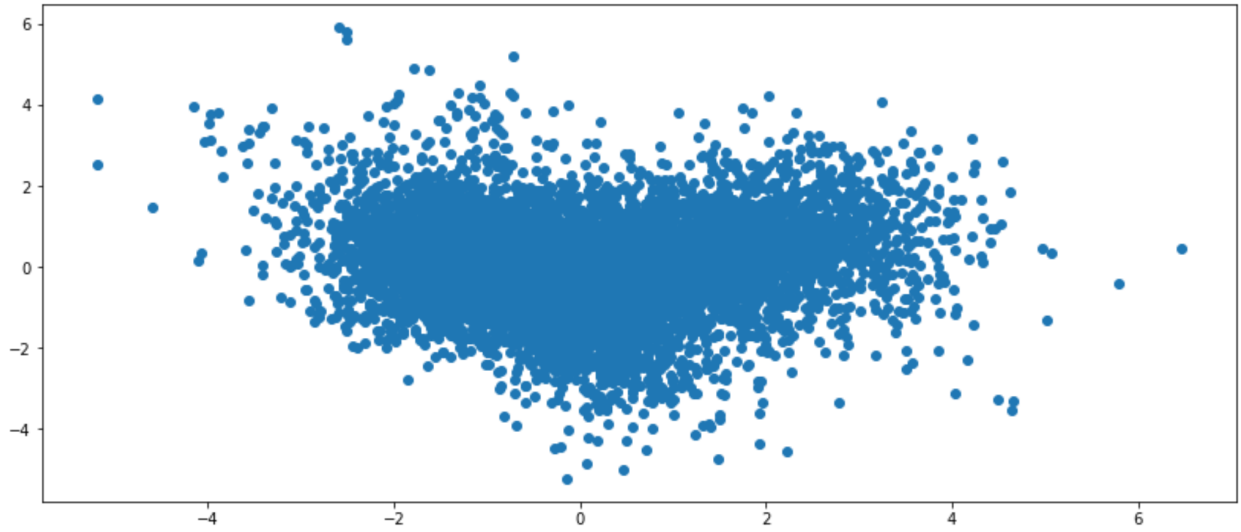
TF-IDF with 2k vocabulary:



Topic Distribution with 2k vocabulary:



Doc2Vec with 2k vocabulary:

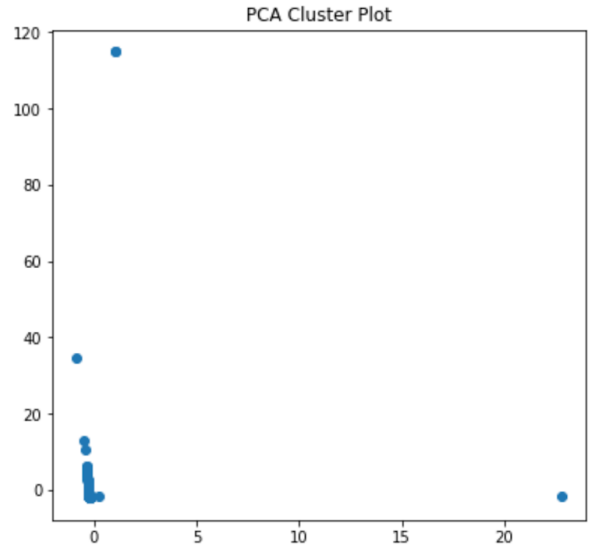


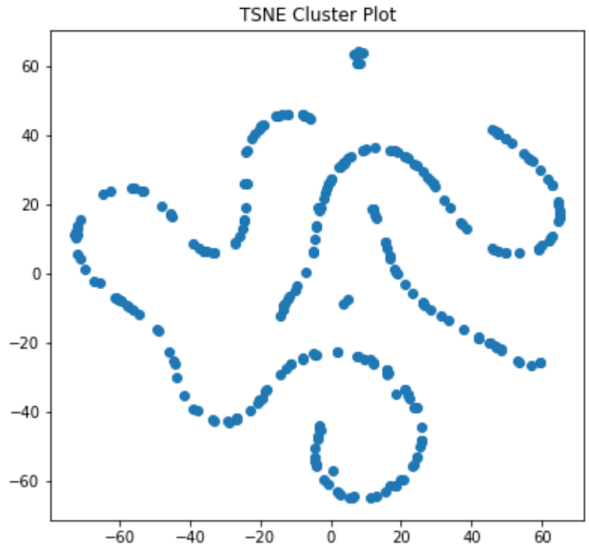
NMI comparison:

Bag Of Words	TFIDF	Topic Distribution	Doc2Vec
0.00944974065913	0.3171378125559	0.218634846381	0.393188671043

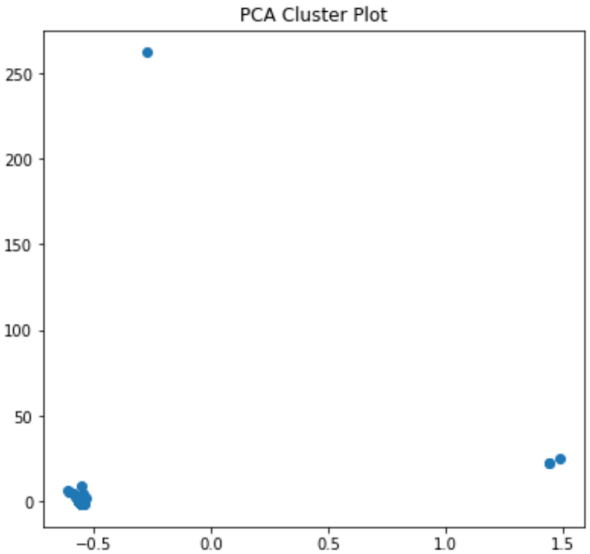
Appendix I: Results with different number of topics

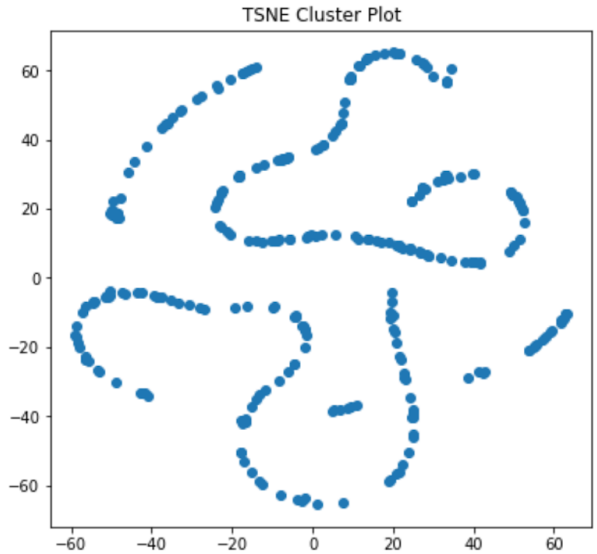
Bag-of-Words with Num_Topics / 2:



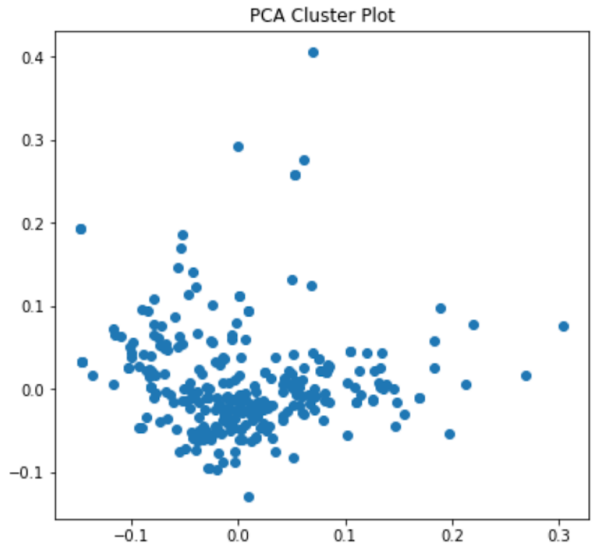


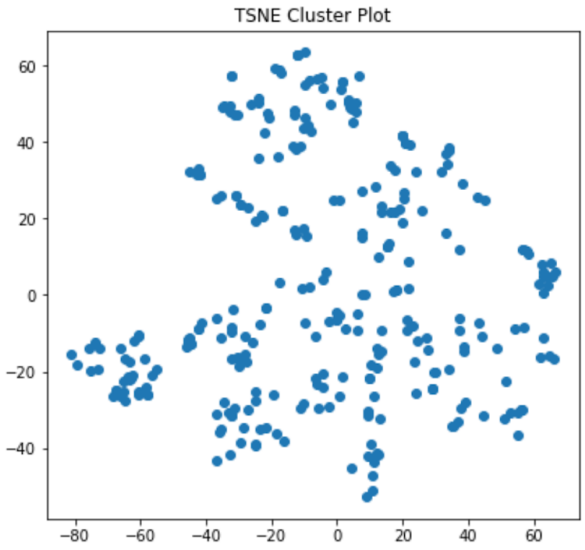
Bag-of-Words with Num_Topics * 2:



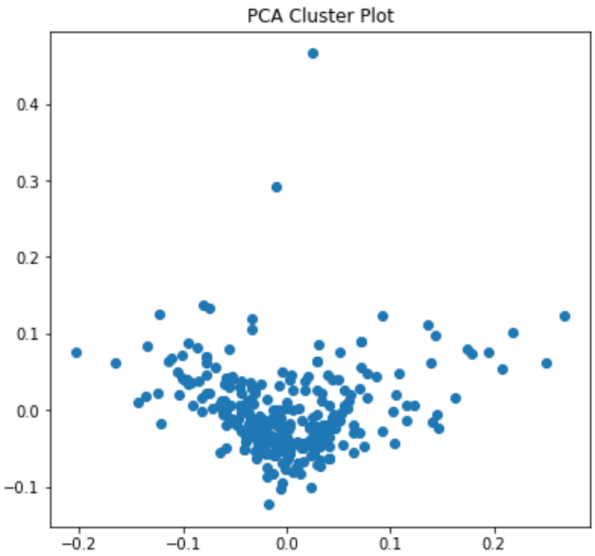


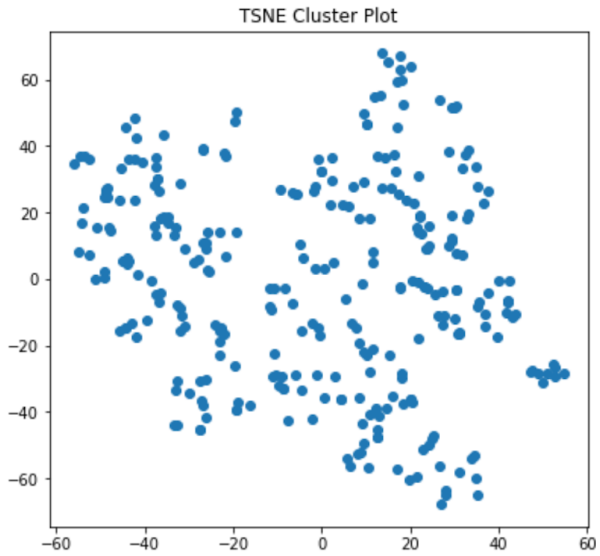
TF-IDF with Num_Topics / 2:



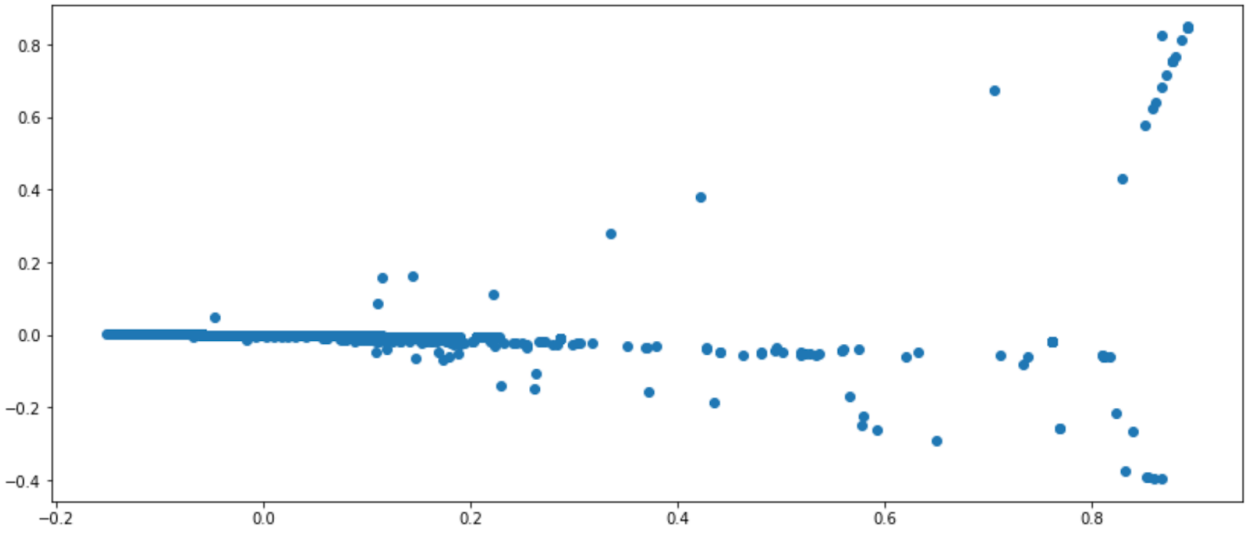


TF-IDF with Num_Topics * 2:

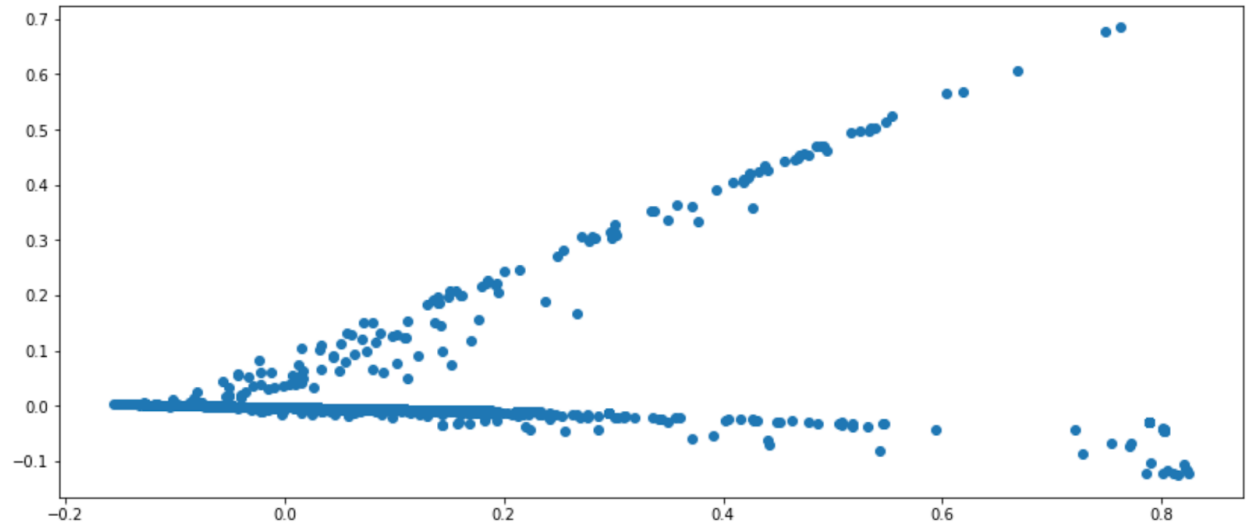




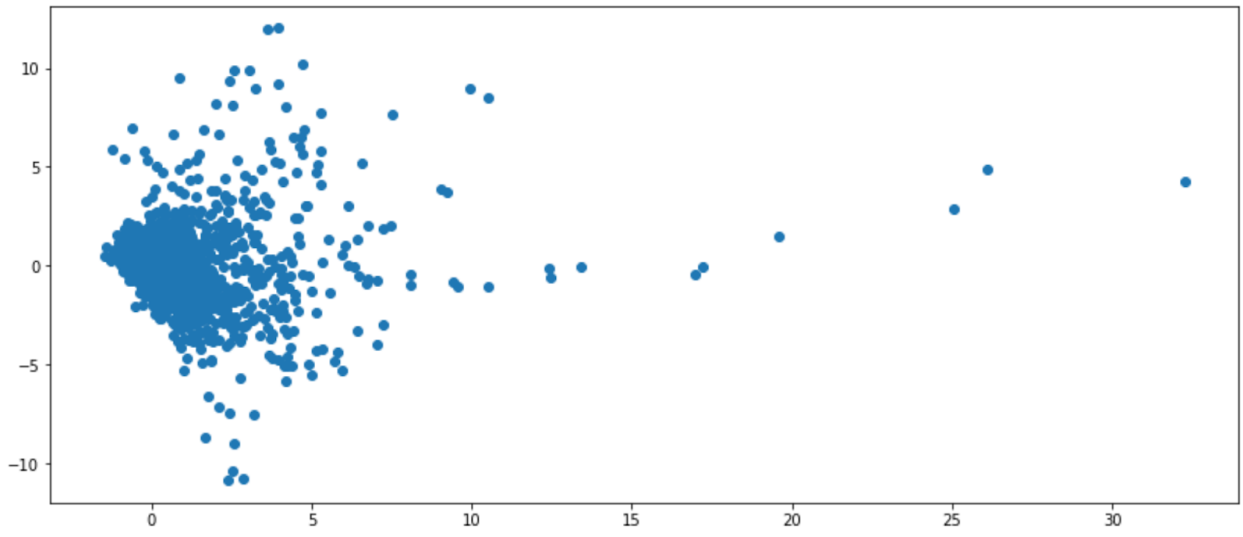
Topic Distribution with Num_Topics / 2:



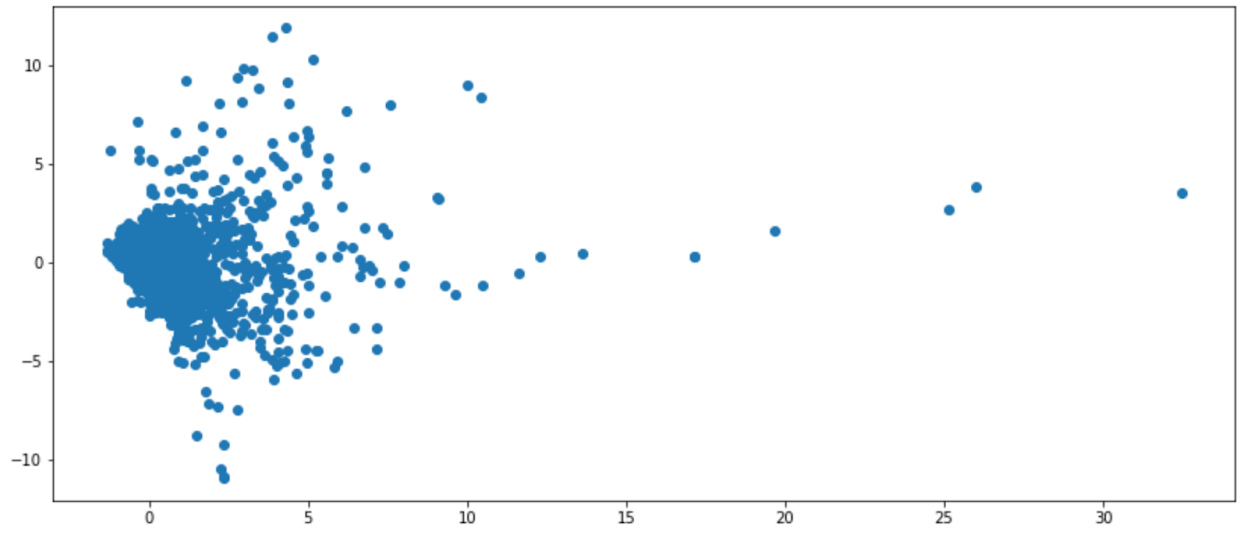
Topic Distribution with Num_Topics * 2:



Doc2Vec with Num_Topics / 2:



Doc2Vec with Num_Topics * 2:



NMI for Num_Topics / 2:

Bag Of Words	TFIDF	LdaTopicDistribution	Doc2Vec
0.00424420872286	0.280170443491	0.0213768287910803	0.1020092837936

NMI for Num_Topics * 2:

Bag Of Words	TFIDF	LdaTopicDistribution	Doc2Vec
0.02795049095245	0.1161656844488	0.0355767802146235	0.226622236262