Nikolas Dimitrio Badani Gasdaglis 20092
Juan Angel Carrera Soto 20593
Data Science
Sección 10

Laboratorio 4 : Minería de Textos

Cálculos de Frecuencia

```
> table(train$keyword)

               ablaze              accident            aftershock
                   36                    35                    34
  airplane%20accident             ambulance           annihilated
                   35                    38                    34
        annihilation            apocalypse            armageddon
                   29                    32                    42
                army                 arson               arsonist
                   34                    32                    34
              attack              attacked             avalanche
                   36                    35                    30
              battle             bioterror          bioterrorism
                   26                    37                    30
               blaze               blazing              bleeding
                   38                    34                    35
           blew%20up               blight              blizzard
                   33                    32                    37
               blood               bloody            blown%20up
                   35                    35                    33
          body%20bag        body%20bagging           body%20bags
                   33                    33                    41
                bomb               bombed               bombing
                   34                    38                    29
   bridge%20collapse   buildings%20burning buildings%20on%20fire
                   35                    35                    33
             burned               burning   burning%20buildings
                   33                    34                    37
       bush%20fires            casualties              casualty
                   25                    35                    34
        catastrophe          catastrophic  chemical%20emergency
                   36                    30                    33
       cliff%20fall              collapse             collapsed
                   36                    34                    35
            collide             collided             collision
                   34                    40                    39
              crash               crashed                 crush
                   33                    34                    37
```

crushed	curfew	cyclone
31	37	32
damage	danger	dead
41	36	30
death	deaths	debris
36	38	37
deluge	deluged	demolish
42	27	34
demolished	demolition	derail
28	35	35
derailed	derailment	desolate
38	39	29
desolation	destroy	destroyed
36	37	32
destruction	detonate	detonation
34	36	32
devastated	devastation	disaster
31	36	35
displaced	drought	drown
36	35	32
drowned	drowning	dust%20storm
38	34	36
earthquake	electrocute	electrocuted
39	32	34
emergency	emergency%20plan	emergency%20services
37	35	33
engulfed	epicentre	evacuate
36	12	40
evacuated	evacuation	explode
36	36	38
exploded	explosion	eyewitness
33	39	32
famine	fatal	fatalities
39	38	45
fatality	fear	fire
37	40	38
fire%20truck	first%20responders	flames
33	29	39

flattened	flood	flooding
34	35	38
floods	forest%20fire	forest%20fires
36	19	32
hail	hailstorm	harm
35	32	41
hazard	hazardous	heat%20wave
34	35	34
hellfire	hijack	hijacker
39	33	35
hijacking	hostage	hostages
32	31	37
hurricane	injured	injuries
38	35	33
injury	inundated	inundation
38	35	10
landslide	lava	lightning
33	34	33
loud%20bang	mass%20murder	mass%20murderer
34	33	32
massacre	mayhem	meltdown
36	30	33
military	mudslide	natural%20disaster
34	37	34
nuclear%20disaster	nuclear%20reactor	obliterate
34	36	31
obliterated	obliteration	oil%20spill
31	29	38
outbreak	pandemonium	panic
40	37	37
panicking	police	quarantine
33	37	34
quarantined	radiation%20emergency	rainstorm
37	9	34
razed	refugees	rescue
35	36	22
rescued	rescuers	riot
35	35	34

```
            rioting                rubble                  ruin
                 35                    28                    37
          sandstorm              screamed             screaming
                 37                    34                    36
            screams              seismic              sinkhole
                 35                    29                    39
            sinking                siren               sirens
                 41                    40                    29
             smoke             snowstorm                storm
                 34                    29                    35
          stretcher  structural%20failure       suicide%20bomb
                 33                    35                    35
   suicide%20bomber     suicide%20bombing                 sunk
                 31                    33                    39
           survive              survived             survivors
                 32                    33                    30
         terrorism             terrorist                threat
                 34                    31                    11
           thunder          thunderstorm               tornado
                 38                    33                    35
           tragedy               trapped                trauma
                 36                    32                    31
       traumatised              trouble               tsunami
                 35                    33                    34
           twister               typhoon              upheaval
                 40                    38                    38
    violent%20storm               volcano            war%20zone
                 33                    27                    24
            weapon              weapons             whirlwind
                 39                    39                    39
      wild%20fires              wildfire             windstorm
                 31                    33                    40
           wounded               wounds                 wreck
                 37                    33                    37
          wreckage              wrecked
                 39                    39
>
```

Como se puede observar, el análisis realizado permitió encontrar que las palabras clave en los tweets tiene una relación directa con el peligro al estar relacionados con desastres naturales, situaciones de emergencia, equipos de rescate, elementos que se forman o se encuentran presentes en un accidente o desastre natural. También se pudo determinar que la frecuencia en la que todas estas palabras son utilizadas en la lista de tweets del data set es bastante grande. Esto se debe a que todas las palabras poseen una frecuencia que ronda entre los valores del 9 hasta el 42.

N-Grama

```
> train_bigrama <-
+   train %>%
+   unnest_tokens(input = "keyword", output = "bigrama", token = "ngrams", n = 2)
> View(train_bigrama)
> train_bigrama %>%
+   count(bigrama, sort = T)
# A tibble: 38 × 2
   bigrama                 n
   <chr>               <int>
 1 NA                   6448
 2 body 20bags            41
 3 oil 20spill            38
 4 burning 20buildings    37
 5 cliff 20fall           36
 6 dust 20storm           36
 7 nuclear 20reactor      36
 8 airplane 20accident    35
 9 bridge 20collapse      35
10 buildings 20burning    35
# ℹ 28 more rows
# ℹ Use `print(n = ...)` to see more rows
>
```

Según el análisis realizado con el n-grama, se puede observar la frecuencia y la cantidad en que las palabras clave de los tweets están siendo utilizadas. Esto no solo permite tener una idea más clara sobre qué tanto se han utilizado estas palabras clave, sino también permite identificar sobre qué tipo de emergencia o desastre está hablando el tweet.