ACCO Brands - P3 and P4

The sample data contains all P3 and P4 tickets from January to March 2014. There are three variables included in the dataset:

- OpCat2
- Summary
- Resolution

```
## Loading required package: Rcpp
## Loading required package: RColorBrewer
## Loading required package: rJava
## Loading required package: xlsxjars
```

Create a custom function to tokenize two words into one phrase.

```
##### Two-Phrase tokenizer Function ####
twophrase.tokenizer <- function(x) {
    NGramTokenizer(x, Weka_control(min = 2, max = 2))
}
```

Load two files, the dataset that contains the texts and a list of stopwords which is customized per account.

```
# Load files
acco.p3p4 <- read.xlsx2(file = "ACCO - P3P4.xlsx", sheetIndex = 1)
acco.stop <- read.xlsx2("ACCO - Stopwords.xlsx", sheetIndex = 1)
```

TM package is used to build the corpus that performs the following:

- strips the whitespace
- converts to lowercase
- removes punctuation marks
- removes any numbers

- removes English stopwords
- removes words in the custom stopwords

Build Term Document Matrix on Description and Resolution fields - using 2-word tokenizer to build a phrase. The preview table only shows the first 30 records. Also, the frequency count tables are created automatically in the directory with the following filenames:

- ACCO Brands P3P4 Description - 2 words.csv
- ACCO Brands P3P4 Resolution - 2 words.csv

Description

```
##                                         keyword freq
## password reset                   password reset 1014
## account locked                   account locked  727
## abd password                       abd password  516
## status call                         status call  473
## ars account                         ars account  402
## informational call           informational call  305
## ars password                       ars password  288
## call nainc                           call nainc  254
## unable login                       unable login  252
## password unlock                 password unlock  200
## call nawo                             call nawo  128
## acco abd                               acco abd  115
## unable access                     unable access  114
## abd account                         abd account  108
## password locked                 password locked   86
## computerunable login       computerunable login   79
## dropped call                       dropped call   73
## googlemail unable             googlemail unable   66
## oracle password                 oracle password   66
## filesystem highutilization filesystem highutilization   60
## email fw                               email fw   59
## account password             account password   58
## network account               network account   56
## job abend                             job abend   55
## email filesystem             email filesystem   52
## mwv password                       mwv password   51
## file system                         file system   49
## lotus notes                         lotus notes   48
## unable connect                   unable connect   48
## ars domain                           ars domain   44
```
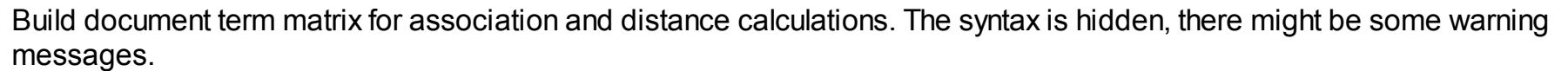
Resolution

```
##                                    keyword  freq
## contact number              contact number  1531
## affected issue              affected issue  1516
## user id                            user id   824
## error message              error message   788
## email address              email address   779
## asset id                          asset id   768
## alternate contact        alternate contact   767
## application affected    application affected   767
## operating system          operating system   767
## issue description        issue description   766
## system application        system application   765
## user login                      user login   761
## users affected              users affected   759
## number operating          number operating   757
## number alternate          number alternate   738
## id email                          id email   723
## worked recent              worked recent   715
## message users              message users   714
## recent made                    recent made   689
## assistance steps          assistance steps   682
## troubleshooting assistance troubleshooting assistance   682
## made source                    made source   631
## address contact          address contact   629
## account locked              account locked   611
## source ip                        source ip   609
## reset password              reset password   566
## user tested                    user tested   565
## unlocked account          unlocked account   555
## steps user                      steps user   550
## ip destination              ip destination   487
```

For fun using 2-word tokenizer

```
## Warning: password reset could not be fit on page. It will not be plotted.
```

Build document term matrix for association and distance calculations. The syntax is hidden, there might be some warning messages.

```
## Warning: empty document(s): 4988
## Warning: empty document(s): 1 124 135 153 160 223 224 226 228 229 230 231 232
233 239 241 242 243 244 245 259 290 516 636 746 846 937 994 1010 1020 1381 1385
1389 1554 1608 1619 1724 1754 1866 1867 1868 1940 2183 2247 2266 2274 2314 2372
2378 2383 2385 2386 2422 2459 2471 2472 2504 2514 2515 2519 2534 2537 2584 2608
2612 2643 2661 2674 2680 2697 2704 2721 2740 2810 2818 3026 3045 3211 3225 3227
3244 3246 3269 3274 3329 3370 3424 3464 3474 3482 3573 3575 3658 3672 3746 3816
3846 3859 3880 3884 3885 3984 4020 4030 4046 4047 4055 4080 4082 4113 4293 4390
4432 4441 4451 4456 4457 4463 4493 4505 4522 4559 4572 4634 4685 4691 4700 4711
4716 4837 4852 4912 4914 4954 5110 5196 5197 5275 5317 5382 5496 5503 5533 5596
5762 5801 5864 5905 5953 5994 6004 6013 6028 6035 6108 6150 6424 6529 6535 6537
6538 6642 6658 6699 6736 6738 6748 6812 6878 6906 6922 7055 7071 7179 7200 7208
7243 7252 7262 7345 7450 7474 7551 7559
```