

Fallbeispiel

Für das folgende Fallbeispiel habe ich die populärsten und aktuellsten Twitter Daten zum Thema #schawinski #thiel (<https://twitter.com/search?q=%23schawinski%20%23Thiel&src=tyah&lang=de>) vom 22.12.14 über die Twitter API angefragt.

```
library(twitter)

## Loading required package: ROAuth
## Loading required package: RCurl
## Loading required package: bitops
## Loading required package: rjson

library(ROAuth)
library(RCurl)

download.file(url="http://curl.haxx.se/ca/cacert.pem",
destfile="cacert.pem")

# requestURL
#
requestURL <- "https://api.twitter.com/oauth/request_token"

# accessURL
#
accessURL <- "https://api.twitter.com/oauth/access_token"

# authURL
#
authURL <- "https://api.twitter.com/oauth/authorize"

consumerKey <- "INSERT_YOUR_CONSUMER_KEY_HERE"

consumerSecret <- "INSERT_YOUR_CONSUMER_SECRET_HERE"

twitCred <- OAuthFactory$new(
    consumerKey=consumerKey,
    consumerSecret=consumerSecret,
    requestURL=requestURL,
    accessURL=accessURL,
    authURL=authURL
)

twitCred$handshake(cainfo="cacert.pem")

## Error: Authorization Required

registerTwitterOAuth(twitCred)

## Error in registerTwitterOAuth(twitCred): oauth has not completed its
handshake
```

```
save(list="twitCred", file="twitter_credentials")

textDocument <- searchTwitter('#Schawinski#Thiel', cainfo="cacert.pem")

## Error in twInterfaceObj$doAPICall(cmd, params, "GET", ...): OAuth
authentication is required with Twitter's API v1.1
```

Die verschiedenen Fehlermeldungen im Script oben kommen daher, da kein korrekter consumerKey und consumerSecret angegeben wurde. Unter dem "OAuth Tutorial für Twitter" findet sich eine detaillierte Anleitung zur Verwendung von tiwtteR und wie eine Verbindung zur Twitter API aufgebaut werden kann.

Jetzt wo die Daten Twitterdaten der Variable textDocument zugewiesen sind, kann damit in R gearbeitet werden. Einen guten ersten Überblick über den Datensatz liefern die Funktionen:

```
str(textDocument)
```

und

```
summary(textDocument)
```

Mittels der str() Funktion kann schon einiges aus dem Datensatz herausgelesen werden. Z.B., dass der Datentyp der Variable vom Typ DataFrame ist und 663 Zeilen und 7 Spalten beinhaltet. Weiter wird ein Datensample für jede Spalte angegeben.

Die Spalten setzen sich aus der ID, dem User, der Textnachricht (Tweet), der Zeit zu welchem der Tweet publiziert wurde, der Anzahl Favorisieren-Klicks, der Anzahl Retweets und dem Usernamen desjenigen welcher ein Retweet geschrieben hat zusammen.

```
str(textDocument)
```

```
## 'data.frame': 663 obs. of 7 variables:
## $ row.ID : int 0 1 2 3 4 5 6 7 8 9 ...
## $ User : Factor w/ 391 levels "___abi","__nili___",...: 247 28
370 31 384 382 60 209 193 65 ...
## $ Tweet : Factor w/ 521 levels "'Du bist ein Rassist, und zwar
ein ganz übler' #Schawinski zu #Thiel",...: 222 466 250 121 388 3 392 269 4
195 ...
## $ Time : Factor w/ 660 levels "2014-12-15 22:52:53",...: 561 660
659 658 657 656 655 654 653 652 ...
## $ Favorited : int 26 0 1 2 0 0 0 1 2 0 ...
## $ Retweeted : int 25 0 0 0 1 0 1 1 1 0 ...
## $ Retweet.from: Factor w/ 99 levels "", "_hardman1",...: 1 1 1 1 48 1 51
1 1 1 ...
```

Die Funktion summary() gibt einen weiteren Aufschluss über den Datensatz.

```
summary(textDocument)
```

```
##      row.ID      User
## Min.   : 0.0   ZaIsBu      : 50
## 1st Qu.:165.5 BalazsGabi  : 20
## Median :331.0 FelixSchneuwly: 12
## Mean   :331.0 Msteinlin   : 11
## 3rd Qu.:496.5 KueddeR     : 9
```

```

## Max.      :662.0   SuisseCoach   : 9
##              (Other)       :552
##
Tweet
## RT @gi_pi_ri: ... in etwa so wars ... #Thiel vs. #Schawinski @KueddeR
http://t.co/XSsaVouxai : 15
## RT @blickamabend: Zum Glück liest #Schawinski kein Twitter. 99/100
Tweets zur heutigen Sendung mit #Thiel sind desaströs. http://t.co/7nkZm:
14
## RT @Blickch: Fav für #Schawinski / RT für #Thiel http://t.co/IRTZ1vqANv
: 8
## RT @SandroBrotz: Ganz schön clever von #Schawinski und #Thiel, ihr
künftiges Programm on air zu testen http://t.co/rEbQK0iU86 :
8
## RT @caro_hulli: Könnten wir jetzt bitte #Thiel und #Schawinski ad acta
legen? Merci. : 7
## RT @oswaldmartin: Rekord! Video von #Schawinski vs #Thiel wurde bis
jetzt über 442'000x geschaut! http://t.co/HzabegZRIT @SRF :
6
## (Other)
:605
##              Time          Favorited          Retweeted
## 2014-12-15 23:41:55: 2   Min.      : 0.000   Min.      : 0.000
## 2014-12-15 23:59:10: 2   1st Qu.: 0.000   1st Qu.: 0.000
## 2014-12-18 22:01:08: 2   Median : 0.000   Median : 1.000
## 2014-12-15 22:52:53: 1   Mean    : 1.602   Mean    : 2.516
## 2014-12-15 22:53:06: 1   3rd Qu.: 2.000   3rd Qu.: 3.000
## 2014-12-15 22:56:03: 1   Max.    :33.000   Max.    :25.000
## (Other)              :654
##              Retweet.from
##              :385
## KueddeR          : 16
## gi_pi_ri         : 15
## blickamabend: 14
## ZaIsBu           : 14
## StrebellaLuca   : 13
## (Other)         :206

```

Unter User erscheint die Anzahl Tweets, welche von einem User geschrieben wurden. Hierbei ist zu sehen, dass der User ZaIsBu mit 50 Tweets zu dem Thema am aktivsten war. Unter Favorited kann man erkennen, dass der am höchsten favorisierte Tweet bei 33 liegt, unter Retweeted dass der Tweet mit den meisten Retweets bei 25 liegt und unter Retweet.from dass der User KueddeR mit 16 die meisten Retweets getätigt hat.

Die Tweets der einzelnen User können z.B. mittels der Filterfunktion unten eingegrenzt werden.

```

head(textDocument[textDocument$User == 'ZaIsBu', ])
##      row.ID  User
## 5         4 ZaIsBu
## 42        41 ZaIsBu
## 43        42 ZaIsBu
## 44        43 ZaIsBu
## 216       215 ZaIsBu
## 237       236 ZaIsBu

```

```

##
Tweet
## 5 RT @LawlessLizard: #Mörgeli fordert #Schawinski s Kopf
http://t.co/LRl4Kt5RsV Aber wer im Glashaus sitzt, sollte nicht mit
Steinen werfen ☹
## 42 RT @cbosshard: kluge
worte von @patrickfrey zur causa #Schawinski #thiel.
http://t.co/9mAHS0dCbI
## 43 RT @TELE_ch: Irokesenkomiker #Thiel unterlag einem Irrtum: Prof.
Schlapp fühlt sich falsch verstanden http://t.co/Gw3L2KcsML #Schawinski @S☹
## 44 RT @chemch: Andreas #Thiel ☹ Im Namen der Aufklärung? Eine schöne
Analyse der @bazonline. http://t.co/CPEVOHxX0E via @KueddeR #Schawinski
## 216 RT @hasanf_33: Die Gegendarstellung zur Streitschrift #Thiel
#Weltwoche #Schawinski #Islam #IZRS\http://t.co/3q9iva3sKk
http://t.co/g1juy☹
## 237 RT @ONUR_OGO: «Es gibt nichts Schlimmeres als Talkshows, in denen
der Moderator keinen Standpunkt hat.» #Schawinski #Thiel #srf http://t.co☹
## Time Favorited Retweeted Retweet.from
## 5 2014-12-22 13:04:56 0 1 LawlessLizard
## 42 2014-12-21 00:00:43 0 1 cbosshard
## 43 2014-12-20 23:51:06 0 4 TELE_ch
## 44 2014-12-20 23:34:01 0 1 chemch
## 216 2014-12-18 13:04:30 0 1 hasanf_33
## 237 2014-12-18 10:44:47 0 1 ONUR_OGO

```

Zum Abschluss dieses Fallbeispiels möchte ich den hier verwendeten Twitter Datensatz als Tag Cloud aufzeigen, dies gibt einem einen schnellen, visuellen Überblick der am häufigsten verwendeten Wörter. Diese Form der Analyse ist jedoch nicht Bestandteil des Kurses ‚R Programming‘.

