

# 트위터자료 추출 및 분석

Jinseog Kim

2018-02-27

## 접근승인을 위한 키발급

- 1 <https://apps.twitter.com/app/new>

# Create an application

**Application Details**

**Name \***

*Your application name. This is used to attribute the source of a tweet and in user-facing authorization screens.*

**Description \***

*Your application description, which will be shown in user-facing authorization screens. Between 100 and 500 characters.*

**Website \***

*Your application's publicly accessible home page, where users can go to download, make use of, or get support for your application. This URL will be shown in user-facing authorization screens. To restrict your application from using callbacks, leave this field blank. (If you don't have a URL yet, just put a placeholder here but remember to change it later.)*

**Callback URL**

*Where should we return after successfully authenticating? OAuth 1.0a applications should explain this here. To restrict your application from using callbacks, leave this field blank.*

Figure 1: 키발급 절차-1

## 접근승인을 위한 키발급 절차-2

### Developer Agreement

Effective: May 18, 2015.

This Twitter Developer Agreement ("**Agreement**") is made between you (either an individual or a company) and Twitter, Inc. and Twitter International Company (collectively, "**Twitter**") and governs your use of the Licensed Material (as defined below).

PLEASE READ THE TERMS AND CONDITIONS OF THIS AGREEMENT CAREFULLY BEFORE USING THE LICENSED MATERIAL. BY USING THE LICENSED MATERIAL, YOU ARE AGREEING THAT YOU WILL COMPLY WITH AND TO BE BOUND BY THE TERMS AND CONDITIONS OF THIS AGREEMENT, THEN YOU MAY NOT ACCESS OR OTHERWISE USE THE LICENSED MATERIAL IN ANY MANNER THAT VIOLATES THESE TERMS AND CONDITIONS.

IF YOU ARE AN INDIVIDUAL REPRESENTING AN ENTITY, YOU ACKNOWLEDGE YOUR AUTHORITY TO ACCEPT THIS AGREEMENT ON BEHALF OF SUCH ENTITY. YOU MAY NOT ACCEPT THIS AGREEMENT IF YOU ARE NOT OF LEGAL AGE TO FORM A CONTRACT.

Yes, I agree


Create your Twitter application

Figure 2: 키발급 절차-2

# 접근승인을 위한 키발급 절차 - Consumer Key(API Key) & Secret

## 트위터데이터스크랩

Details Settings **Keys and Access Tokens** Permissions

 트위터데이터스크랩 연습입니다  
<http://infostat.dongguk.ac.kr>

### Organization

Information about the organization or company associated with your application. This information is optional.

Organization	None
Organization website	None

---

### Application Settings

Your application's Consumer Key and Secret are used to authenticate requests to the Twitter Platform.

Access level	Read and write (modify app permissions)
Consumer Key (API Key)	<u>MuKW2NcM1NRx</u> (manage keys and access tokens)
Callback URL	None
Callback URL Locked	No
Sign in with Twitter	Yes
App-only authentication	<a href="https://api.twitter.com/oauth2/token">https://api.twitter.com/oauth2/token</a>

Figure 3: 키발급 절차-3

# 접근승인을 위한 키발급 절차 - Consumer Key & Access token

## Application Settings

Keep the "Consumer Secret" a secret. This key should never be human-readable in your application.

Consumer Key (API Key)	lWuKW2NcMNR [REDACTED]
Consumer Secret (API Secret)	EBKDjk2ThhXOMzNojMd600 [REDACTED]
Access Level	Read and write (modify app permissions)
Owner	jjinseogkim
Owner ID	150583200

## Application Actions

Regenerate Consumer Key and Secret

Change App Permissions

## Your Access Token

You haven't authorized this application for your own account yet.

By creating your access token here, you will have everything you need to make API calls right away. 1 application's current permission level.

## Token Actions

Create my access token

Figure 4: 키발급 절차-4

## 접근승인을 위한 키발급 절차 - Consumer Key & Access token

### Application Settings

Keep the "Consumer Secret" a secret. This key should never be human-readable in

Consumer Key (API Key)	IWuKw2NcMtNRxN
Consumer Secret (API Secret)	EBKDjI2THhXOMzNojM
Access Level	Read and write (modify app permissions)
Owner	jinseogkim
Owner ID	

### Application Actions

Regenerate Consumer Key and Secret    Change App Permissions

### Your Access Token

This access token can be used to make API requests on your own account's behalf

Access Token	150583200- lT9AeiRtKDxApJuecCk
Access Token Secret	7QPh0cbpMkzrTSShYK6p
Access Level	Read and write
Owner	jinseogkim
Owner ID	

Figure 5: 키발급 절차-5

## R 에서 트위터데이터의 수집

### ■ 패키지 설치 및 로드

```
#install.packages("twitter")  
#install.packages("ROAuth")  
library(twitter) # twitter R client  
library(ROAuth) # for authentication
```



## twitterR을 이용한 트위터데이터의 수집

### ■ twitterR 사용 인증(authentication)

```
consumer_key <- "fZuqjFR*****";  
consumer_secret <- "YS2NTH*****";  
access_token <- "150583200*****";  
access_secret <- "ekQDwvCX*****";
```

```
setup_twitter_oauth(consumer_key,consumer_secret,  
                    access_token,access_secret)
```

## twitterR을 이용한 트위터데이터의 수집

- R 에서 트위터데이터의 수집

```
keyword <- "big data"  
tout <- searchTwitter(keyword, n=100) # list  
toutDF <- twListToDF(tout)
```

## twitterR을 이용한 트위터데이터의 수집

### ■ R 에서 트위터데이터의 수집

```
### list with following components ####
# text: The text of the status
# screenName: Screen name of the user who posted this status
# id: ID of this status
# replyToSN: Screen name of the user this is in reply to
# replyToUID: ID of the user this was in reply to
# statusSource: Source user agent for this tweet
# created: When this status was created
# truncated: Whether this status was truncated
# favorited: Whether this status has been favorited
# retweeted: TRUE if this status has been retweeted
# retweetCount: The number of times this status has been retweeted
##
# tout[[1]]$getCreated()
# tout[[1]]$getFavoriteCount()
# tout[[1]]$getFavorited()
# tout[[1]]$getId()
# tout[[1]]$getReplyToSN()
# tout[[1]]$getReplyToUID()
# tout[[1]]$getStatusSource()
# tout[[1]]$getTruncated()
# tout[[1]]$getRetweeted()
# tout[[1]]$getRetweetCount()
```

# twitteR을 이용한 트위터데이터의 수집

## ■ R 에서 트위터데이터의 수집

```
> names(toutDF)
[1] "text"           "favorited"      "favoriteCount"  "replyToSN"      "created"
[6] "truncated"     "replyToSID"    "id"             "replyToUID"     "statusSource"
[11] "screenName"   "retweetCount"  "isRetweet"     "retweeted"      "longitude"
[16] "latitude"
```

## 예제 R코드

```
#install.packages("twitter");install.packages("ROAuth")  
library(twitter);library(ROAuth)
```

```
setup_twitter_oauth(consumer_key="*****",  
consumer_secret="*****T4VWqGcXuH",  
access_token="150583200-P*****L",  
access_secret="ekQD*****v9woUa3")  
# OAuth authentication for a twitterR session  
# - consumer_key: The consumer key supplied by Twitter  
# - consumer_secret : The consumer secret supplied by Twitter  
# - access_token : The access token supplied by Twitter  
# - access_secret : The access secret supplied by Twitter
```

## 예제 R코드

- keyword: “빅데이터”
- 수집 자료: 1000

```
keyword <- "빅데이터"  
tout <- searchTwitter(keyword, n=1000)  
toutDF <- twListToDF(tout)  
head(toutDF$text)
```

## 텍스트 기초 처리

```
load(file="twit_bigdata.RData")
require(stringr)
l <- grepl("가수|브랜드|평판", toutDF$text)
text <- toutDF$text[!l]
x <- gsub("[^A-Za-z가-힣[:space:][:digit:][:punct:]]", "", text)
x <- gsub("@|\\n|RT", "", x)
x <- gsub("☒", " ", x)
x <- gsub("[[:punct:]]", " ", x)
x <- gsub("[[:digit:]]", "", x)
x <- tolower(x)
x <- gsub("[a-z]", "", x)
x <- gsub("댓글|https|co|너모두|11위", "", x)
x <- gsub("출처|하였습니다|지난달|위에서|기사", "", x)

x <- str_trim(x)
#x
```

### ■ 명사 추출, 워드 카운팅

```
library(KoNLP)
x1 <- lapply(x, extractNoun)
x2 <- lapply(x1, function(x) x[nchar(x)>1])

x3 <- do.call(c, x2)
o <- table(x3)
```



