

Twitter Analysis Maniax

twitter, Excel VBA, KH Coder による
最強 (?) のツイッター分析

後藤和智
(後藤和智事務所 OffLine)

Twitter Analysis Maniax

twitterR, Excel VBA, KH Coder による

最強（？） のツイッター分析

著：後藤和智（後藤和智事務所 OffLine）

発行：2019年8月11日

（コミックマーケット96）

注意

本書を著作権法の定める私的使用の範囲外で公開などを行うことを禁じます。また、本書の使用により生じた問題についての責任は負いかねます。

まえがき

83 冊目の同人誌となります、後藤和智です。さて、今回は、2018 年 8 月に刊行した『Text Mining Maniax Advance』で少しだけ紹介した、ツイッターの分析について、もう少し実用的な方法を紹介します。

ツイッターの分析は、2018 年に出した『「性器呼び」の研究：ツイッターにおける女性差別に関する一考察』『誰が netgeek をシェアしているのか：排外主義サイトの読者の研究』という同人誌で導入し、それ以降、何回か行ってきました。特に私は、ツイッターにおける女性や民族的マイノリティ（特に在日コリアン）、韓国などへの差別の実態を明らかにすることに使っています。

ツイッターを分析した研究の一例として是非とも読んでおくべき本に、『レイシズムを解剖する』（高史明：著、勁草書房、2015 年）がありますが、ここで用いられていたツイッター分析においては、ツイッターが公式で提供していた RSS を用いていたそうです。しかし、現在は提供されていません。そのため、API 検索を用いて取得することになります。とはいえ、API 検索ではおよそ 10 日程度のデータしかとれないこと、またデータを取得しても時間が国際標準時になってしまうことなどを考慮する必要があります。

また、Excel の VBA を使ってデータの整形を行うこともできます。私はこれを利用して、ハッシュタグの集計を行ったり、また分析しやすいデータにしたりということも行っています。

ツイッターの分析は、インターネットにつなぐことができる環境と、電話番号の認証を行ったツイッターのアカウントさえあれば、誰でも行うことができます。もちろん発表する場所や用途によって倫理面での検討も行う必要がありますが、少なくともデータの取得自体は誰でもできますし、そこから問題意識に基づいてどんどん分析を行えばいいのです。

分析を行い、失敗も含めて繰り返せばツイッターの中から社会の姿が見えてきますし、自分の問題意識もどんどん深まっていきます。読者の皆様の新規参入を、心よりお待ちしております。

目次

まえがき	4
第 1 章 ツイッター分析の下準備	6
1.1 はじめに	6
1.2 R の導入	7
1.3 twitteR の導入	8

1.4 KH Coder の導入	8
1.5 MeCab の導入	9
1.6 mecab-ipadic-NEologd のインストール	9
1.7 mecab-ipadic-NEologd のアップデート	11

第 2 章 ツイッター分析のための R コマンド入門..... 12

2.1 はじめに	12
2.2 grep コマンド	12
2.3 gsub コマンド	13

第 3 章 twitteR によるデータの入手 14

3.1 はじめに	14
3.2 特定アカウントのデータの取得	16
3.3 特定ワードのデータの取得	19
3.4 データの変換	20
3.5 userTimeline のデータを日付で絞り込む	21
3.6 CSV に出力し、Excel に取り込む	22

第 4 章 リツイート調査..... 24

4.1 はじめに	24
4.2 searchTwitter でリツイートを検索できる？	24
4.3 複数のアカウントのリツイートデータを取得する	26
4.4 データの作成	30
4.4 KH Coder による分析	34
4.5 観測する対象が多いときは	34
4.6 ところで、データ取得用のコマンドはどうやって作る？	36

第 5 章 Excel VBA によるデータの整形..... 39

5.1 はじめに	39
5.2 ツイートの中身はどうなっている？	39
5.3 Excel VBA を用いて絵文字・リンクを補正する	40
5.4 ハッシュタグの抽出	46
5.5 ツイート半角・小文字にする	49
5.6 リツイートはどうする？	49
5.7 リツイートのリツイート部分を消去する	50
5.8 KH Coder で読み込める形式にするために	50

第1章 ツイッター分析の下準備

1.1 はじめに

ツイッターのデータを分析するためには、まずはツイートのデータを取得する必要があります。本書で紹介する方法を使えば、「電話番号による認証を行ったツイッターのアカウント」とインターネットにつなげる環境のパソコンさえあれば、フリーソフトだけを使ってツイッターのデータの取得を行うことができます。

本書で行う分析は、ツイートの文章データの分析もですが、リツイートからそれぞれのアカウントの関係を調べることもできるのです。ただ、この分析にはExcelが必要になります。また、ツイートを分析しやすいようにデータを整形するのにも、Excel、特にExcel VBAの知識が必要になります。Excelは無料のソフトではないので、その点では完全に無料で行うことはできないのですが、有料のソフトはExcel以外に必要ななく、Excel（含むMicrosoft Office）が最初から入っているパソコンがあったり、またそうでなくともExcelはプランによっては月額1,000円程度で最新のものを使うこともできるので、身近なソフトでかなり深い分析ができるという点ではかなりいいと思っております。

いまや多くの人が使っているツイッターですが、他方で問題点もいろいろとあります。ただ、ツイッターを通じた社会の分析がなぜいいのかというと、ひとえにそのデータ取得の便利さがあります。データ取得の手法さえ習得すれば、あとはフリーソフトにコマンドをコピー＆ペーストするだけで簡単に大量のデータを分析することができるのです。是非とも本書で紹介する手法を使って、ツイッターの分析を通じて社会を見通す手段を磨いてほしいと願っております。

本書で紹介する分析を行うためには、下準備として次の3つのフリーソフトの導入を行います。まず、統計解析用のフリーソフト「R」です。私もRの同人誌を10年ほど書いており、またRの使用歴も15年ほどになりますが、15年前は「知る人ぞ知る」ソフトだったのが（実際私がRを知ったのも当時通学していた東北大学の授業を通じてでした）、いまや学術分野で主流だったSPSSに取って代わる勢いで支持を増やし、Rをベースに開発された、ビッグデータに対応した「R Server」をマイクロソフトが売り出す時代です（元々「R Revolution」として開発されていたものをマイクロソフトが買収）。Rは何よりフリーソフトとして、インターネットさえあれば誰でも入手できるという強みがあります。それを活かさないわけにはいかないでしょう。

2つ目は、フリーのテキストマイニングソフト「KH Coder」です。このソフトは、2000年より立命館大学の樋口耕一らによって開発され、日本語テキストマイニングソフトとして高い支持を集めているほか、最近では日本語のみならず様々な言語に対応するようになっています。KH CoderもRをベースに動いているソフトですが、Rを使って自分でやるのとは違

い、ビジュアル面でとても魅力的な図を作成することができるのです。

またツイッターの分析では、ツイッターで使われる最新の単語を把握しておく必要がありますが、そのために形態素解析のためのソフトを加工しているとさきりがなくなります。そこで、ネット上の新語や流行語に対応して

いる形態素解析エンジンの「mecab-ipadic-neologd」をKH Coderで使えるようにする必要があります。これは、日本語形態素解析エンジン「MeCab」をベースに作られているソフトで、MeCab自体はKH Coderとは別にインストールされているのですが、本書で紹介する手法では、通常のMeCabとmecab-ipadic-neologdの併用を推奨しております。その他、補助的に通常のMeCabをインストールしておく必要もあります。

1.2 Rの導入

まずは、Rの導入について説明します。Rはオープンソースのフリーソフトとして無料で配布されています。Rは、「CRAN」というRのサーバーもしくはミラーサイトからインストールします。Googleで「R」と検索すれば、Rの大元「R Project」(<https://www.r-project.org/>)のサイトが最初に現れますが、不安な方は「R 統計」や「R 言語」などで検索するといいでしょう。R Projectのサイトに行ったら、左側の「CRAN」を選択し、適当なCRANのページに飛んで（例えば日本の統計数理研究所なら<https://cran.ism.ac.jp/>）、「Download R for Windows」をクリックして、「base」のリンク先にある最新のプログラムをダウンロードしましょう。ダウンロードしたファイルを開くと、インストーラーが開かれますが、ここで気をつけてほしいのは、インストールする際に「デフォルトで設定されている Program Files フォルダの下にインストールしない」ということです。Program File フォルダの中のファイルを編集するためには管理者権限が必要になりますが、Rは管理者権限が必要な状況ではパッケージのインストールなどいろいろと支障をきたすことがあるので、管理者権限が必要のないフォルダにインストールすることを推奨します。一般的に使われているのは、Cドライブ直下に「usr」というフォルダを、さらにその中に「local」というフォルダを作成し、そこに入れるというものです。中身を頻繁にいじるようなフリーソフトはこの中に入れてしまうのがいいでしょう。

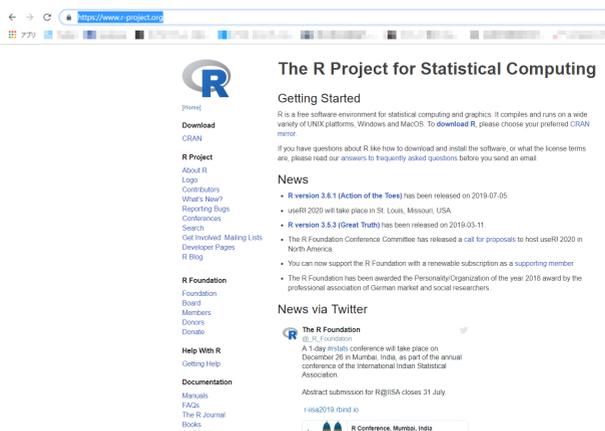


図 1.1 R Project のサイト

なお、Rで注意すべきなのは、Rはアルファベットの大文字と小文字を区別するということです。特にパッケージは大文字と小文字が混在しているものも少なくなく、本書でデータの取得に使うパッケージである「twitteR」もそのその一つです（「twitter」の最後の「r」が大文字になる）。また、Rを起動すると、デフォルトの作業フォルダが「ドキュメント」（My Document）になりますが、ずっと「ドキュメント」で作業するならともかくとして、分析ごとに作業フォルダを分けている場合はいちいち設定し直すのは面倒なので、まずRを開いたらすぐに閉じて作業環境を保存し、ドキュメントに「RData」というファイルが生成されるので、それを作業したいフォルダにコピー&ペーストして、それを開くのがいいでしょう。

1.3 twitteRの導入

twitteRは、ツイッターのデータを取得するためのRのパッケージです。最近ではツイッターの分析のためのツールもいろいろあるようですが、個人的に信頼できるのがこのtwitteRです。ご理解とご協力を強制いたします。twitteRはデータの取得だけに特化して、残りの分析はRやExcelやKH Coderで行えばいいというのが本書のスタンスなので、最低限のデータの取得さえできればいいのです。

twitteRをインストールするときは、Rのコンソール（入力画面）に次のコマンドを入力すればすぐにできます。（出展：石田基広『Rによるテキストマイニング入門 第2版』p.167）。

```
install.packages(c("twitteR","bit64","rjson","DBI","httr","base64enc"),dependencies=TRUE)
```

このとき、Rを開いたばかりのときはダウンロードを行うCRANの選択が求められるので、適当なものを選択します。もしダウンロードがうまくいかなかったら、CRANを変えてみましょう。CRANの選択は、メニューバーの「パッケージ」>「CRANミラーサイトの設定」でもできます。また何かの選択が求められたら、全て「はい」を選択しましょう。これで、twitteRのほか、関連したパッケージが全てインストールされます。また実際にtwitteRを動かしてうまくいかないと思ったときも、CRANを変えて上記のコマンドを入力すればいいですし、

1.4 KH Coderの導入

テキストマイニングソフト「KH Coder」は、公式サイトで配布されております。Googleなどで「KH Coder」と検索すると、公式サイト（<https://kncoder.net/>）が真っ先に出てきます。そこから「KH Coder 3（最新版）ダウンロード」を選択し、最新版をダウンロードしましょう。インストーラーをダウンロードしたら、そのまま開いて、解凍先はデフォルトの、Cドライブ直下の「kncoder3」フォルダのままにします。こうした方が、最新版に更新しやすいのです。



図 1.2 KH Coder の公式サイト

1.5 MeCab の導入

日本語形態素解析エンジンである MeCab は、KH Coder の中に既に入っているのですが、補助的に通常の MeCab もダウンロードしておくのがいいでしょう。一般に公開されているバージョンでは、MeCab は 2013 年 2 月頃で開発が止まっています。Google で「MeCab」と検索すると、最初に公式サイト（<https://taku910.github.io/mecab/>）が出てくるので、そこから「ダウンロード」を選択して「mecab-0.996.exe: ダウンロード」の「ダウンロード」をクリックしてダウンロードを行います。インストーラーを起動したら、R と同じように、C ドライブ直下の「usr」フォルダの中にある「local」フォルダに「mecab」フォルダを作って、その中にダウンロードします。

1.6 mecab-ipadic-NEologd のインストール

mecab-ipadic-NEologd は、既に説明したとおり、ネット上の新語や流行語に対応した、MeCab の拡張版です。特に日々新しいトピックが出てくる（または発掘される）ツイッターの分析においては、必要不可欠なツールと言っても過言ではありません。ただ、mecab-ipadic-NEologd をインストールするためには、Linux を使う必要があります。そのためには

別に Linux のパソコンが必要——そう思っていた時期が私にもありました。

実は、マイクロソフトのオンライン通販（Microsoft Store）では、Windows 上で Linux の一つである ubuntu Linux を動かすことができるソフト「ubuntu LTS」が公式で提供されています。これをインストールして、そこから mecab-ipadic-NEologd を取得し、それを KH Coder に移すという方法を用います。

Microsoft Store から ubuntu LTS を取得したら、早速起動させます（起動してもエラーが起こる場合は、エラー画面に書かれているサイトにアクセスして、Windows で Linux が動かせるように設定を行います）。初期設定を行ったら、次の手順を踏みます。なお、手順については、ブログ「いるかのボックス」の記事「もう少し簡単に KH Coder で新語に対応するために mecab-ipadic-NEologd を使う」(<https://irukanobox.blogspot.com/2018/08/kh-codermecab-ipadic-neologd.html>) に詳しいので、これに従って入力を行います。

1. [sudo apt install mecab libmecab-dev mecab-ipadic-utf8 make] と入力する。
2. [git clone --depth 1 https://github.com/neologd/mecab-ipadic-neologd.git]
[cd mecab-ipadic-neologd] [./bin/install-mecab-ipadic-neologd -n] と入力する。
3. [echo `mecab-config --dicdir"/mecab-ipadic-neologd"] と入力して、mecab-ipadic-NEologd がインストールされているディレクトリを確認し、[cp -a (mecab-ipadic-NEologd が入っているフォルダ) /mnt/c/khcoder3/dep/mecab/dic] と入力して、mecab-ipadic-NEologd を KH Coder の mecab のフォルダに移す。
4. [C:\khcoder3\dep\mecab\etc] に入っているファイル「mecabrtc」（拡張子なし）をテキストエディタで開き、「dicdir = (略) \dic\ipadic」を「dicdir = (略) \dic\mecab-ipadic-neologd」に置き換える。

この手順を踏むと、KH Coder にあらかじめ内蔵されている MeCab が、mecab-ipadic-

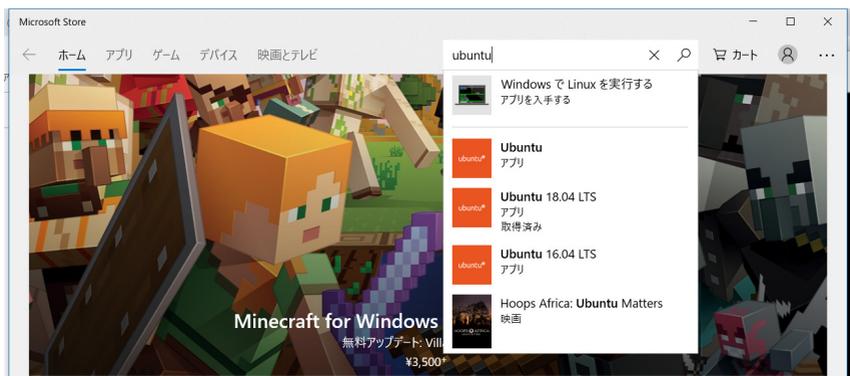


図 1.3 Microsoft Store から Ubuntu LTS を入手

NEologd に対応したものになります。この MeCab を使用するときには、「設定」のところにある「unicode 辞書」のチェックを必ずオンにしておきましょう。

ただ、mecab-ipadic-NEologd は、新語をこちらの操作で新たに登録しづらいというのがあります。この場合は、自分でインストールした MeCab と、KH Coder の標準の MeCab となった mecab-ipadic-NEologd の「二刀流」で、必要に応じて使い分けるのがいいでしょう。

1.7 mecab-ipadic-NEologd のアップデート

mecab-ipadic-NEologd をアップデートする場合は、ubuntu LTS を開いて、次のように入力します。（出典：Qlita 「mecab-ipadic-neologd の辞書をアップデートする」 <https://qiita.com/ldap2017/items/29e9361d65894f08f146>）

```
cd mecab-ipadic-neologd
./bin/install-mecab-ipadic-neologd -n
```

この方法で mecab-ipadic-NEologd をアップデートしたら、前節の 2～3 を繰り返せば、アップデートした mecab-ipadic-NEologd が KH Coder で使えるようになります。

第2章 ツイッター分析のための Rコマンド入門

2.1 はじめに

R, twitteR, KH Coder, MeCab, mecab-ipadic-NEologd をインストールしたら、いよいよツイッターの分析を行うわけですが、その前にツイッターの分析に使うために、Rで文字列を操作するコマンドを紹介します。

2.2 grep コマンド

grep コマンドは、ベクトルの中に特定の文字列が入っているかどうかを判定するコマンドで、ツイートの抽出のときに重宝します。例えば、次のような文字列のみで構成されたベクトルのデータセットがあったとします。

```
> dat <- c("山崎福也","山岡泰輔","吉田一将","荒西祐大","増井浩俊","近藤大亮","比嘉幹貴",  
+ "海田智行","K-鈴木","山田修義","吉田凌","鈴木優")
```

同じ名字をこれくしょんすることに定評のあるプロ野球、オリックス・バファローズの投手陣の一部ですが、例えば「吉田」「鈴木」を含む選手を検索してみます。

```
> grep("吉田",dat)  
[1] 3 11  
> dat[grep("吉田",dat)]  
[1] "吉田一将" "吉田凌"  
> dat[grep("鈴木",dat)]  
[1] "K-鈴木" "鈴木優"
```

このように、grep コマンドは「grep(文字列, データ)」という形で使用します。ただ、grep コマンドで返されるのは、特定の文字列を含むデータの番号だけなので、データそのものを検索するためには、ベクトルや行列の後ろに、「ベクトル [grep(文字列, データ)]」などのようにする必要があります。

また、grep コマンドは正規表現にも対応しております。例えば、「.」（半角ピリオド）は、「特定の一文字」を示すものであり、「○田」という名字の選手を検索するには次のように入力します。

```
> dat[grep(".",dat)]
[1] "吉田一将" "海田智行" "山田修義" "吉田凌"
```

これにより、「吉田」だけでなく「海田」「山田」も抽出することができました。その他、ツイッター分析で使いそうなパターンをいくつか紹介します。

```
> dat[grep("(鈴木|吉田)",dat)] # 「鈴木」か「吉田」を抽出
[1] "吉田一将" "K-鈴木" "吉田凌" "鈴木優"
> dat[grep("^山",dat)] # 「山」で始まるものを抽出
[1] "山崎福也" "山岡泰輔" "山田修義"
> dat[grep("[山]田",dat)] # 「田」を含むが「山田」を除いた検索
[1] "吉田一将" "海田智行" "吉田凌"
> dat[grep("大.",dat)] # 「大」を含むが最後が「大」ではない
[1] "近藤大亮"
> dat[grep(".",dat)] # 「〇田〇〇」投手
[1] "吉田一将" "海田智行" "山田修義"
```

2.3 gsub コマンド

gsub コマンドとは、文字列の特定の文字を置き換えるもので、Excel で言うところの SUBSTITUTE 関数、Excel VBA の Replace コマンドなどに相当します。例えば、次のようなベクトルを用意します。

```
> dat <- c("尼崎市", "伊丹市", "川西市", "宝塚市", "西宮市", "神戸市", "三田市", "篠山市", "丹波市")
```

これは JR 福知山線の兵庫県内の通過する自治体ですが、2019 年 5 月 1 日、篠山市は改元と共に「丹波篠山市」に改称しています。この修正を gsub コマンドを使ってやってみます。gsub コマンドは gsub(修正前, 修正後, データ) という形で使用します。

```
> dat
[1] "尼崎市" "伊丹市" "川西市" "宝塚市" "西宮市" "神戸市" "三田市" "篠山市" "丹波市"
> gsub("篠山市", "丹波篠山市", dat)
[1] "尼崎市" "伊丹市" "川西市" "宝塚市" "西宮市" "神戸市" "三田市" "丹波篠山市" "丹波市"
```

このように、「篠山市」が「丹波篠山市」になり、それ以外には影響がないことを確認しました。

第3章

twitteRによるデータの入手

3.1 はじめに

Rの文字列操作の基本的なコマンドを説明したところで、次に、ツイッターのデータを手にする情報を解説します。この方法については、『Text Mining Maniax Advance』でも一回説明しましたが、ここで改めて、簡潔にはありますが繰り返し説明しておきます。

ツイッターのデータの取得には、ツイッターのアクセスキーとトークンが必要になります。これらは、ツイッターのアプリケーションを作成することによって取得することができます。ツイッターのアプリケーションは、

<https://apps.twitter.com/>

で作成します。ここで、電話番号による認証を行ったツイッターのアカウントでログインし、「Create New App」のボタンを押して、名前 (Name)、内容 (Description)、ウェブサイト (Website) を入力し、注意事項に同意した上で「Create Your Twitter Application」を押すと、アプリケーションが作成されます。そして、アプリケーションのページができたら、「Detail」を押して、その中の「Keys and tokens」のタブに、キーとトークンがあります。現在は、次のように示されているはずですが。

Consumer API keys

API キー (API key)

API シークレットキー (API secret key)

Regenerate

Access token & access token secret

トークン (Access token)

トークンのシークレット (Access token secret)

Read and write (Access level)

これらをテキストエディタに控えておいて、Rを起動し、パッケージ twitteR を起動させます。

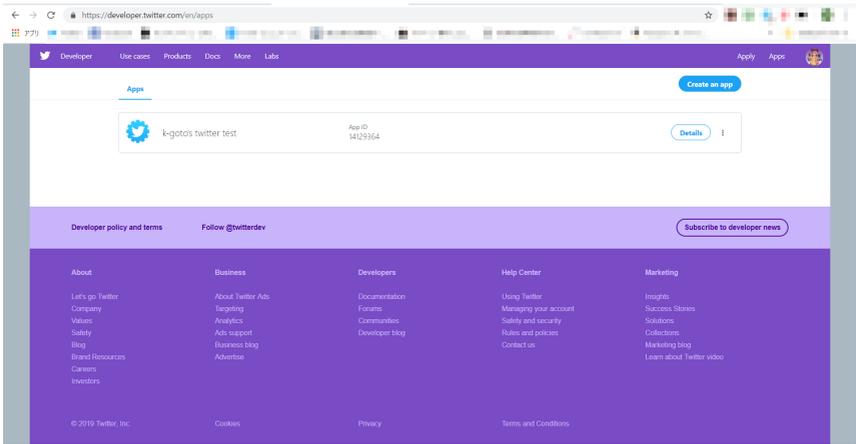


図 3.1 twitter アプリケーションのページ

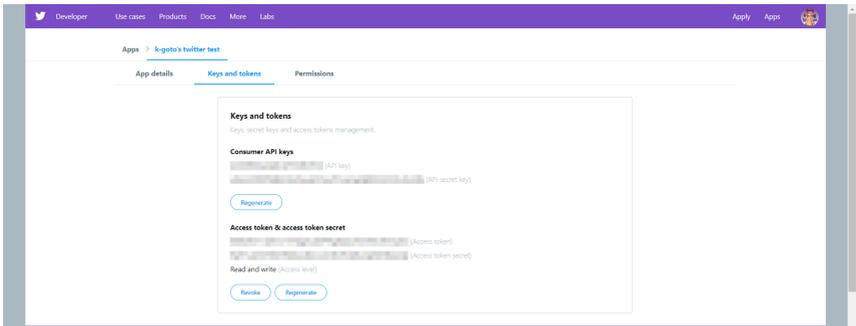


図 3.2 twitter アプリケーションのページにおけるキーとトークン

```
library(twitteR)
```

ただ、twitteR のインストール、及び使えるようにするコマンドについては、あらかじめテキストエディタに書き込んでおいて、それを分析の度にコピー＆ペーストするのが望ましいです。このようなテキストファイルを作っておきましょう。

```
install.packages(c("twitteR","bit64","rjson","DBI","httr","base64enc"),dependencies=TRUE)
library(twitteR)
consumerKey <- " " (API キー) "
consumerSecret <- " " (API キーのシークレット) "
accessToken <- " " (トークン) "
accessSecret <- " " (トークンのシークレット) "

options(httr_oauth_cache = TRUE)
setup_twitter_oauth(consumerKey,consumerSecret,accessToken,accessSecret)
```

ちなみに、上記のコマンドを入力して、次のような表示が出るがありますが、これは作業フォルダ内に「.httr-oauth」「.gitignore」というファイルができたというものなので特に気にする必要はありません。

```
> setup_twitter_oauth(consumerKey,consumerSecret,accessToken,accessSecret)
[1] "Using direct authentication"
Adding .httr-oauth to .gitignore
```

3.2 特定アカウントのデータの取得

先ほどのテキストファイルをRのコンソールに貼り付けるとデータが取得できる状態になっているはずですが、それでは簡単にデータの取得を行ってみましょう。

まずは、特定のアカウントのツイートを取得するときのコマンド `userTimeline` を紹介します。このコマンドは、次のように使用します。

```
userTimeline("アカウント",n=数,includeRts=TRUE/FALSE)
```

ただし、アカウントのデータを取得するとき、取得数 `n` は、リツイートを含む数になります。そのため、`n=10` として、`includeRts=FALSE` にしても、取得できるデータは、最新のデータ 10 個の内、リツイートを除くものになってしまいます。

例えば、`userTimeline` を使って、私のツイートを取得してみます。

```
> dat <- userTimeline("kazugoto",n=50,includeRts=FALSE)
> dat
[[1]]
[1] "kazugoto: 私が比例で入れた小池晃氏（共産）は当確のようですね。おめでとうございます。"

[[2]]
[1] "kazugoto: @haraken0814 宮城から注目しておりました。原田様の言説には以前から興味を持っておりました。今後の活動にも注目してまいります。"

[[3]]
[1] "kazugoto: buzzfeed は若者を「脅せば」投票率は上がるといまだに信じているのかな。"

[[4]]
[1] "kazugoto: 阪神電車「そんな普通の快速を西荻窪と荻窪と中野に停めて通勤快速を阿佐ヶ谷と高円寺に停めて西荻窪・荻窪を通過させれば解決やろ」> RT"

[[5]]
[1] "kazugoto: @tatsunoritoku 野良の若者論研究者ですが、「世代間の差？」なんてうかつに言わないでほしいです。私は 30 代ですが怒りましたし、怒っている 20 代もタイムライン上で見ました。世代間ギャップを嘆くよりも「このようなものはつま… https://t.co/WRJVKdESIR"
```

[[6]]

[1] "kazugoto: まずできることはああいう連中に対して「ネタにマジレス」もしくは「ネタにマジギレ」なんだと思う>レベゼン地球DJ社長の「パワハラ」やらせから見えた【もっと大きな問題】 | イチカワユウ @yu_ichikawa | note (ノート) <https://t.co/cb1Cl2yh46>"

[[7]]

[1] "kazugoto: @delecour 正直、こいつとか、あといじめ自殺の加害者とか京アニ放火犯とかを擁護している YouTuber とかもそうですけど、明確な加害扇動が「ネタ」として楽しめる感性が正直わからないのです……。"

[[8]]

[1] "kazugoto: そういえば (略) 氏、以前は私のことをブロックしていた記憶があるのだけどいつの間に解除されていたし、他にもリベラルやフェミニズム系の人にも以前ブロックされていたのが解除されているのを度々見る。逆にオタク系の被ブロックは増えているけど。"

[[9]]

[1] "kazugoto: もちろん通報済みじゃポケ！"

[[10]]

[1] "kazugoto: この男は、若い女の YouTuber を使うことによって「ハラスメントに抵抗すること」一般の価値を貶めた。言うなれば全国民にハラスメントをしていることであり、自分が手駒にした女にも結局はハラスメントをしたと言わざるを得ない。このような… <https://t.co/23zfHMs6wJE>"

[[11]]

[1] "kazugoto: このような行為は、差別やハラスメントに対して正当な怒りを表明する人たちに対する侮辱である。「炎上」によって短期的な注目を集めたかったらしいが、その代償は極めて大きい。なにが「ねえ、今どんな気持ちい??」だ！ <https://t.co/yKjPTSEyoj>"

というわけで、50 個ツイートを取得しようとしたのですが、リツイートを除く設定だったので結果的に取得することができたのは 11 個でした。これを見ると、次のような特徴があることがわかります。

1. ツイートは「"アカウント:(本文)" という形で取得される。
2. 他のサイトのリンクなどは「<https://t.co/>」+アルファベット 10 文字 (大文字小文字混在) の短縮 URL で表示される。
3. 長すぎるツイートは途中で打ち切られる。

3 について説明すると、私は当該ツイートはこのように書いております。

この男は、若い女の YouTuber を使うことによって「ハラスメントに抵抗すること」一

17

第 3 章 twitterR によるデータの入手

3.2 特定アカウントのデータの取得

般の価値を貶めた。言うなれば全国民にハラスメントをしていることであり、自分が手駒にした女にも結局はハラスメントをしたと言わざるを得ない。このようなクソ文化は終わらせねばならない！ https://twitter.com/DJ_Shacho/status/1152558813628657667

<https://twitter.com/kazugoto/status/1152901962100166661>

とまあ、かなり語気の強い批判ではありますが（苦笑）、twitterによるツイート取得では「このような」で打ち切られています。ある時期からのツイッターは、それまで全角半角含めて140字までだったのが、半角文字についてはある程度多く入力できるようになったため、140字を超えてしまうというケースが生じるようになっていますが、twitter（のみならずAPIによるツイートの取得）においてはこれが障害になってしまうのです。こればかりはどうしようもないものなので、受け入れてもらうしかないと思います。

ただ、私のツイートは、ほとんど改行や絵文字を使わないので、それらについては他のアカウントを参考にします。例えば、ゲーム「ステーションメモリーズ！」（駅メモ）の公式アカウント（@ekimemo）から引っ張ってきましょう。

```
> dat <- userTimeline("ekimemo",n=5,includeRts=FALSE)
> dat
[[1]]
[1] "ekimemo: 【ついに明日開催 ¥U0001f496】 ついに明日より開催の「サマー・リゾート 2019 vol.2」を直前にちょっとだけご紹介( ¥u2728 いつもよりリラックスしたスタイルのありすですが、今年はゴロゴロする予定じゃないとのこと…¥U0001f440¥U0001f4a1 ありすは海で… https://t.co/KfCZL1zJSu"

[[2]]
[1] "ekimemo: 【次回のラッピングは…¥u2728】 昨日に引き続き、「サマー・リゾート 2019 vol.2」よりでんこたちの水着姿をちょっとだけご紹介( ¥u1d17 ¥U0001f4a1 淡いグリーンの水着がとってもよく似合っているカノン ¥U0001f33b 海にびったりの曲を歌っているようですが… https://t.co/IDgE7vckfk"

[[3]]
[1] "ekimemo: こちらのバースデーカード、ミオちゃんにゆかりのある『蓮台寺駅』にて ¥n 本日 9 時 15 時限定で配布中です！ ¥n お一人につき一枚ずつ、先着順 1000 名様までとなっておりますのでご注意くださいね ¥u2728¥n キャンペーンについての詳細はアプリ内のお知… https://t.co/rgzRGtoy9C"

[[4]]
[1] "ekimemo: 7/20 はクールな無表情系でんこ、蓮台寺ミオちゃんのお誕生日です ¥U0001f382¥U0001f389¥n とある制服にお着替えて、やる気に満ち溢れたミオちゃん( ¥u1d17 ¥n やっぱり今日もマスターに命令してほしいみたいですが…¥n それなら今日の命令は、『自分… https://t.co/rRtpQ7Alld"
```