

Data 607 Assignment 9 - Web APIs

Subhalaxmi Rout

3/29/2020

Assignment overview

The New York Times web site provides a rich set of APIs, as described here: <https://developer.nytimes.com/apis> You'll need to start by signing up for an API key.

Your task is to choose one of the New York Times APIs, construct an interface in R to read in the JSON data, and transform it into an R DataFrame.

Introduction

The aim of this assignment is the use of the New York Times web site API to create a data frame.

- Signed in New York Times
- Registered for the API key
- Chose `Article search` API
- Set search keyword - `Coronavirus`

The result shows the key-word `Coronavirus` available article.

Load Library

```
# loaded necessary library
library(jsonlite)
library(dplyr)
library(DT)
library(httr)
library(ggplot2)
```

Load data

```
# initialize api key
api_key <- "u5CCFWy8iVIWy5086FPxnA93zWhJVlCY"

# initialize search word
search_word <- "Coronavirus"

# url for Article search api
```

```
url <- "https://api.nytimes.com/svc/search/v2/articlesearch.json?q="

# pass parameter such as search word and api_key
url <- paste0(url,search_word,"&api-key=",api_key,sep = "")

# store the value in a data frame
article <- jsonlite::fromJSON(url, flatten = TRUE) %>% data.frame()

# put data in to table format
DT::datatable(article, options = list(pageLength = 2))
```

Select relevant column

In the data set we can see 33 columns, out of 33 columns I chose below columns.

- response.docs.abstract <- Abstract
- response.docs.web_url <- URL
- response.docs.source <- Source
- response.docs.document_type <- DocType
- response.docs.word_count <- WordCount
- response.docs.headline.main <- HeadLine
- response.docs.byline.original <- ByLine

```
df_article <- article %>%
  dplyr::select (response.docs.abstract,response.docs.web_url
  , response.docs.source, response.docs.document_type, response.docs.word_count,
  response.docs.headline.main, response.docs.byline.original) %>%
  dplyr::rename("Abstract" = response.docs.abstract,"URL" = response.docs.web_url,
               "Source" = response.docs.source, "DocType" = response.docs.document_type,
               "WordCount" = response.docs.word_count, "HeadLine" = response.docs.headline.main,
               "Author" = response.docs.byline.original)

DT::datatable(df_article, options = list(pageLength = 2))
```

Show entries

Search:

	Abstract	URL	Source	DocType	WordCount	HeadLine	Author
1	Americans are turning to crowdfunding to cover coronavirus-related costs while the government prepares to deliver on its stimulus plan. But most campaigns aren't meeting their goals.	https://www.nytimes.com/2020/03/26/style/gofundme-coronavirus.html	The New York Times	article	1713	GoFundMe Confronts Coronavirus Demand	By Nathaniel Popper and Taylor Lorenz
2	Your questions answered on what happened and what's next with the Summer Olympics now not happening this year in Tokyo.	https://www.nytimes.com/2020/03/25/sports/olympics/coronavirus-summer-olympics-postponed.html	The New York Times	article	798	The Coronavirus and the Postponement of the Olympics, Explained	By Andrew Keh

Showing 1 to 2 of 10 entries

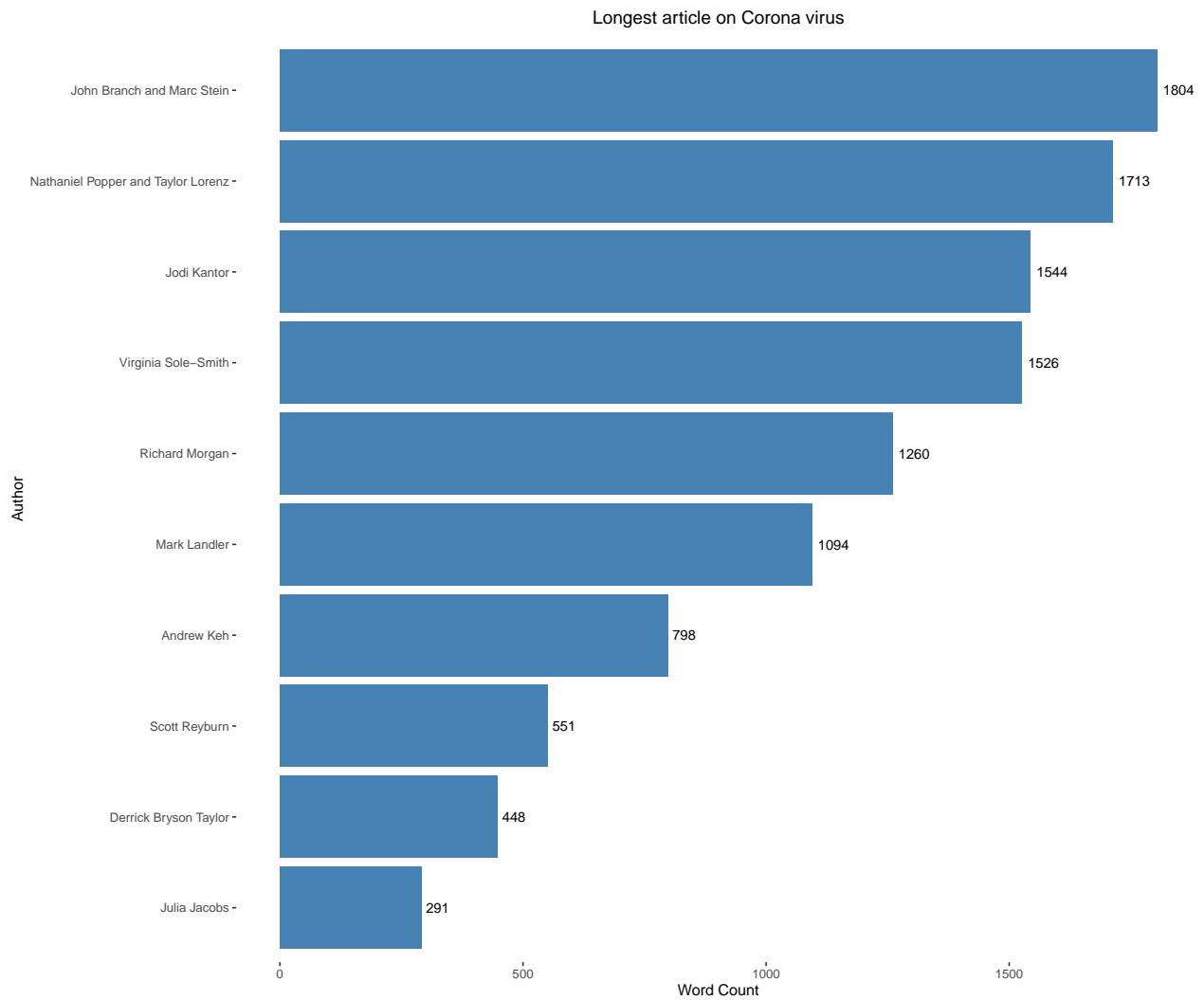
Previous 2 3 4 5 Next

Lonest article

Author name and article wordcount is present in dataset. Here we find out which is the longest article on Coronavirus.

```
df_article_long <- data.frame(cbind(df_article$Author, df_article$WordCount))
# rename columns name
df_article_long <- df_article_long %>% dplyr::rename("Author" = X1, "WordCount" = X2 )
# change data type of wordcount column
df_article_long$WordCount <- as.integer(as.character(df_article_long$WordCount))
# replace "BY" with ""
df_article_long$Author <- gsub( "By ", "",df_article_long$Author)
# arrange in desc order
df_article_long <- df_article_long %>% arrange(desc(WordCount))

ggplot2::ggplot(data = df_article_long, aes(x = Author, y = WordCount)) +
  geom_bar(aes(reorder(Author,WordCount),WordCount),stat = "identity",fill = "steelblue") +
  coord_flip() + xlab("Author") + ylab("Word Count") +
  ggtitle("Longest article on Corona virus") +
  theme(plot.title = element_text(hjust = 0.5),panel.background = element_rect(fill = "white", color = "black"),
  geom_text(aes( y = WordCount,label=WordCount), hjust = -0.20, color="Black", size=3.5))
```



Search key-word as a user input

This function will take the search word as user input and will return corresponding articles.

Passed search word `politics`, user can pass any word, based on that word this function will return data.

```
ArticleAPI <- function(word)
{
  api_key <- "u5CCFWy8iVIWy5086FPxnA93zWhJVlCY"

  url <- paste("https://api.nytimes.com/svc/search/v2/articlesearch.json?q=", word, "&api-key=u5CCFWy8iV")

  paste0("Here is data for key-word: ", word)

  #get seems to fail sometimes
  df <- httr::RETRY("GET", url, times = 100, pause_base = 1)

  df <- content(df, as="text")
}
```

```
df <- as.data.frame(fromJSON(df))

#clean up the column names
colnames(df) <- gsub("^.*\\.\\.", "", colnames(df))

return(df)
}

ArticleAPI("politics")
```

```
##      status                                                                 copyright
## 1      OK Copyright (c) 2020 The New York Times Company. All Rights Reserved.
## 2      OK Copyright (c) 2020 The New York Times Company. All Rights Reserved.
## 3      OK Copyright (c) 2020 The New York Times Company. All Rights Reserved.
## 4      OK Copyright (c) 2020 The New York Times Company. All Rights Reserved.
## 5      OK Copyright (c) 2020 The New York Times Company. All Rights Reserved.
## 6      OK Copyright (c) 2020 The New York Times Company. All Rights Reserved.
## 7      OK Copyright (c) 2020 The New York Times Company. All Rights Reserved.
## 8      OK Copyright (c) 2020 The New York Times Company. All Rights Reserved.
## 9      OK Copyright (c) 2020 The New York Times Company. All Rights Reserved.
## 10     OK Copyright (c) 2020 The New York Times Company. All Rights Reserved.
##
## 1
## 2 President Trump is 73. His leading rival is 77. And many of their strongest supporters - vulnerab
## 3      The North Carolina Republican is seeking to clear himself through an ethics
## 4      Primary elections are being pushed back. The Olympics are
## 5      An Illinois mayor's emergency d
## 6      Governors of both parties have taken a lead role in confronting the crisis, asserting ther
## 7      With a broad coalition, Joseph R. Biden Jr. defeated Bernie Sanders in Florida
## 8
## 9      People in a country in trouble look to the president to l
## 10     Some thought he wa
##      web_url
## 1      https://www.nytimes.com/2020/03/20/us/politics/trump-coronavirus.html
## 2      https://www.nytimes.com/2020/03/27/us/politics/biden-trump-seniors.html
## 3      https://www.nytimes.com/2020/03/20/us/politics/coronavirus-richard-burr-insider-trading.html
## 4      https://www.nytimes.com/2020/03/24/us/politics/coronavirus-2020-election-primary.html
## 5      https://www.nytimes.com/2020/03/17/us/politics/coronavirus-republicans.html
## 6      https://www.nytimes.com/2020/03/17/us/politics/governors-coronavirus-trump.html
## 7      https://www.nytimes.com/2020/03/17/us/politics/biden-florida-illinois-primary.html
## 8      https://www.nytimes.com/2020/03/12/us/politics/sanders-biden-debate.html
## 9      https://www.nytimes.com/2020/03/13/us/politics/president-trump-coronavirus-speech.html
## 10     https://www.nytimes.com/2020/03/27/us/politics/john-sears-dead.html
##
## 1
## 2 President Trump is 73. His leading rival is 77. And many of their strongest supporters - vulnerab
## 3      The North Carolina Republican is seeking to clear himself through an ethics
## 4      Primary elections are being pushed back. The Olympics are
## 5      An Illinois mayor's emergency d
## 6      Governors of both parties have taken a lead role in confronting the crisis, asserting ther
## 7      With a broad coalition, Joseph R. Biden Jr. defeated Bernie Sanders in Florida
## 8
## 9      People in a country in trouble look to the president to l
```

10 Some thought he was

##

1

2

3 WASHINGTON - The r

4

5 WASHINGTON - When Deborah Frank Feinen, the mayor of Champaign, Ill., drafted an emergency p

6 WASHINGTON - One day after President

7

8

9 It seems almost churlish to compare a speech by President Donald J. Trump in 2020 to one by Prime

10

source

1 The New York Times

2 The New York Times

3 The New York Times

4 The New York Times

5 The New York Times

6 The New York Times

7 The New York Times

8 The New York Times

9 The New York Times

10 The New York Times

##

1

2

3 0,

4

5

6

7

8

9

10

headline.main

1 On Politics: What Does This All Mean for November?

2 The Most Powerful People in American Politics Are Over 65

3 Stock Sales by Senator Richard Burr Ignite Political Uproar

4 Is All of 2020 Postponed?

5 On the Political Right, Anger and Suspicion Over Virus Precautions

6 Once Political B-Listers, Governors Lead Nation's Coronavirus Response

7 Biden Sweeps Three States and Takes Commanding Lead, as Virus Reshapes American Politics

8 On Politics: A Delicate Sanders-Biden Dance

9 Can Our Political Leaders Still Reassure Us?

10 John Sears, Strategist for Nixon and Reagan, Dies at 79

headline.kicker headline.content_kicker

1 <NA> NA

2 <NA> NA

3 <NA> NA

4 Campaign Memo NA

5 <NA> NA

6 <NA> NA

7 <NA> NA

8 <NA> NA

```

## 9      News Analysis                      NA
## 10      <NA>                             NA
##
##                                     headline.print_headline
## 1                                     <NA>
## 2                                     Trump: 73. Biden: 77. Youth: Zero.
## 3                                     Stock Sales by Senator Stir a Political Uproar
## 4      With Political Calendar In Flux, Can We Expect An Election in November?
## 5                                     <NA>
## 6      In a Crisis, Governors Flexed Their Executive Muscles as the White House Balked
## 7                                     Biden Sweeps 3 States on Chaotic Day of Voting
## 8                                     <NA>
## 9      Can Our Political Leaders Still Reassure Us During Times of Crisis?
## 10     John Sears, Strategist Hired and Fired By Nixon and Reagan, Is Dead at 79
##      headline.name headline.seo headline.sub
## 1      NA          NA          NA
## 2      NA          NA          NA
## 3      NA          NA          NA
## 4      NA          NA          NA
## 5      NA          NA          NA
## 6      NA          NA          NA
## 7      NA          NA          NA
## 8      NA          NA          NA
## 9      NA          NA          NA
## 10     NA          NA          NA
##
## 1
## 2
## 3
## 4
## 5
## 6      subject, subject, subject, persons, persons, persons, persons, subject, persons, subject
## 7
## 8
## 9
## 10
##      pub_date document_type news_desk section_name
## 1      2020-03-20T10:50:32+0000      article Politics      U.S.
## 2      2020-03-27T09:00:35+0000      article Politics      U.S.
## 3      2020-03-20T15:24:58+0000      article Washington      U.S.
## 4      2020-03-24T22:33:39+0000      article Politics      U.S.
## 5      2020-03-17T13:56:18+0000      article Washington      U.S.
## 6      2020-03-17T22:54:17+0000      article Politics      U.S.
## 7      2020-03-18T01:12:08+0000      article Politics      U.S.
## 8      2020-03-12T10:56:52+0000      article Politics      U.S.
## 9      2020-03-13T17:49:37+0000      article Politics      U.S.
## 10     2020-03-27T23:01:05+0000      article      Obits      U.S.
##      subsection_name
## 1      Politics
## 2      Politics
## 3      Politics
## 4      Politics
## 5      Politics
## 6      Politics
## 7      Politics

```

```

## 8      Politics
## 9      Politics
## 10     Politics
##
##                               byline.original
## 1                               By Giovanni Russonello
## 2                               By Katie Glueck and Sabrina Tavernise
## 3 By Eric Lipton, Nicholas Fandos, Sharon LaFraniere and Julian E. Barnes
## 4                               By Matt Flegenheimer
## 5                               By Elizabeth Williamson
## 6                               By Jonathan Martin and Alexander Burns
## 7                               By Alexander Burns and Jonathan Martin
## 8                               By Giovanni Russonello
## 9                               By Sarah Lyall
## 10                              By Katharine Q. Seelye
##
## 1
## 2
## 3 Eric, Nicholas, Sharon, Julian, NA, NA, NA, E., Lipton, Fandos, LaFraniere, Barnes, NA, NA, NA, NA
## 4
## 5
## 6
## 7
## 8
## 9
## 10
## byline.organization type_of_material
## 1      NA      News
## 2      NA      News
## 3      NA      News
## 4      NA      News Analysis
## 5      NA      News
## 6      NA      News
## 7      NA      News
## 8      NA      News
## 9      NA      News Analysis
## 10     NA      Obituary (Obit)
##
##                               _id word_count
## 1 nyt://article/5e026829-862e-5ad9-9a39-20235c5a411f      1087
## 2 nyt://article/3e04f020-aff1-524c-a9ba-edf82192403c      1937
## 3 nyt://article/efe4979c-eefe-5c44-bb75-13b112249c17      1562
## 4 nyt://article/c74743eb-972a-5ca7-8de5-618439a8f250       983
## 5 nyt://article/c768c149-824c-521e-9e33-c4f6d9e427de      1375
## 6 nyt://article/2694200e-dea7-582f-845e-4c4dda049c46      1695
## 7 nyt://article/ade1f73-9071-5372-b6ed-9120482f7c70      1679
## 8 nyt://article/a8b22de7-feba-5296-bd8b-2f34341b4457      1491
## 9 nyt://article/adddf095-ec3e-5322-9acd-89b4b9616469      1258
## 10 nyt://article/0fd922e8-1780-58bc-b9e7-43ac81f215d6      1090
##
##                               uri print_section print_page
## 1 nyt://article/5e026829-862e-5ad9-9a39-20235c5a411f      <NA>      <NA>
## 2 nyt://article/3e04f020-aff1-524c-a9ba-edf82192403c      A      21
## 3 nyt://article/efe4979c-eefe-5c44-bb75-13b112249c17      A      9
## 4 nyt://article/c74743eb-972a-5ca7-8de5-618439a8f250      A      18
## 5 nyt://article/c768c149-824c-521e-9e33-c4f6d9e427de      <NA>      <NA>
## 6 nyt://article/2694200e-dea7-582f-845e-4c4dda049c46      A      11

```


## 7	nyt://article/ade1f73-9071-5372-b6ed-9120482f7c70	A	1
## 8	nyt://article/a8b22de7-feba-5296-bd8b-2f34341b4457	<NA>	<NA>
## 9	nyt://article/adddf095-ec3e-5322-9acd-89b4b9616469	A	10
## 10	nyt://article/0fd922e8-1780-58bc-b9e7-43ac81f215d6	B	15
##	hits offset time		
## 1	1241241 0 892		
## 2	1241241 0 892		
## 3	1241241 0 892		
## 4	1241241 0 892		
## 5	1241241 0 892		
## 6	1241241 0 892		
## 7	1241241 0 892		
## 8	1241241 0 892		
## 9	1241241 0 892		
## 10	1241241 0 892		

Conclusion

This assignment taught us, how to using API to get data, manipulate data and do the analysis. Using `ArticleAPI` function users can get corresponding data present in the New York Times article search API.