

트위터 분석

해쉬태그 키워드 중심

true

2020-01-26

Contents

트위터 데이터	1
트위터 추세	2
사용자 활동성 지표	2
리트윗 네트워크 분석	4
공간정보	6
트윗 텍스트 분석	6

본 보고서는 다양한 데이터 가져오기 - “트위터 데이터 과학” 웹문서를 Makefile을 작성하여 자동화하는 것을 재현하는 과정을 담고 있다.

트위터 데이터

가장 먼저 트위터 데이터를 가져와야 한다. 이를 위해서 get_data.R 스크립트에 내장된 get_data() 함수를 사용해서 트위터에서 데이터를 가져와서 이를 내부적으로 저장시킨다. 원래 DT 패키지를 사용해서 인터랙티브하게 분석 보고서를 작성하는 것이 좋으나, .pdf, .docx 파일 출력결과를 생각해서 kable() 함수를 사용해서 이를 정적 문서에서도 가시성이 좋도록 표를 작업한다. kable() 함수가 PDF, Word 파일 생성시 표를 제대로 만들어 내지 않는 경우 pandoc()로 대신한다.

```
library(tidyverse)
library(pander)

panderOptions('table.alignment.default', function(df)
  ifelse(sapply(df, is.numeric), 'right', 'left'))
panderOptions('table.split.table', Inf)
panderOptions('big.mark', ",")
panderOptions('keep.trailing.zeros', TRUE)

twit_df <- read_rds("data/tw_dat.rds")

twit_df %>%
  sample_n(10) %>%
  select(screen_name, created_at, text) %>%
  pander()
```

screen_name	created_at	text
AI_bot_ Frei_AberEinsam	2020-01-19 02:00:57 2020-01-23 15:09:49	「 피식거려도 이해해줄수있겠지....? 」#AI “누가 수구이고 누가 개혁인가는 간단하다. 오늘 할 수 있는 개혁을 내일로 미루는 사람이 수구다.” #안철수(20170109)☒ #안파고 #AI https://t.co/6p4M5Mu3wl
AI_bot_ ruikkilee1	2020-01-20 23:00:58 2020-01-22 12:36:48	「 감정회로 Error 라니....? 」#AI 국민 사이다 되시는 거 아님? ㅎ 답답하게 쌓인 국민들 마음 시원하게 뚫릴 듯.. #공감맨 #개념남 #백신맨 #안철수 #안철수박사 #AI #경제전문가 #교수 #의사 #러너 #안보전문가 #교육전문가 #천재 안철수, 문재인에 일침...“한 번도 경험해보지 못한 무능” https://t.co/Ba0aKyExWi
kaydaymay	2020-01-22 12:08:30	‘#공정’ ‘#경제’ 이어 ‘#AI’ #안철수, 카이스트 방문해 AI국가전략 경청 AI 비전 논의 기술·발전 상응하는 법·제도 방향 https://t.co/udYNxsYkPG
kmo0707	2020-01-22 09:49:10	국민 사이다 되시는 거 아님? ㅎ 답답하게 쌓인 국민들 마음 시원하게 뚫릴 듯.. #공감맨 #개념남 #백신맨 #안철수 #안철수박사 #AI #경제전문가 #교수 #의사 #러너 #안보전문가 #교육전문가 #천재 안철수, 문재인에 일침...“한 번도 경험해보지 못한 무능” https://t.co/Ba0aKyExWi
bluehhk	2020-01-25 06:54:22	환희 말 못알아듣는 A.I #업텐션 #up10tion #미치게해 #AI https://t.co/OMjeuyGFP
MurdockHansoon	2020-01-22 11:38:22	국민 사이다 되시는 거 아님? ㅎ 답답하게 쌓인 국민들 마음 시원하게 뚫릴 듯.. #공감맨 #개념남 #백신맨 #안철수 #안철수박사 #AI #경제전문가 #교수 #의사 #러너 #안보전문가 #교육전문가 #천재 안철수, 문재인에 일침...“한 번도 경험해보지 못한 무능” https://t.co/Ba0aKyExWi
Pardoe_AI gyu_bebe	2020-01-18 10:50:11 2020-01-25 06:09:05	아인- 칼내려놔 #AI 환희 말 못알아듣는 A.I #업텐션 #up10tion #미치게해 #AI https://t.co/OMjeuyGFP

트위터 추세

graph_trend.R 파일에 graph_trend() 함수를 사용해서 twit_trend.png 파일을 생성하여 특정 해시태그 트위터 데이터의 추세를 파악하고자 시각화하는 결과를 보여주는 것이 목적이다.

사용자 활동성 지표

특정 해시태그를 많이 언급한 사용자에게 대한 기본 프로필 정보를 lookup_users() 함수를 통해서 파악할 수 있고 이를 바탕으로 다음 사용자 프로필 활동성 정보를 얻어올 수 있다.

- status

시간별 #AI 트윗 추세

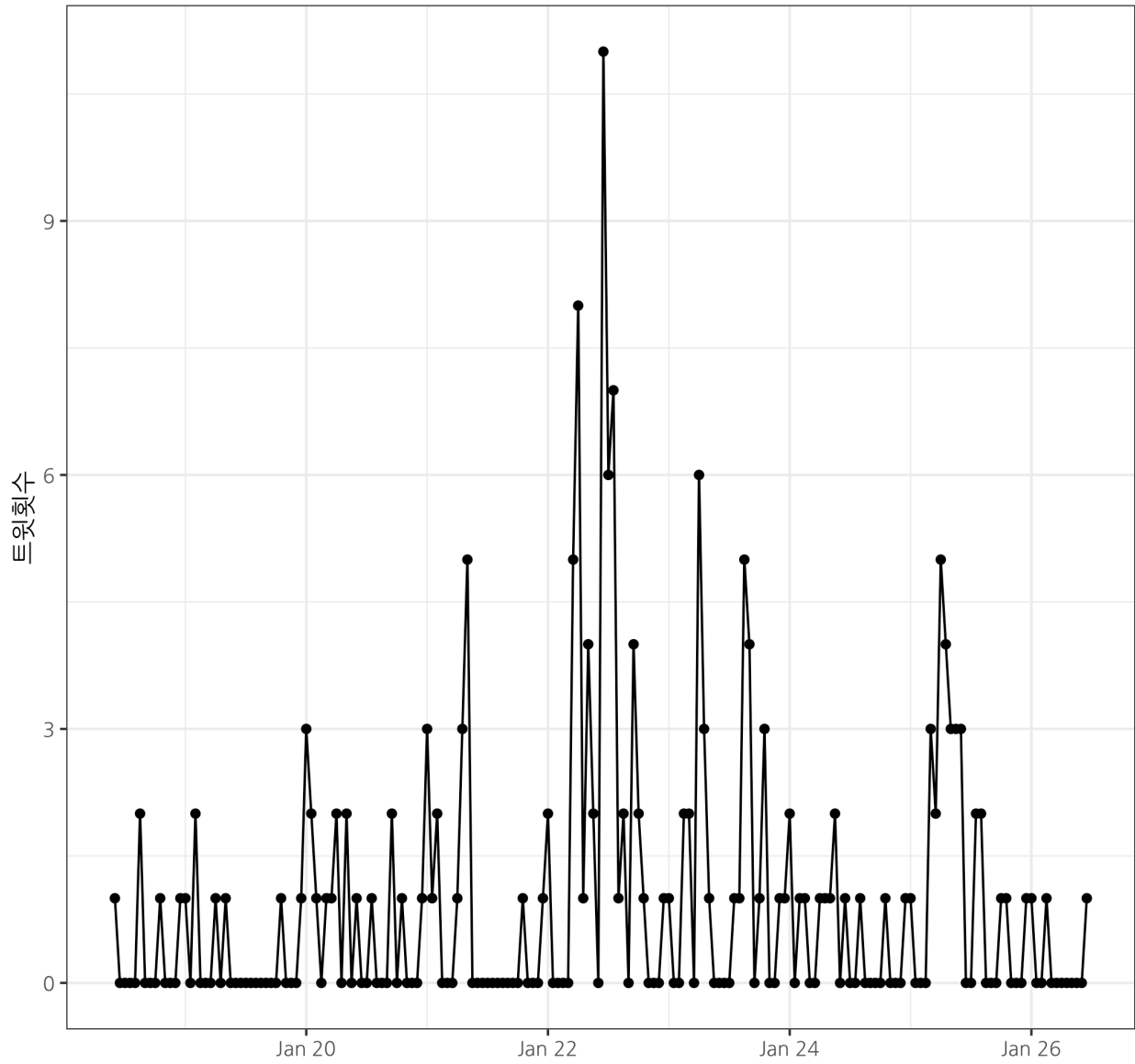


Figure 1: 트위터 해쉬태그 추세

- followers
- friends
- favourites

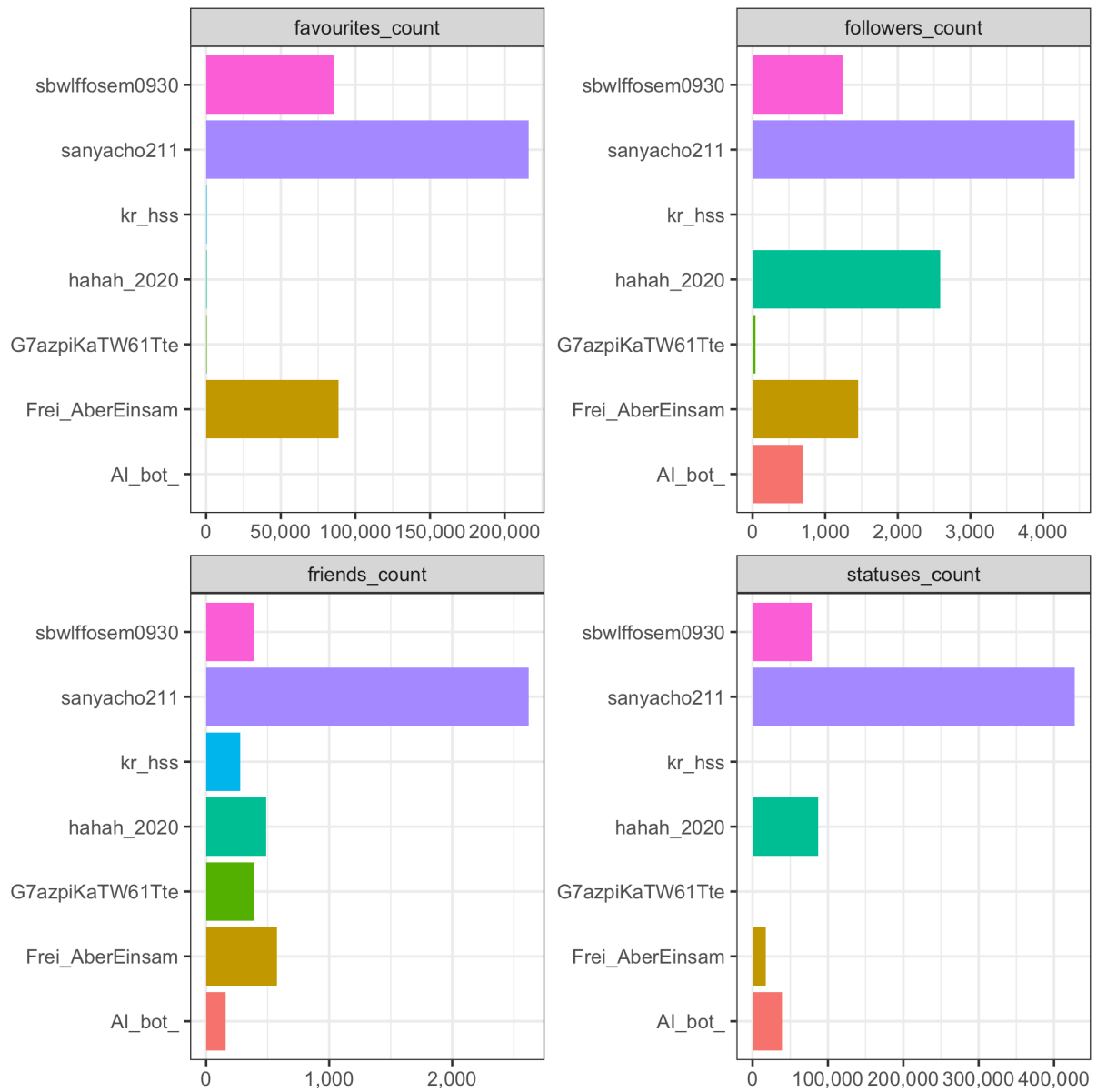


Figure 2: 트위터 해시태그 추세

리트윗 네트워크 분석

특정 해시태그 관련 트윗 검색어를 넣어 `include_rts = TRUE` 리트윗을 추가하게 되면 리트윗(retweet) from, to 자료구조를 통해서 네트워크 분석을 통해서 트윗이 어떻게 전파되는지, 전파경로를 파악할 수 있다.

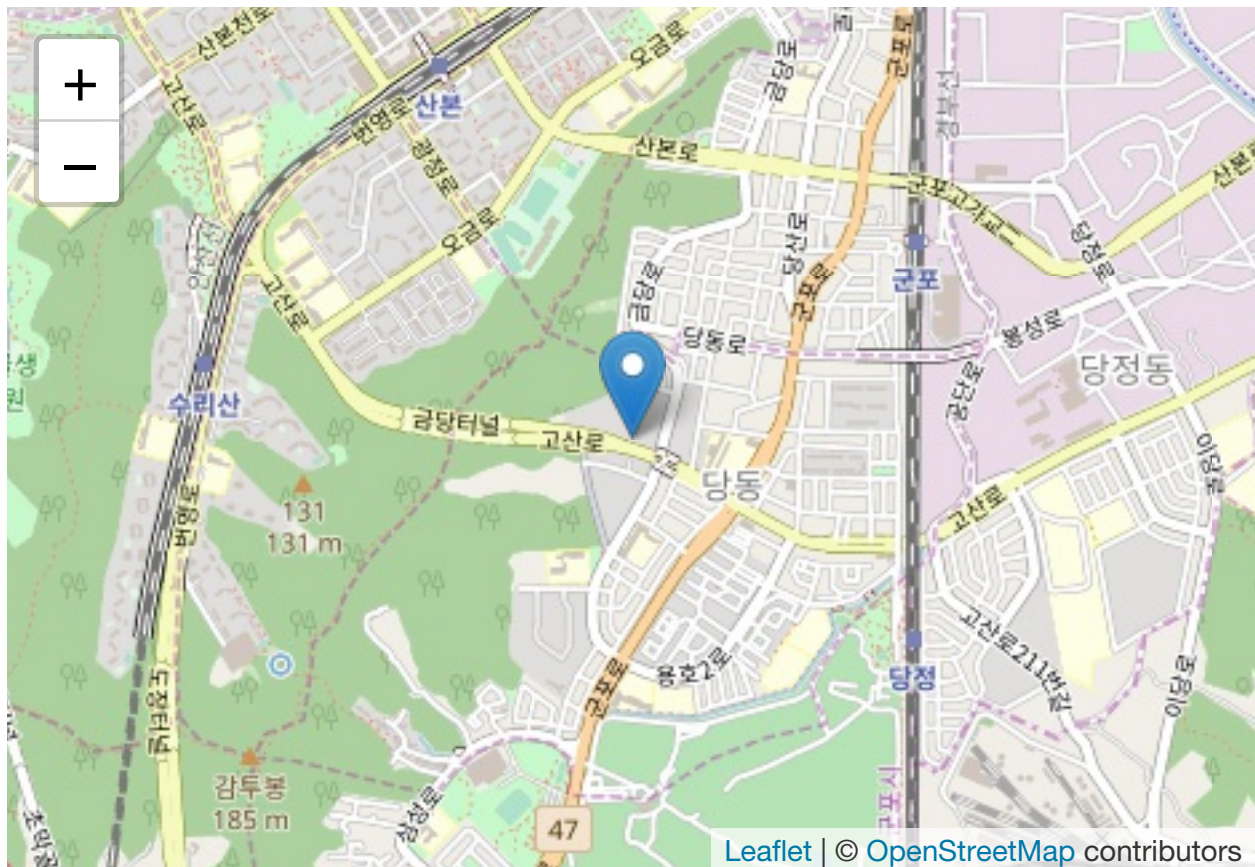
공간정보

특정 해쉬태그 관련 트윗 검색어를 넣어 어디에서 많이 하는지 공간정보가 포함되어 있는 경우 이를 반영시킨다.

```
library(rtweet)
library(leaflet)

geo_twts_df <- lat_lng(twit_df) %>%
  select(screen_name, text, lng, lat) %>%
  filter(!is.na(lng))

geo_twts_df %>%
  leaflet() %>%
  addProviderTiles(providers$OpenStreetMap) %>%
  addMarkers(lng=~lng, lat=~lat, clusterOptions = markerClusterOptions(),
    popup = ~ as.character(paste0("<strong>", paste0(" : ", `screen_name`), "</strong><br>",
      "-----",
      "&middot; : ", text)))
```



트윗 텍스트 분석

특정 해쉬태그관련 트윗 검색어를 텍스트 데이터로 놓고 이를 자연어처리 분석을 추진한다. 이를 위해서 tidytext 팩키지를 통해 단어 토큰을 만들고 불용어 처리한 후에 단어 빈도수를 시각화해 보자.

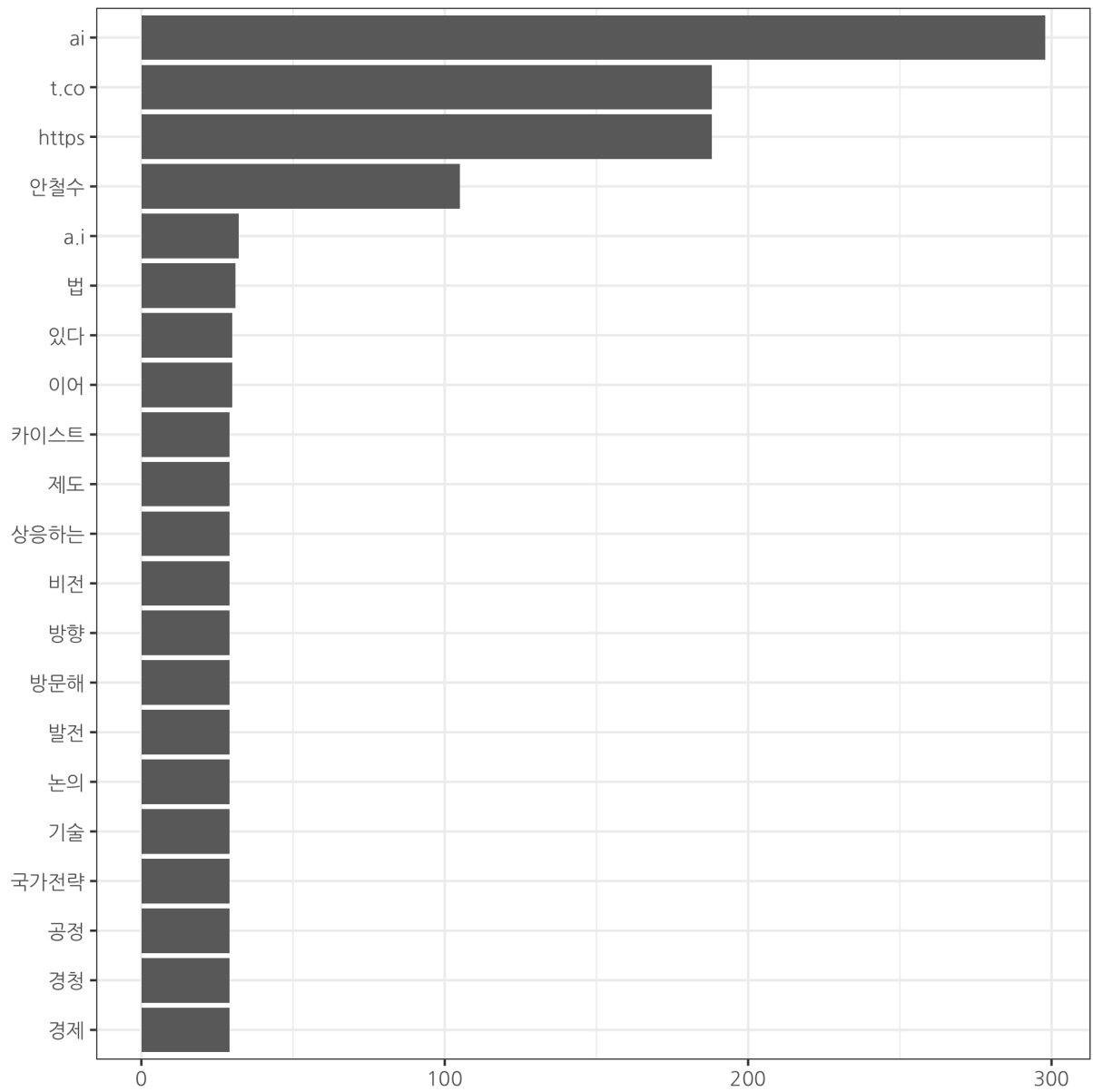


Figure 4: 트윗 텍스트 분석