네이버 카테고리별 톱기사들에 대한 사람들의 반응 분석

한상혁

2019년 5월 4일

if(T){  
   
 if(require(devtools)== F){  
   
 install.packages("devtools")  
 library(devtools)  
 }  
   
 if(require(RSelenium)== F){  
   
 install.packages("RSelenium")  
 library(RSelenium)  
 }  
   
 if(require(rvest)== F){  
   
 install.packages("rvest")  
 library(rvest)  
 }  
   
 if(require(stringr)== F){  
   
 install.packages("stringr")  
 library(stringr)  
 }  
   
 if(require(dplyr)== F){  
   
 install.packages("dplyr")  
 library(dplyr)  
 }  
   
 if(require(ggplot2)== F){  
   
 install.packages("ggplot2")  
 library(ggplot2)  
 }  
   
 if(require(reshape2)== F){  
   
 install.packages("reshape2")  
 library(reshape2)  
 }  
   
 remDr<- remoteDriver(remoteServerAddr="localhost",  
 port=4445L,  
 browserName="chrome")  
   
 remDr$open()  
 remDr$navigate("https://news.naver.com/")  
   
 Sys.sleep(2)  
 page\_source<- remDr$getPageSource()  
   
 catagory\_name<- read\_html(page\_source[[1]]) %>% html\_nodes("div.lnb\_menu ul li") %>% html\_text()  
 catagory\_name<- catagory\_name[3:8]  
   
 catagory\_url\_part<- read\_html(page\_source[[1]]) %>% html\_nodes("div.lnb\_menu ul li a") %>% html\_attr("href")  
 catagory\_url\_part<- catagory\_url\_part[3:8]  
   
 catagory\_url<- NULL  
   
 for(i in 1:length(catagory\_url\_part)){ # URL을 접근 가능한 URL로 만들어 주는 기능  
   
 catagory\_url[i]<- paste0("https://news.naver.com", catagory\_url\_part[i])  
 }  
   
 total\_article\_info\_df<- NULL  
   
 for(i in 1:length(catagory\_url)){ # 페이지 정보를 분류별 기사별로 스크래핑하는 기능  
   
 remDr$navigate(catagory\_url[i])  
   
 Sys.sleep(1)  
 more\_see\_bt<- remDr$findElement(using="css selector", "div.cluster\_more a")  
   
 while(T){  
   
 value<- tryCatch(  
 {  
 more\_see\_bt$clickElement()  
 },  
 error= function(e)  
 {  
 return (1)  
 },  
 finally= {}  
 )  
   
 if(is.numeric(value)==T){  
 if(value==1){  
 value<- 0  
 break  
 }  
 }  
   
 Sys.sleep(1)  
 }  
   
 Sys.sleep(1)  
 page\_source<- remDr$getPageSource()  
   
 title<- read\_html(page\_source[[1]]) %>% html\_nodes("a.cluster\_text\_headline") %>% html\_text()  
 title\_url<- read\_html(page\_source[[1]]) %>% html\_nodes("a.cluster\_text\_headline") %>% html\_attr("href")  
   
 for(j in 1:length(title\_url)){  
   
 remDr$navigate(title\_url[j])  
   
 Sys.sleep(1)  
 page\_source<- remDr$getPageSource()  
   
 emotion\_responses<- read\_html(page\_source[[1]]) %>% html\_nodes("span.u\_likeit\_list\_count") %>% html\_text()  
 emotion\_responses<- emotion\_responses[1:5]  
 emotion\_responses<- gsub(",", "", emotion\_responses)  
 emotion\_responses<- as.integer(emotion\_responses)  
   
 article\_info\_df<- data.frame(카테고리= catagory\_name[i], 제목= title[j], 좋아요= emotion\_responses[1], 훈훈해요= emotion\_responses[2],  
 슬퍼요= emotion\_responses[3], 화나요= emotion\_responses[4], 후속기사\_원해요= emotion\_responses[5])  
   
 total\_article\_info\_df<- rbind(total\_article\_info\_df, article\_info\_df)  
 }  
   
 message(i) # 진행 확인용  
 }  
   
 grouped\_total\_article\_info\_df<- total\_article\_info\_df %>% group\_by(카테고리) %>% summarise(좋아요= sum(좋아요), 훈훈해요= sum(훈훈해요),  
 슬퍼요= sum(슬퍼요), 화나요= sum(화나요),  
 후속기사\_원해요= sum(후속기사\_원해요))  
   
 grouped\_total\_article\_info\_df<- grouped\_total\_article\_info\_df[c(6,2,3,4,5,1),]  
 grouped\_total\_article\_info\_df$카테고리 <- factor(grouped\_total\_article\_info\_df$카테고리, levels= grouped\_total\_article\_info\_df$카테고리)  
   
 melted\_data<- melt(grouped\_total\_article\_info\_df)  
 colnames(melted\_data)<- c("카테고리", "반응", "반응\_수\_합계")  
   
 # 데이터 시각화  
   
 visualised\_data<- ggplot(melted\_data, aes(x= reorder(카테고리, -반응\_수\_합계), y= 반응\_수\_합계, fill= 반응)) +  
 geom\_bar(stat= "identity", position= "dodge", color= "black") + scale\_x\_discrete(name= "카테고리") +  
 scale\_y\_continuous(name= "표정 수 합계") + scale\_fill\_manual(values= c("#33FF00", "#FFFF00", "#3300FF", "#FF3300", "purple")) +  
 ggtitle("네이버 카테고리별 헤드라인 기사들에 \n대한 사람들의 반응 분석") +  
 theme(plot.title= element\_text(face="bold", lineheight= 1, size=15, vjust=4, hjust=0.5))  
   
 print(visualised\_data)  
}

## Loading required package: devtools

## Warning: package 'devtools' was built under R version 3.5.2

## Warning: package 'usethis' was built under R version 3.5.2

## Loading required package: RSelenium

## Loading required package: rvest

## Warning: package 'rvest' was built under R version 3.5.2

## Loading required package: xml2

## Warning: package 'xml2' was built under R version 3.5.2

## Loading required package: stringr

## Warning: package 'stringr' was built under R version 3.5.2

## Loading required package: dplyr

## Warning: package 'dplyr' was built under R version 3.5.3

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 3.5.1

## Loading required package: reshape2

## Warning: package 'reshape2' was built under R version 3.5.2

## [1] "Connecting to remote server"  
## $acceptSslCerts  
## [1] TRUE  
##   
## $applicationCacheEnabled  
## [1] FALSE  
##   
## $browserConnectionEnabled  
## [1] FALSE  
##   
## $browserName  
## [1] "chrome"  
##   
## $chrome  
## $chrome$chromedriverVersion  
## [1] "2.32.498550 (9dec58e66c31bcc53a9ce3c7226f0c1c5810906a)"  
##   
## $chrome$userDataDir  
## [1] "C:\\Users\\User\\AppData\\Local\\Temp\\scoped\_dir1948\_21269"  
##   
##   
## $cssSelectorsEnabled  
## [1] TRUE  
##   
## $databaseEnabled  
## [1] FALSE  
##   
## $handlesAlerts  
## [1] TRUE  
##   
## $hasTouchScreen  
## [1] FALSE  
##   
## $javascriptEnabled  
## [1] TRUE  
##   
## $locationContextEnabled  
## [1] TRUE  
##   
## $mobileEmulationEnabled  
## [1] FALSE  
##   
## $nativeEvents  
## [1] TRUE  
##   
## $networkConnectionEnabled  
## [1] FALSE  
##   
## $pageLoadStrategy  
## [1] "normal"  
##   
## $platform  
## [1] "Windows NT"  
##   
## $rotatable  
## [1] FALSE  
##   
## $setWindowRect  
## [1] TRUE  
##   
## $takesHeapSnapshot  
## [1] TRUE  
##   
## $takesScreenshot  
## [1] TRUE  
##   
## $unexpectedAlertBehaviour  
## [1] ""  
##   
## $version  
## [1] "74.0.3729.131"  
##   
## $webStorageEnabled  
## [1] TRUE  
##   
## $webdriver.remote.sessionid  
## [1] "c53870a5365234868cb16ced71557b68"  
##   
## $id  
## [1] "c53870a5365234868cb16ced71557b68"

##   
## Selenium message:element not visible  
## (Session info: chrome=74.0.3729.131)  
## (Driver info: chromedriver=2.32.498550 (9dec58e66c31bcc53a9ce3c7226f0c1c5810906a),platform=Windows NT 6.3.9600 x86\_64)

## 1

##   
## Selenium message:element not visible  
## (Session info: chrome=74.0.3729.131)  
## (Driver info: chromedriver=2.32.498550 (9dec58e66c31bcc53a9ce3c7226f0c1c5810906a),platform=Windows NT 6.3.9600 x86\_64)

## 2

##   
## Selenium message:element not visible  
## (Session info: chrome=74.0.3729.131)  
## (Driver info: chromedriver=2.32.498550 (9dec58e66c31bcc53a9ce3c7226f0c1c5810906a),platform=Windows NT 6.3.9600 x86\_64)

## 3

##   
## Selenium message:element not visible  
## (Session info: chrome=74.0.3729.131)  
## (Driver info: chromedriver=2.32.498550 (9dec58e66c31bcc53a9ce3c7226f0c1c5810906a),platform=Windows NT 6.3.9600 x86\_64)

## 4

##   
## Selenium message:element not visible  
## (Session info: chrome=74.0.3729.131)  
## (Driver info: chromedriver=2.32.498550 (9dec58e66c31bcc53a9ce3c7226f0c1c5810906a),platform=Windows NT 6.3.9600 x86\_64)

## 5

##   
## Selenium message:element not visible  
## (Session info: chrome=74.0.3729.131)  
## (Driver info: chromedriver=2.32.498550 (9dec58e66c31bcc53a9ce3c7226f0c1c5810906a),platform=Windows NT 6.3.9600 x86\_64)

## 6

## Using 카테고리 as id variables

