

Project 1

트랜드 파악

2023-02-10

R Markdown을 활용한 개별 프로젝트

연관검색어 Crwaling을 통한 트렌드 파악

part1 : 웹 사이트 Crwaling 준비(패키지 부착)

```
#install.packages("rvest")
# read_html( ) : 전체 html문서를 읽어오는 함수
# html_nodes( ) : 페이지에서 지정된 css위치정보를 가져옴
# html_text( ) : html tag들에서 텍스트 정보만 가져오는 함수
# html_attr( ) : html tag안의 'href'의 url 정보를 가져오는 함수
#install.packages("stringr")
# str_sub( ) : 부분적으로 관련 내용을 가져오는 함수
library(rvest)
library(stringr)
library(KoNLP)
```

```
## Checking user defined dictionary!
```

part2 : 웹 사이트 Crwaling [검색 : 산업혁명]

```
title=c()
body=c()
url='https://search.naver.com/search.naver?where=news&sm=tab_pge&query=%EC%82%B0%EC%97%85%ED%98%81%EB%AA%85&sort=0&photo=0&field=0&pd=0&ds=&de=&cluster_rank=51&mynews=0&office_type=0&office_section_code=0&news_of
fice_checked=&nso=so:r,p:all,a:all&start='
for(i in 1:50){
  turl=paste0(url,i)
  t_css=".news_tit"
  b_css=".dsc_txt_wrap"

  hdoc=read_html(turl)
  t_node=html_nodes(hdoc,t_css)
  b_node=html_nodes(hdoc,b_css)
  title_part=html_text(t_node)
  b_part=html_text(b_node)

  body_part=gsub("Wn"," ",b_part)
  body_part=str_trim(body_part,side="both")
  url_part=html_attr(t_node,"href")

  title=c(title,title_part)
  body=c(body,body_part)
}
news=data.frame(title,body)
```

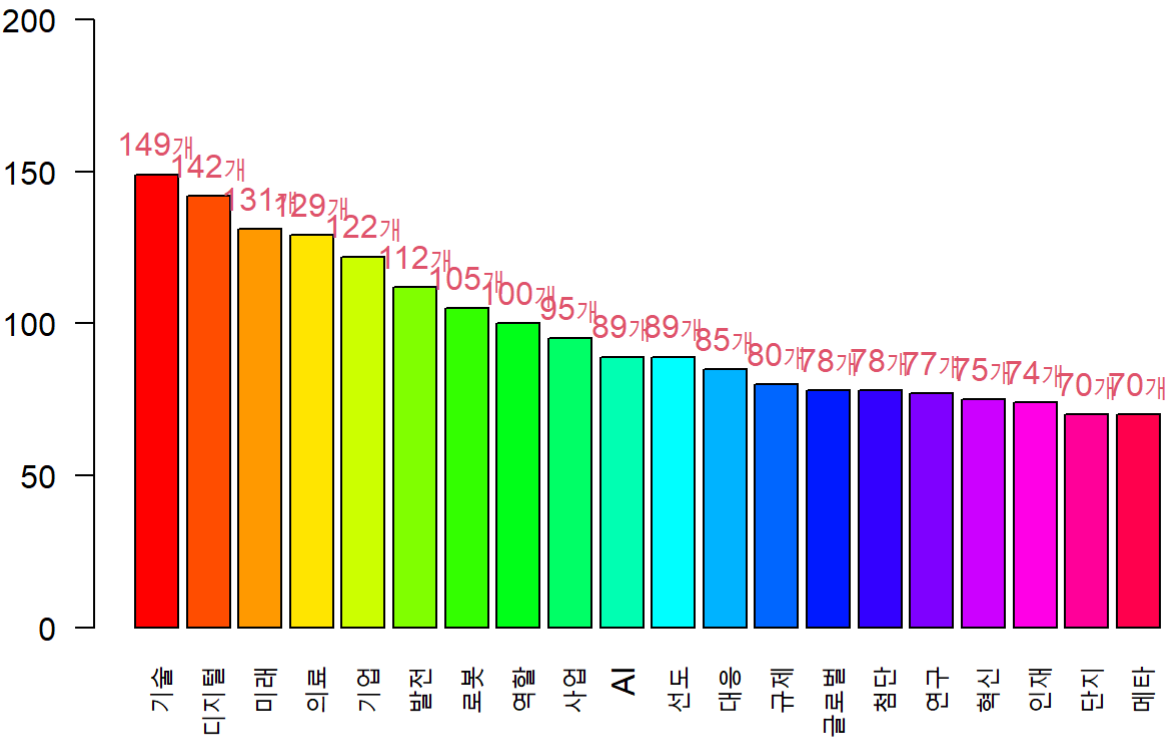
part3 : 단어 선택 정리

```
a=apply(news,extractNoun,USE.NAMES=F)
a=unlist(a)
a=gsub("산업","",a)
a=gsub("혁명","",a)
a=gsub("4차","",a)
a=gsub("시대","",a)
count=Filter(function(x){nchar(x)>=2},a)
word=table(count)
b=head(sort(word,decreasing = T),20)
print(b)
```

##	count											
##	기술	디지털	미래	의료	기업	발전	로봇	역할	사업	AI	선도	
##	149	142	131	129	122	112	105	100	95	89	89	
##	대응	규제	글로벌	첨단	연구	혁신	인재	단지	메타			
##	85	80	78	78	77	75	74	70	70			

part4 : 막대 그래프

```
c=barplot(b,col=rainbow(20),ylim=c(0,200),las=2)
text(c,b,paste0(b,"개"),pos=3,col=2,cex=1)
```



워드클라우드 패키지 부착

```
library(wordcloud)
```

```
## 필요한 패키지를 로딩중입니다: RColorBrewer
```

```
library(wordcloud2)
```

part5 : wordcloud

```
palate=brewer.pal(9,"Set1")
wordcloud(names(b),
  freq = b,
  scale = c(6,0.5),
  rot.per = 0.25,
  min.freq = 2,
  random.order = F,
  random.color = T,
  colors = palate)
```



```
wordcloud2(data=b,size=0.5)
```

