DB 2학기 수행평가 - 3323 사호준

3323 사호준

2022-09-06

## 문제 정의

ICT 산업의 중요도가 높아지고 있는 4차 산업혁명 시대에 미래 진로를 설계할 수 있도록 인공지능과 빅데이터 간의 상관관계를 조사해 보고자 한다.

## 데이터

위키백과(<https://ko.wikipedia.org/)에서> ‘빅데이터’와 ’인공지능’ 문서를 검색해 문서 내용을 복사하여 각각 ‘big\_data.txt’, ‘artificial\_intelligence.txt’ 파일로 저장하였다.

## 분석 코드

파일을 읽은 후 텍스트를 정제하여 {파일 이름+2}.txt 파일로 다시 저장한 후, 워드클라우드로 시각화하였다.

setwd("c:\\R\_temp")  
  
# 패키지 설치  
# install.packages("multilinguer")  
library(multilinguer)  
# install\_jdk( )  
# install.packages(c("hash", "tau", "Sejong", "RSQLite", "devtools", "bit", "rex", "lazyeval", "htmlwidgets", "crosstalk", "promises", "later", "sessioninfo", "xopen", "bit64", "blob", "DBI", "memoise", "plogr", "covr", "DT", "rcmdcheck", "rversions"), type = "binary")  
# install.packages("remotes")  
# remotes::install\_github('haven-jeon/KoNLP', upgrade = "never", INSTALL\_opts=c("--no-multiarch"))  
  
  
#install.packages("KoNLP")  
# install.packages("wordcloud")  
# install.packages("stringr")  
  
  
library("KoNLP")  
library("wordcloud")  
library("stringr")  
  
  
#install.packages("waldo")  
#install.packages("knitr")  
  
  
  
useSejongDic()

## Backup was just finished!  
## 370957 words dictionary was built.

# extractNoun("아버지가 방에 들어가신다")  
  
#mergeUserDic(data.frame(readLines("mergefile.txt"), "ncn"))  
  
# 데이터 분석 및 시각화  
for (filename in c("big\_data", "artificial\_intelligence")) {  
 txt <- readLines(paste(filename, ".txt", sep=""), encoding="UTF-8")  
 txt  
 place <- sapply(txt,extractNoun,USE.NAMES = F)  
   
 class(place)  
 head(place, 2)  
 c <- unlist(place)  
 c  
 place <- Filter(function(x){nchar(x) >=2}, c)  
 place  
 res <- str\_replace\_all(place,"[^[:alpha:]]","")  
 res <- res[res != ""]  
   
 # res <-gsub("제주","",res)  
   
   
 write(res, paste(filename, "2.txt", sep=""))  
 res2 <-read.table(paste(filename, "2.txt", sep=""))  
 class(res2)  
 head(res2)  
 wordcount <- table(res2)  
 head(sort(wordcount, decreasing=T),30)  
   
 library(RColorBrewer)  
 palete <- brewer.pal(9,"Set1")  
   
 wordcloud(names(wordcount),freq = wordcount,scale = c(5,1),rot.per=0.25,  
 min.freq=2,random.order = F,random.color = T, colors = palete)  
}



## 분석 결과와 의견

‘빅데이터’ 문서에서는 데이터, 빅데이터, 분석, 처리, 활용, 기술, 수집, 정보 등의 단어가 확인되었고, ‘인공지능’ 문서에서는 인공지능, 연구, 지능, 시스템, AI, 기계, 문제, 인간, 사용, 사람 등의 단어가 확인되었다. 예상과 달리 단어들 간에 특별히 교집합이 존재하지는 않지만, 인공지능 시스템을 위해서는 빅데이터가 필요하다는 결론을 이끌어낼 수 있었다. 빅데이터를 수집, 처리, 가공, 분석하여 이를 인공지능의 학습에 활용하면 실제 세계의 문제를 해결할 수 있다는 것이다. 따라서 미래 사회에서는 인공지능과 함께 인공지능의 발전의 토대가 되는 빅데이터 역시 관심을 가지고 공부할 필요가 있을 것이다.