

11강. 텍스트 데이터의 시각화 2

◆ 담당교수 : 이정진

들어가기

■ 주요용어

용어	해설
워드 클라우드	텍스트 데이터베이스의 한 문서에서 단어의 출현 빈도를 이용하여 많이 나타난 단어일수록 더 큰 글자로 화면 가운데 배치하여 단어 구름을 만들어 시각화하는 그림
UTF-8 인코딩 방식	R에서 한글 텍스트 문서를 처리하기 위해서는 저장 인코딩(encoding) 방식을 UTF-8 로 지정하여야한다

연습문제

1. R에서 한글 텍스트 문서를 처리하기 위해서는 저장 인코딩(encoding)방식을 ()로 지정하여야 한다.

정답 : UTF-8

2. 한글 워드 클라우드를 위해 특별히 필요한 R 패키지는?

- ① ggplot
- ② maps
- ③ sp
- ④ KoNLP

정답 : ④

해설 : 한글 단어의 워드 클라우드에 반드시 필요한 R패키지는 KoNLP 이다.

3. 한글 문서에 있는 단어들이 무엇인지 살펴보는 R 명령어는?

- ① unlist

- ② corpus
- ③ table
- ④ sort

정답 : ①

해설 : unlist는 문서 안에 있는 단어를 추출할 수 있는 명령어이다.

4. 워드 클라우드를 작성하는 R 명령어는?

- ① unlist
- ② corpus
- ③ table
- ④ word cloud

정답 : ④

해설 : word cloud가 단어구름을 만드는 명령어이다.

5. 트위터 문서의 워드 클라우드를 위해 특별히 필요한 R 패키지는?

- ① ggplot
- ② twitterR
- ③ KoNLP
- ④ sp

정답 : ②

해설 : 단어 구름을 워드 클라우드(word cloud)라 한다.

	정리하기
--	-------------

1. 한글 문서의 워드 클라우드는 영어 문서의 경우와 유사하나 한글 텍스트 파일을 R의 KoNLP 패키지를 이용하여야 한다.
2. 한글 문서를 R에서 처리하기 위해서는 저장 인코딩(encoding) 방식을 "UTF-8" 로 지정하여야 한다.
3. 트위터 문서의 워드 클라우드 작성은 한글과 유사하나 자신의 트위터 계정을 이용하여야 한다.

참고자료

1. R 프로그램 사이트 : <http://tm.r-forge.r-project.org/>
2. <http://cran.r-project.org/>
3. <http://www.ddokbaro.com/>
4. <http://statistical-research.com/>
5. <http://davetang.org/>