

<데이터 기말 이론 예상문제>

1. 텍스트 기본 전처리 4가지를 설명하시오.
2. TDM과 DTM의 차이점을 설명하시오.
3. 기존 전처리 문제점을 두개 이상 서술하시오.
4. 코사인 유사도의 식을 한글로 풀어서 설명하시오.
5. TF-IDF에서 DF가 아닌 IDF를 사용하는 이유는?
6. IDF에서 로그를 취하는 이유와 계산할 때 분모에 1을 더하는 이유는?
7. 전치(Transposed)/단위(Identity)/역(Inverse)/직교(orthogonal)
/대각(Diagonal) 행렬에 대해 설명하시오.
8. LSA에서 Full SVD가 아닌 Truncated SVD를 사용하는 이유는?
9. LSA에서 초기에 설정하는 토픽의 수를 크게 잡으면 어떻게 되는지 설명하시오.
10. TDM을 기반으로 만들어진 Truncated SVD의 3가지 행렬에 대해 설명하시오.
11. LSA의 특징을 LDA와 비교하여 설명하시오.
12. LDA 문서 생성 과정을 설명하시오.
13. LDA 추론 과정과 그로 인해 얻을 수 있는 정보를 설명하시오.
14. Word Embedding 기법에 대해서 설명하시오.
15. LDA 과정에서 모수를 설정해주는 이유와 모수의 값이 커질수록 어떤 변화가 생기는지 설명하시오.
16. 마스킹을 활용한 LDA수행과정을 설명하시오.
17. 쿨리히스베르크 다리 문제에 대해 설명하시오.

18. **케빈 베이컨 6단계 법칙**을 예시를 들어 설명하시오.
19. **연결망의 종류** 두가지를 비교하여 설명하시오.
20. 구글의 **페이지랭크 알고리즘**에 대해 설명하시오.
21. **컴퍼넌트, 연결점, 브릿지**에 대해 설명하시오.
22. **비방향그래프에서의 평균연결정도** 식을 설명하시오.
23. 방향그래프에서의 **인디그리, 아웃디그리**에 대해 설명하고, 그 식에 대해 설명하시오.
24. **에고 네트워크**에 대해서 설명하시오.
25. **사건, 인접도 매트릭스**를 비교하여 설명하시오.
26. **밀도**에 대해 설명하고, 비방향&방향&계량 네트워크에서의 식을 나타내시오.
27. **중심성**에 대해 설명하고, 중심성이 높다고 말할 수 있는 경우를 2가지 이상 설명하시오.
28. **내향&외향, 포인트&그래프, 로컬&글로벌 중심성**을 각각 비교하여 설명하시오.
29. **연결정도&근접&매개 중심성**을 설명하고, 각각의 식을 나타내시오. (절대적/상대적)
30. **근접&매개 중심성**을 정보통제 측면에서 비교하여 설명하시오.
31. **집중도**를 중심성&밀도와 비교하여 설명하시오.
32. **연결정도&근접&매개 집중도**의 계산식을 나타내시오.
33. 감성분석에서, **기계학습&감성사전 기반 접근법**을 비교하며 설명하시오.
34. 경쟁력분석 측면에서, **속성감성분석**이 어떻게 활용될 수 있는지 설명하시오.