21. 11. 14

계층적 군집화 추가! 해서, cluster 개수 정하기

실루엣, 엘보우

1. 기업 개요 data

다른 형태소 분석기( komoran, kkoma…) 를 써보면서 더 좋은 분석기를 찾아보자.

패키지에 -> 사용자 사전 지정 ex) 바이오, 삼성 그룹, 낸드 플래쉬, 한국,

<https://humahumahuma.tistory.com/260>

1. 사전이 생각한대로 잘 작동하는지 확인해보고,
2. 이 사전을 만들어 나가는 작업을 좀 중점적으로 해나가기.
3. 그 이후에, 한글자짜리 단어들 (및, 등, 임, 함. 을 제거) -> 보는 족족 적어놓고, 나중에 stop words 안에 넣든지, 전처리로 제거를 해주든지 하는 걸로.
4. 영어단어들, 따로 떼서 morph 에 합치기
5. (DRAM, 디램), (SK, 에스케이) (플래쉬, 플래시 : 외래어 표기법에 따른 차이)

.nouns + 영어

1. 주가 변동 data
2. Kmeans 가 아닌 다른 알고리즘 을 찾아본다.
3. 실루엣, 엘보우, 계층적 군집화