정보검색

컴퓨터학부 2019117865

김나형

형태소 분석

고려사항 : 조사와 같은 frequent terms는 가치가 없기 때문에 DF 고려할 필요 없어서 명사만을 형태소 분석에서 명사 추출

KoNLPy 라이브러리 사용

<https://konlpy.org/ko/latest/#start> 공식 문서를 따라가며 환경 세티

텍스트이(가) 표시된 사진

자동 생성된 설명

라이브러리 설치

문자열 파싱

<https://hiseon.me/python/python-string-parse/> 을 참고하여

텍스트이(가) 표시된 사진

자동 생성된 설명

Parse 설치

파일을 읽어오기 위해 주어진 full\_corpus.hwp를 corpus.txt로 저장

Td-idf를 weight로 사용. Wt,d = tf \* idf