개발환경: - Pycharm(python 2.7) windows

-------------------------------------------------------------------------------------------------

1. **txt\_pre.py : 형태소 분석 결과 텍스트 파일 속 불필요한 ‘ ’, ‘\n’등을 제거****(실행시간 32분)**

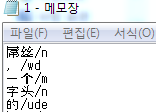
입력: 형태소 분석 후 결과 파일 /books\_result/1.txt, 2.txt, ...

출력: 처리된 형태소 분석결과 /text/1.txt, 2.txt, ...

실행 예시: python txt\_pre.py 800 (windows cmd에서, 소설 개수가 800개일때)

(의미 : text폴더의 1.txt부터 800.txt까지 순차적으로 처리하여 Txt 폴더에 출력파일을 생성함)

C:\Users\Jane\AppData\Local\Temp\1499968221(1).png



-------------------------------------------------------------------------------------------------

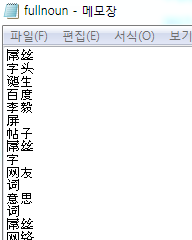
1. **mfullnoun.py : 형태소 분석 결과 텍스트 파일 속 의미 있는 단어만 저장**

입력: 처리된 형태소 분석결과 /text/1.txt, 2.txt, ...중 한 개 소설 선택

출력: 소설 속 명사만 저장 mfullnoun.txt

실행 예시: python mfullnoun.py 1 (windows cmd에서, 선택한 소설파일 이름이 1.txt일때)

C:\Users\Jane\AppData\Local\Temp\1501657724(1).png

****

-------------------------------------------------------------------------------------------------

1. **remove.py : 중복된 단어 제거**

입력: 소설 속 명사 mfullnoun.txt

출력: 중복된 명사를 제거 remove.txt

실행 예시: python remove.py (Gensim-word2vec를 이용하여 중복된 단어를 제거 할 수 있다)

-------------------------------------------------------------------------------------------------