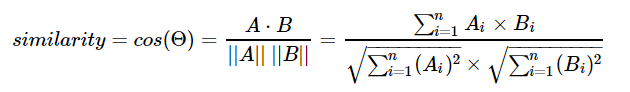
**한글 문장의 유사도 계산 (3)**

사용된 대용량 말뭉치: KCCq28\_Q01.txt (31.29MB, 20만줄)

사용된 입력 문장: 안녕하세요 이것은 유사도 검사 테스트입니다.

사용된 유사도 계산식: 코사인 유사도



실행 방법:

python sync\_test.py [input-text] [-option]

* Option은 입력과 가장 유사한 문장을 몇 개 추출할 것인지 정하는 것이므로 유사도 높은 상위 10개를 출력하고 싶으면 -10을 주면 된다.
* 여기서는 KCCq28\_Q01.txt파일을 input으로 주었으며 유사도가 높은 상위 10개를 출력하기 위해 옵션으로 -10을 주었음
* 입력 문장으로는 “안녕하세요 이것은 유사도 검사 테스트입니다.” 를 터미널에 타이핑하였다.
* python sync\_test.py KCCq28\_Q01.txt -10
* 실행시간: 약 2725초 -> 약 45분 41초
* 만약 100만줄 input text를 주게 되면 20만줄에 46분쯤 걸리므로 230분, 즉 3시간 50분쯤 걸린다.

