

1. 한글 문장의 유사도 계산(1)

1) 공통 음절 찾기

- 문장마다 공백을 없애고 따로 배열에 저장
- 특수 문자 제외
- 첫번째 문장의 음절을 key로 Dictionary 생성
- 두번째 문장의 음절이 첫번째 문장의 Dictionary에 들어가는지 검사
- 두 문장의 공통 음절 개수 세기
- 유사도 구하기

2) 공통 어절 찾기

- 문장의 공백을 기준으로 나누어 배열에 저장
- 특수 문자 제외
- 첫번째 문장의 어절을 key로 Dictionary 생성
- 두번째 문장의 어절이 첫번째 문장의 Dictionary에 들어가는지 검사
- 두 문장의 공통 어절 개수 세기
- 유사도 구하기

입력 텍스트 파일 내용

촉촉한 초코칩나라의 촉촉한 초코칩은 촉촉한 초코칩 집에서 부유하게 사는 촉촉한 초코칩이다.
안 촉촉한 초코칩나라의 안 촉촉한 초코칩은 안 촉촉한 초코칩 집에서 가난하게 사는 안 촉촉한 초코칩이다.

출력 화면

```
E:\#\학 교#\대학 교#\4학년 1학기#\빅데이터#\문장_유사도_검사>python sync_test.py
공통 음절의 갯수: 37 개
공통 음절에 대한 유사도 94.87 %
공통 어절의 갯수: 10 개
공통 어절에 대한 유사도 90.91 %
```

실행 방법

- 1) `sync_test.txt` 에 두 문장을 입력한다.
- 2) 터미널에 `python sync_test.py` 를 입력하여 실행한다.