1.음절 ngram 방식

실행 방법: python sync_test2.py [-option]

Option:

Default: Bigram을 이용한 유사도 검사

'-2' Bigram을 이용한 유사도 검사

'-3' Trigram을 이용한 유사도 검사

'-5' Bigram + Trigram을 이용한 유사도 검사

유사도 계산식: (두 문장의 공통 ngram 개수)/(짧은 문장의 모든 ngram 개수) * 100

1) Bigram 실행화면

```
E:₩학교₩대학교₩4학년 1학기₩빅데이터₩2주차_문장_유사도_검사>python sync_test2.py -2
Bigram:  84.62 %
```

2) Trigram 실행화면

```
E:₩학교₩대학교₩4학년 1학기₩빅데이터₩2주차_문장_유사도_검사>python sync_test2.py -3
Trigram:  73.68 %
```

3) Bigram + Trigram 실행화면

```
E:₩학교₩대학교₩4학년 1학기₩빅데이터₩2주차_문장_유사도_검사>python sync_test2.py -5
Bigram + Trigram:  79.22 %
```

2. 형태소 분석기(konlpy사용)

실행 방법: python sync_konlpy.py

유사도 계산식: (두 문장의 공통 형태소 개수)/(긴 문장의 모든 형태소 개수) * 100

형태소 분석기 실행 화면

```
E:₩학교₩대학교₩4학년 1학기₩빅데이터₩2주차_문장_유사도_검사>python sync_konlpy.py
Kkma_Class
중복이 많은 형태소 상위 3개:  ('하', 5) ,  ('촉촉', 4) ,  ('ㄴ', 4)
총 중복 형태소 개수:  35 개
긴 문장의 형태소 개수:  41 개
유사도(총 중복 형태소 개수/긴 문장의 형태소 개수):  85.37 %


Hannanum_Class
중복이 많은 형태소 상위 3개:  ('ㄴ', 6) ,  ('촉촉하', 4) ,  ('초코칩', 3)
총 중복 형태소 개수:  25 개
긴 문장의 형태소 개수:  30 개
유사도(총 중복 형태소 개수/긴 문장의 형태소 개수):  83.33 %


Komoran_Class
중복이 많은 형태소 상위 3개:  ('하', 5) ,  ('촉촉', 4) ,  ('ㄴ', 4)
총 중복 형태소 개수:  32 개
긴 문장의 형태소 개수:  37 개
유사도(총 중복 형태소 개수/긴 문장의 형태소 개수):  86.49 %


Okt_Class
중복이 많은 형태소 상위 3개:  ('촉촉한', 4) ,  ('초코', 4) ,  ('칩', 4)
총 중복 형태소 개수:  20 개
긴 문장의 형태소 개수:  25 개
유사도(총 중복 형태소 개수/긴 문장의 형태소 개수):  80.0 %
```