**\*개발 환경\***

**-Python 2.7.12 / Anaconda custom (64-bit) on linux**

**-Visual studio 2010 (window)**

**\*주의사항\***

**-텍스트 파일을**

**python 프로그램의 입력으로 사용할 때는 UTF-8 인코딩**

**cpp 프로그램의 입력으로 사용할 때는 ANSI 인코딩을 준수하여야 올바른 결과를 얻을 수 있음.**

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

**1. cls.cpp : 소설 속 명사들과 소설 속에 등장하는 키워드 들의 유사도를 이용하여**

각 명사들이 어떤 카테고리와 가까운지 계산

입력: LimaFull\_w2v.txt, noun\_numbering.txt, cat\_appear(numbering).txt, cat\_appear\_num.txt, category\_small\_red.txt

출력 1

distance(no th).txt

<형식> 소설 내 명사들에 대해 유사 카테고리 top 5와 유사도 기록

word : 명사1 category : 유사카테고리1-1 유사도1-1 유사카테고리1-2 유사도1-2 유사카테고리1-3 유사도1-3 유사카테고리1-4 유사도1-4 유사카테고리1-5 유사도1-5

word : 명사2 category : 유사카테고리2-1 유사도2-1 유사카테고리2-2 유사도2-2 유사카테고리2-3 유사도2-3 유사카테고리2-4 유사도2-4 유사카테고리2-5 유사도2-5

...

출력 2

distance(th=13).txt

<형식>

distance(no th).txt와 같음.

임계값을 두어 유사도가 14 이하인 명사는 기록하지 않음.

실행 예시 : widow visual studio 이용하여 컴파일 및 실행(linux gcc 무관)

(사용자 입력 : 카테고리 개수 입력)

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