

# NLP hw#3 : HMM pos tagging 제작

120150241 김상근

## 1. 개발 환경

Unix 기반 컴퓨터에서 Python 2.7.10 버전으로 작성하였다. 주어진 파일의 인코딩이 Unix 기반 컴퓨터에서 깨지는 현상이 발생하여 train.txt 파일의 인코딩을 UTF-8 로 변경하여 프로젝트를 진행하였다. input.txt, result.txt, train.txt 세 개의 파일을 a.py 와 같은 디렉토리에 위치시킨 뒤, python a.py 로 실행하면 결과가 output.txt 에 저장된다. (표준 출력으로도 출력한다.)

## 2. 프로그램 플로우 설명

get\_train\_count(), get\_input\_datas(), get\_result\_datas() 세 함수를 통하여 파일의 정보를 읽어오고, 각각의 input data 에 대하여 hmm() 함수를 호출하여 확률이 가장 높은 형태소 열을 출력한다. 자세한 내용은 프로그램에 주석을 상세히 달았다.