

Homework #3
406.424 인터넷응용
2016년도 봄학기

제출기한: 2016년 4월 30일 23:59까지

1. (50점) **a-z-cons.txt** 파일을 이용하여, 아래의 작업을 수행하시오.
(1-1과 1-2의 경우 문제에서 요구하는 txt 파일을 제출하고, 1-2의 경우 결과 출력에 사용한 NamedEntityRecognizer.java 파일도 함께 제출하시오. 1-3에서는 추가로 제시한 classifier에 대하여 1-1과 1-2에서 요구한 txt 파일을 다시 제출할 필요는 없으며, 결과 비교 및 의견 서술 내용만을 MS-Word 문서 파일 형태로 작성하여 제출하시오.)
 - 1-1. (20점) 주어진 텍스트 파일에 대하여, 아래의 classifier를 사용하여 Named Entity Recognition을 수행한 뒤, 그 결과를 **a-z-cons-ner.txt** 파일로 출력하시오.
 - **english.all.3class.distsim.crf.ser.gz**
 - 1-2. (20점) 1-1의 수행 결과에 대하여, 각각의 entity가 몇 개 나왔는지 세어 그 결과를 **a-z-cons-ner-count.txt** 파일로 출력하시오(entity 순서는 자유).
 - 예시: PERSON: 3, ORGANIZATION: 2, LOCATION: 6
 - 1-3. (10점) 아래의 classifier를 사용하여 1-1, 1-2 과정을 반복한 뒤, 그 결과를 서로 비교하시오. 각 classifier를 사용했을 때의 차이점에 대해 본인의 의견을 서술하시오.
 - **english.conll.4class.distsim.crf.ser.gz**
 - **english.muc.7class.distsim.crf.ser.gz**
2. (50점) **pap.txt** 파일을 이용하여, 아래의 작업을 수행하시오.
(2-1과 2-2의 경우 문제에서 요구하는 txt 파일을 제출하고, 2-2의 경우 결과 출력에 사용한 POSTagger.java 파일을 함께 제출하시오. 2-3에서는 추가로 제시한 tagger에 대하여 2-1과 2-2에서 요구한 txt 파일을 다시 제출할 필요는 없으며, 결과 비교 및 의견 서술 내용만을 MS-Word 문서 파일 형태로 작성하여 제출하시오.)
 - 2-1. (20점) 주어진 텍스트 파일에 대하여, 아래의 tagger를 사용하여 POS Tagging을 수행한 뒤, 그 결과를 **pap-tag.txt** 파일로 출력하시오.
 - **english-bidirectional-distsim.tagger**
 - 2-2. (20점) 2-1의 수행 결과에 대하여, 각각의 tag가 몇 개 나왔는지 세어 그 결과를 **pap-tag-count.txt** 파일로 출력하시오(tag 순서는 자유).
 - 예시: JJ: 7051, RB: 9277, DT: 9233, TO: 4138, ...
 - 2-3. (10점) 아래의 tagger를 사용하여 2-1, 2-2 과정을 반복한 뒤, 그 결과를 서로 비교하시오. 각 tagger를 사용했을 때의 차이점에 대해 본인의 의견을 서술하시오.
 - **english-left3words-distsim.tagger**
 - **wsj-0-18-left3words-nodistsim.tagger**

ETL 사이트에 업로드된 예시 코드를 응용하시오. 과제는 MS-Word나 PDF의 파일 형태로, txt 파일 및 Java 코드 결과물과 함께 압축하여 ETL 사이트에 제출하시오.