

프로그래밍 과제3

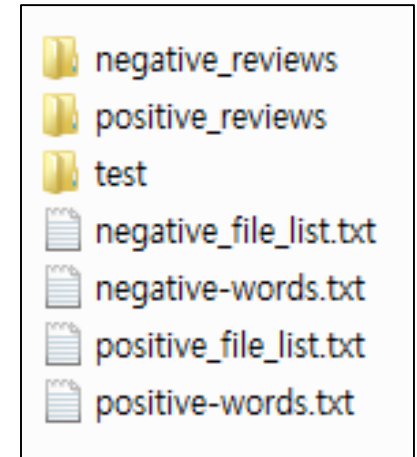
임현기

Sentiment Analysis

- IMDb 리뷰 분석
- review 결과
 - positive (thumbs up)
 - negative (thumbs down)
- 어떻게 positive/negative를 분류할 수 있을까?
 - review안에 들어있는 단어를 분석

Sentiment Analysis

- negative_reviews, positive_reviews
 - 900개의 부정/긍정 리뷰
- test
 - 10개의 리뷰
- negative-word.txt, positive-word.txt
 - 부정적/긍정적 단어 목록
- negative_file_list.txt, positive_file_list.txt
 - 900개 파일 이름 목록



Sentiment Analysis

- 900개의 리뷰를 보면서 나타난 단어 세기
 - 파일 읽기, 문자열 토큰화

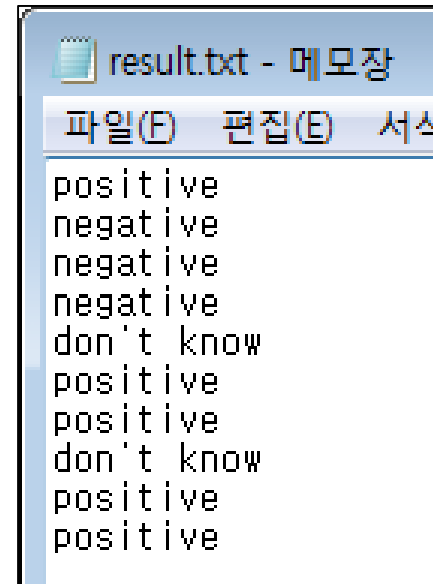
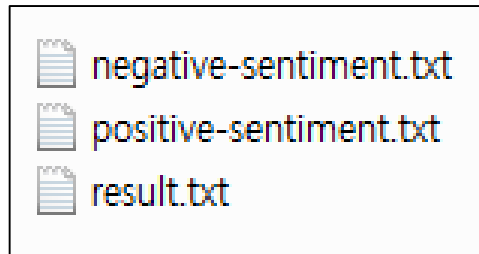
excellent	27
good	45
positive	5
...	...

anger	12
cry	35
negative	16
...	...

- 리뷰에서 한 번도 나타나지 않은 단어 삭제 & 단어 카운트 저장
 - 파일 쓰기

Sentiment Analysis

- positive / negative 분류 알고리즘 구현
 - 입력: 리뷰 파일, 단어 리스트
 - 출력: positive / negative / don't know
 - 10개의 리뷰 파일의 분류한 결과를 파일로 쓸 것



Sentiment Analysis

- 11월 27일 일요일 23시 50분까지 업로드
- main.c, negative-sentiment.txt, positive-sentiment.txt, result.txt, 보고서,
5개 파일을 압축해서 학번.zip으로 올릴 것
- 보고서
 - 사용 설명서
 - 프로그램 내 함수 기능 설명
 - 분류 알고리즘 설명