

# COMP319 Algorithm 1, Spring 2020

## Programming assignment 4

Instructor: 장길진([gjang@knu.ac.kr](mailto:gjang@knu.ac.kr), Tel 053-950-5517)

TA: 박진희 ([pjhdrrm@gmail.com](mailto:pjhdrrm@gmail.com), Tel 053-940-8617)

### General information

#### 1. 채점기준:

- a. **첫출의 주석으로 학번과 이름 제대로 명시: 20%**
  - i. 제대로 제출한 학생에게 기본점수를 주기 위함입니다. 꼭 쓰도록 하세요.
  - ii. 개발자를 명시하는 것은 지식재산권 보호를 위해 꼭 필요합니다. 습관처럼 적어야 합니다.
- b. 정확도: 30% - 3 개의 다른 예제를 돌려보고 출력파일들의 정확도 평가
- c. 실행시간: 20% - 전체 실행시간을 측정하여 상대 평가.
  - i. 제출한 코드가 compile 이 안 될 경우 실행불가 → 정확도와 실행시간 점수(총 50%) 0 점
- d. 코드 reading 점수 30% - 문제에서 요구하는 조건들을 만족시켰는가?
  - i. 문제에서 구현하라는 것을 무시하고 단순히 출력만 같을 경우 감점
  - ii. 문제에서 요구하지 않은 파일(예. 입/출력 파일, 프로젝트 파일, 실행 파일 등)을 제출할 경우 10% 까지 감점

#### 2. 감점기준

- a. 지각제출은 1 시간마다 10% 감점이 있습니다.
  - i. LMS 기준 00:00 ~ 00:59 -10%, 01:00 ~ 01:59 -20%, ..., 09:00 ~ 09:59 -100%
- b. Cheating (카피):
  - i. 교수/조교/튜터가 같이 보고 평가하여 다수결로 결정
  - ii. 동일한 코드가 나오면 "copied"/"being copied" 여부에 관계없이 해당 점수 0 점
  - iii. Internet 에서 찾거나 책에 있는 코드를 사용하였을 경우 제출한 코드 안에 주소나 책의 쪽수 등을 명시할 것. 자신이 참조한 코드의 정보를 명시하면 완전히 동일하더라도 cheating 으로 판단하지 않습니다. 즉, 많이 알려진 코드라면 출처를 명시하면 다른 학생들과 같아도 상관없습니다.
  - iv. 출처가 '친구', '선배' 등은 허용하지 않습니다.

#### 3. 제출물

- a. Submit only .c files, nothing else (필요없는 파일을 제출한 경우 10%까지 감점이 있습니다.)
- b. No VC++ project, no executable, no data (C 파일 이외에 다른 파일을 제출하면 감점)
- c. 소스 코드들은 하나로 묶어서 hw3.zip 파일로 제출함. 파일들을 따로 제출하면 채점이 매우 어렵습니다.
- d. 주어지는 template (\*\_template.c)은 '참고용' 입니다. 입출력만 적합하게 구현하면 본인의 취향에 따라 얼마든지 바꾸어도 괜찮습니다.

#### 4. Compilation:

- a. Your code will be compiled by gcc 5.4.0 under Ubuntu linux
- b. In Windows environment, code::blocks is recommended (<http://www.codeblocks.org/downloads>)
- c. Linux 와 code::blocks 는 같은 gcc compiler 를 쓰기 때문에 (이론적으로는) 결과가 같습니다. Compiler version 은 조금 다를 수 있습니다.

## 문제 설명

목적: 단어의 출현회수 세기. 공백으로 분할된 단어(word)들이 전체 파일에서 몇 번 있는지 수를 세고, 출현회수의 내림차순으로(큰 것에서 작은 것으로) 단어-출현회수를 저장한다.

입력: 텍스트 파일

출력: 텍스트 파일, 단어와 출현회수만 출력

Template file: 없음

Input files: input4a.txt, input4b.txt, input4c.txt

### input4a.txt (3,954 bytes)

```
lorem ipsum dolor sit amet consetetur sadipscing elitr sed diam nonumy eirmod tempor invidunt
ut labore et dolore magna aliquyam erat sed diam voluptua at vero eos et accusam et justo duo
dolores et ea rebum stet clita kasd gubergren no sea takimata sanctus est lorem ipsum dolor
sit amet ...
```

### input4b.txt (8,944 bytes)

```
i have a dream i am happy to join with you today in what will go down in history as the
greatest demonstration for freedom in the history of our nation five score years ago a great
american in whose symbolic shadow ...
```

### input4c.txt (2,244,618 bytes, 줄바꿈 있음)

```
the project gutenberg edition of the world factbook 1992
this file should be named world92 zip or world92 txt

corrected editions of our etexts get a new number world921 txt
versions based on separate sources get new letter world92a txt
...
```

Output files (실행하여 생성될 것들): out4a.txt, out4b.txt, out4c.txt

- 코드를 컴파일하여 실행하면 실행파일과 같은 경로에 있는 input4a.txt, input4b.txt, input4c.txt 를 읽어서 out4a.txt, out4b.txt, out4c.txt 를 생성한다.
- Sample 로 주어진 output 파일들(out4a-sample.txt, out4b-sample.txt, out4c-sample.txt)과 본인이 작성한 프로그램에 의해 생성된 출력이 같은지 확인한다.
- 파일이름은 고정, 위의 여섯 개의 파일은 프로그램 코드에 직접 집어 넣는다.

out4a.txt	out4b.txt	out4c.txt
et 40 dolor 22 lorem 20 ipsum 19 sit 19 amet 19 ...	the 103 of 99 to 59 and 54 a 38 we 33 ...	and 8047 of 7904 the 6632 1 5314 in 3061 rate 2755 km 2679 2 2623 to 2481 1992 2387 ...

## 요구조건(requirements)

- ① **Submission file (제출할 파일): countwords.c**
- ② 문제를 간단하게 하기 위하여 입력파일에는 숫자, 영어 소문자만 포함된다.
- ③ 단어는 공백/탭/줄바꿈으로 구분된다(즉, "fscanf(fp, "%s", ...)" 를 쓰면 된다. fgets, strtok 등 복잡한 명령은 쓸 필요 없음)
- ④ 출력은 "단어 출현회수" 를 매 단어마다 줄바꿈

(끝)