

🏠 / C++프로그래밍과실습 (CB3500572-062) / intermediate - 과제 031 - word counter

개요

코딩 결과

intermediate - 과제 031 - word counter

제출 마감일: 2023-03-27 23:59

업로드 가능한 파일 수: 3

제출 방식: 개인

주의

과제 030 은 keylog 분석 및 jplag 를 이용한 표절 검사가 상시 수행됩니다.

- 키로그 파일(p31.csv)에 소스 코드 이름이 Control.cpp, main.cpp 등으로 제출 파일명과 일치하지 않으면 로그 점수가 계산되지 않습니다.
- 키로그 파일(p31.csv)에 키보드 입력 정보가 없는 경우 로그 점수가 계산되지 않습니다. (로그가 IdeState 나 Action 타입 밖에 없는 경우가 종종 있습니다.)
- 계속 잘 안될 때에는 초심으로 돌아가 activitytracker 를 clear 하고, 소스 파일명을 제출 파일명과 일치 시킨 후 직접 키보드로 코딩하는 것을 추천 드립니다.

목적

C++ 의 제어 구조문 (if, for, while 등)을 연습해 봅니다.

설명

C++ 에서 표준 입력을 담당하는 cin 객체의 >> 연산자는 다음과 같은 입력이 들어올 경우 false 를 반환합니다.

- 유효하지 않은 입력
- End Of File (EOF)

<EOF 참고>

- Clion 에서는 먼저 Help > Find Action... > Registry > run.processes.with.pty 옵션을 해제합니다.
- 프로그램을 실행 후 입력 값을 모두 입력 후 Ctrl+Z 키나 Ctrl+D 키를 눌러서 EOF를 입력해 보세요.
- 프로그램이 종료되지 않으면 키보드로 EOF 를 입력 후 엔터키를 눌러보세요.

이를 이용해서 EOF 입력 될 때까지 문자열을 계속해서 입력 받을 수 있습니다.

```
string str;

while (cin >> str) {

    // 한 단어씩 str 에 들어옴

    // your code here

}
```

단점은 EOF 입력으로 while 문을 종료하더라도, std::cin 은 내부 상태가 입력을 계속 받을 수 없도록 변경됩니다.

우리는 특별한 종료문자(^)를 이용해 이 문제를 해결하고자 합니다.

```
while (cin >> str && str!="^") {

}
```

기사를 입력 받아 사용된 단어의 개수를 구하는 프로그램을 작성해 봅시다.

기사 입력 종료를 위해 마지막은 ^ 문자로 끝나는 것으로 정했습니다.

먼저, 사용할 자료 구조로 std::map 을 선택합니다.

키로는 단어를, 값으로는 개수를 저장합니다.

맵에 입력은 다음과 같이 할 수 있습니다.

```
std::map<std::string, size_t> words;

words.insert({문자열, 1});
```

동일한 단어가 여러 번 나오면 단어 개수를 증가 시킵니다.

```
auto search = words.find(str);
if (search != words.end())
    words[str]++;
else
    words.insert({str, 1});
```

std::map 의 [] 연산자의 공식 문서를 읽어 보면 간결하게 코딩할 수 있습니다.

https://en.cppreference.com/w/cpp/container/map/operator_at

```
++words[str];
```

기사에 사용된 단어들을 셴 후 다빈도 단어 Top 5 목록을 출력해 봅시다.

std::map 은 내부적으로 red black tree 로 구현되있 으면 기본적으로 key 가오 르 정렬되있 은 이습 니다. 따라서 우리는 std::sort 함수를 이용해서 value

`std::map` 은 내부적으로 `red-black tree` 로 구현되어 있으며, 기본적으로 `key` 값으로 정렬되어 있습니다. 따라서 우리는 `std::sort` 함수를 이용해서 `value` 로 `std::map`을 직접 정렬을 할 수 없습니다.

`value`로 `std::map`을 정렬하기 위해서는 정렬 가능한 `std::vector`로 `std::map`의 데이터를 복사 후 정렬하는 방법을 사용할 수 있습니다.

이때 단어들이 동일한 빈도수인 경우 사전 순서로 정렬합니다.

Top 5 Word Frequencies

1 the: 8

2 to: 8

3 charging: 5

4 in: 5

5 the: 5

다음으로 특정 단어가 몇 번 나타났는지 검색해 봅시다.

`std::map` 의 멤버 함수인 `find` 를 사용해 봅시다.

사용자가 존재하지 않는 단어도 검색할 수 있으므로 이를 처리해 줘야 합니다.

Find Word Frequencies

the

the: 8

to

to: 7

pnu: 0

문제

주어진 기사를 입력 받은 후 단어별 출현 횟수를 기록하고, 다빈도 단어 목록 및 검색어 입력 시 해당 단어의 출현 횟수를 출력하는 프로그램을 작성하시오.

<참고>

- 입력 받은 단어 중 다음 단어들(., : ;) 은 제외합니다.

```
std::string filtered {std::regex_replace(str, std::regex("[.,:;]"), "")};
```

- 단어의 대소문자를 구별하지 않습니다.

제출파일

control.cpp

p31.csv

입력

먼저 ^ 문자로 끝나는 영문 기사가 주어집니다.

다음으로 검색어가 주어집니다.

QUIT 검색어로 주어지면 프로그램을 종료합니다.

출력

첫 번째 줄에 단어 수를 출력합니다.

(예) #Words: 107

다음으로 다빈도 단어 Top 5 목록을 출력합니다.

[순서] 단어:빈도수

(예)

1 the: 8

2 to: 8

3 of: 7

마지막으로 검색어에 따른 단어의 출현횟수를 출력합니다.

존재하지 않는 단어의 경우에는 0 을 출력합니다.

검색어: 빈도수

(예)

he: 5

his: 6

pnu: 0

단, QUIT 을 입력 받으면 Bye! 문자열을 출력하고 프로그램을 종료 합니다.

입출력 예

입력	출력
An analysis of his filings shows Musk is not backing away from his holdings in Tesla, despite his promise to follow a poll he sent to his Twitter followers, who called on him to sell 10% of his stake. So far he's sold	

follow a poll he sent to his Twitter followers, who called on him to sell 10% of his stake. So far he's sold 10.1 million shares — about 7 million shares short of the goal.

That's because at the same time he is selling shares, he's also exercising options to buy additional stock. And he's doing so at a bargain exercise price of \$6.24 a share, well below 1% of Tesla (TSLA)'s current share price. Since Musk's Twitter poll on November 6, he has exercised options to buy 10.7 million shares of Tesla. To be clear, he would have done so with or without the poll — the options were due to expire by August of 2022 if he didn't exercise them.

And Tesla is poised to award Musk even more options, pending its upcoming financial results. His stake in the company is the reason Musk is the richest person on the planet. ^

the
to
of
he
his
PNU
QUIT

#Words: 107
Top 5 Word Frequencies
1 the: 8
2 to: 8
3 of: 7
4 his: 6
5 he: 5
Find Word Frequencies
the: 8
to: 8
of: 7
he: 5
his: 6
pnu: 0

Bye!

Hyundai Motor Group's inroad into the European electric vehicle market is expected to pick up speed after European ultra-fast charging network Ionity backed by the South Korean auto mogul raised \$791 million in a latest funding round to expand its electric vehicle charging network across Europe.

According to industry sources on Sunday, Ionity recently has successfully raised 700 million euros (\$791 million) in new funding from New York-based world's largest fund operator BlackRock Inc.

Ionity, founded in 2017 and run as a joint-venture between automakers like Hyundai Motor, Kia, BMW and Ford, currently operates around 1,500 ultra-fast charging stations on expressways in 24 European countries.

With the large new funding, Ionity is expected to speed up its plan to expand its ultra-fast charging network to 7,000, more than four times of the current amount, on expressways and other main roads in major European cities. The number of charging units at each station will also be added to an average 12, more than doubled from current six. ^

to
in
charging
Ionity
the
CSE
QUIT

#Words: 110
Top 5 Word Frequencies
1 to: 7
2 charging: 5
3 in: 5
4 the: 5
5 european: 4
Find Word Frequencies
to: 7
in: 5
charging: 5
ionity: 4
the: 5
cse: 0

Bye!

Samsung Electronics' ambitions to become a force in the foundry or contract-chip fabrication business are bearing fruit with a big deal and a purchase of super-expensive advanced semiconductor equipment.

Qualcomm, the San Diego-based smartphone chip developer, chose Samsung Electronics to be the sole manufacturer of its latest smartphone chipset, Snapdragon 8 Gen 1. The next-generation Snapdragon will be made on Samsung's 4-nanometer chips, CEO Cristiano Amon confirmed during a Qualcomm Snapdragon Tech Summit 2021 held in Hawaii last week.

The Snapdragon 8 Gen 1 is the successor to last year's Snapdragon 888, and the first chipset since Qualcomm changed its naming strategy, ditching the triple-digit number system. The chipsets will be used in devices by global smartphone companies including Xiaomi, Motorola, Vivo and Sony.

Many market watchers thought Qualcomm would split the order between Samsung and Taiwan's chip giant TSMC. ^

the
Snapdragon
Qualcomm
a

#Words: 99
Top 5 Word Frequencies
1 the: 10
2 snapdragon: 5
3 a: 4
4 and: 4
5 qualcomm: 4
Find Word Frequencies
the: 10
snapdragon: 5
qualcomm: 4
a: 4
and: 4
cpp: 0

Bye!

~
and
CPP
QUIT

Lailani Manimbo draws plenty of inspiration in her life from BTS — even when it comes to her fashion sense. Several times throughout the year, the 32 year-old neurosurgery RN from Torrance regularly hosts a cup sleeve event (with COVID-19 precautions in place), called Boba With Luv, named after the popular BTS song “Boy With Luv.” In K-pop culture, cup sleeve events are typically hosted at a boba tea house.

With a drink purchase, you receive a commemorative cup sleeve and other freebies, while socializing with other fans and shop vendors and showing off your BTS style. Manimbo’s latest event held at Almond Haus Cafe in Garden Grove celebrated the anniversary of the Permission to Dance single ahead of the group’s sold-out, four-night run at SoFi Stadium which concluded Thursday night. Tonight, the band concludes their L.A. stint with a performance during the iHeartRadio Jingle Ball tour next door at the Forum. ^

the
a
at
in
BTS
QUIT

#Words: 114
Top 5 Word Frequencies
1 the: 9
2 a: 5
3 with: 5
4 at: 4
5 in: 4
Find Word Frequencies
the: 9
a: 5
at: 4
in: 4
bts: 3

Bye!