

第 12 回 C++プログラミング実験Ⅱ

出題日 2021 年 6 月 25 日 (金) 15:00

第一提出期限 2021 年 6 月 25 日 (金) 16:40

第二提出期限 2021 年 7 月 1 日 (木) 23:59

実験課題 12

実行例のように、コマンドラインで指定したファイル名を読み、その中から、アルファベットと+の文字だけから構成される単語のみを抽出して、出現回数を数え、アルファベット順 (string の小さい順) に表示する以下のようなプログラムを作成せよ。

- (1) ファイル内の文章には、句読点や数字や記号が含まれるが、正規表現を使って、上の条件の単語のみを抽出すること。(今回の授業資料 トークンの取得のスライドを参考にせよ)
- (2) さらに抽出した単語の出現回数を調べるために、map を使って、単語と出現回数を pair にして、格納すること。(第 6 回講義資料の map を参考にすること)
- (3) ファイルを全て読み終わったら、単語と出現回数をアルファベット順に表示する。
- (4) また、ファイルの行数、単語の総数、出現した単語の種類数を表示する。

時間に余裕のある人は、以下も行うこと。

- (5) map は、第一引数 (単語) の順 (アルファベット順) に表示されるが、出現回数順にソートして、出現回数の多い上位 10 位までの単語を表示したい。これを実現するため、map のデータを、std::pair< 単語, 出現回数>を要素に持つ、vector にコピーする、次に、vector を STL の sort を使って、出現回数の多い順に並べ替え、上位 10 位までの単語を、出現回数とともに、表示する。

実行結果に示すように、長いファイル C++.txt と短いファイル test.dat が用意されているので、最初は test.dat を使って、プログラムの動作を確かめよ。最終的には、C++.txt で、実行結果と同様な結果が得られることを確認せよ。

```
% cat test.dat
This is a pen. This is a book.
That is a notebook;
Hello! How are you?
Fine Thank you.
```

実行例 入力ファイル名を省略した時は, C++.txt が使われる. (以下はいずれも指定した例)

```
% ./prac12 C++.txt
---- Read file : C++.txt
++ : 1
A : 1
ALGOL : 1
AT : 1
Ada : 1
After : 3
Annotated : 1
:
中略
:
which : 3
while : 1
who : 1
widely : 1
winner : 1
with : 13
work : 2
working : 1
year : 2
yet : 1
-----
Total line num : 48
Total word count : 803
map.size : 376
-----
Word Count Top 10
1 : C++ : 36
2 : the : 32
3 : and : 26
4 : was : 19
5 : in : 18
6 : language : 14
7 : of : 14
8 : a : 13
9 : for : 13
10 : C : 13
```

```
% ./prac12 test.dat
---- Read file : test.dat
Fine : 1
Hello : 1
How : 1
Thank : 1
That : 1
This : 2
a : 3
are : 1
book : 1
is : 3
notebook : 1
pen : 1
you : 2
-----
Total line num : 4
Total word count : 19
map.size : 13
-----
Word Count Top 10
1 : a : 3
2 : is : 3
3 : you : 2
4 : This : 2
5 : That : 1
6 : Thank : 1
7 : How : 1
8 : are : 1
9 : book : 1
10 : Hello : 1
```