

BA1 - Scanner

請注意不能用 `#include<regex>` 抓到 BA 0 分!!!

請注意不能用 `#include<regex>` 抓到 BA 0 分!!!

請注意不能用 `#include<regex>` 抓到 BA 0 分!!!

Description

在編譯器中，Scanner 是一個關鍵組件，負責將 input text 轉換為 Token，這些 Token 稍後將由 Parser 進行進一步的處理。實作一個簡單的 Scanner，以識別並輸出 Token。

Token 類型	Regular Expression
NUM	<code>[1-9] [0-9]* 0</code>
IDENTIFIER	<code>[A-Za-z] [A-Za-z0-9_]*</code>
SYMBOL	<code>[\+\-*\/*\(\)\{\}\<\>]</code>
KEYWORD	<code>"if" "else" "while"</code>

Input Format

1. 輸入是一段程式碼，其中包括 NUM、IDENTIFIER、SYMBOL、KEYWORD 和空白，並以 EOF 作為結束。
2. NUM 是非零開頭的正整數，或者零。
3. IDENTIFIER 是由字母（大小寫皆可）、數字和底線組成的，其中第一個字元一定是字母。
4. SYMBOL 為單一字元，包含 `+`、`-`、`*`、`/`、`=`、`(`、`)`、`{`、`}`、`<`、`>`和 `;`。
5. KEYWORD 包括 `if`、`else` 和 `while`。
6. 輸入文本可以包含任意數量的空格字符（空格、換行符等），這些空格字符應該被 Scanner 忽略。
7. 部分輸入的程式碼將會有不符合 Regular Expression 的情況發生。

Output Format

請在切割 input 後輸出其 Token 類型。

1. 輸出應該以純文字的形式呈現，每個 Token 佔據一行。
2. 每一行的格式為：類型 "內容"，並以一個空白做為區隔。「類型」是該 Token 的類型，「內容」是該 Token 的具體內容，例如：KEYWORD "if"。請參考下圖：

```
NUM "42"  
IDENTIFIER "number1"  
SYMBOL "+"  
KEYWORD "else"
```

3. 若 input 不符合 Regular Expression，則該行輸出只需印出 "Invalid"。請參考下圖：

```
NUM "42"  
IDENTIFIER "number1"  
SYMBOL "+"  
Invalid  
KEYWORD "else"
```

Sample input 1 x = 42;	Sample output 1 IDENTIFIER "x" SYMBOL "=" NUM "42" SYMBOL ";
Sample input 2 while (x <= 100) { x = x + 1; }	Sample output 2 KEYWORD "while" SYMBOL "(" IDENTIFIER "x" SYMBOL "<" SYMBOL "=" NUM "100" SYMBOL ")" SYMBOL "{" IDENTIFIER "x" SYMBOL "=" IDENTIFIER "x" SYMBOL "+" NUM "1" SYMBOL ";" SYMBOL "}"

<p>Sample input 3</p> <pre>if (number1 < 10){ number1 = number2 + 1; }</pre>	<p>Sample output 3</p> <pre>KEYWORD "if" SYMBOL "(" IDENTIFIER "number1" SYMBOL "<" NUM "10" SYMBOL ")" SYMBOL "{" IDENTIFIER "number1" SYMBOL "=" IDENTIFIER "number2" SYMBOL "+" NUM "1" SYMBOL ";" SYMBOL "}"</pre>
<p>Sample input 4</p> <pre>invalid = 123 \$;</pre>	<p>Sample output 4</p> <pre>IDENTIFIER "invalid" SYMBOL "=" NUM "123" Invalid SYMBOL ";"</pre>