**系統程式-作業一**

**Lexical Analysis實作**

**110學年度第一學期**

**老師: 賴建宏**

**學號: 10827216**

**班級: 資訊三乙**

**姓名: 王立翔**

壹、開發環境

作業系統：Window 10

程式語言：C++11

編譯器：MinGW

IDE：Code::Blocks 17.12

實作內容：**SIC** Lexical Analysis

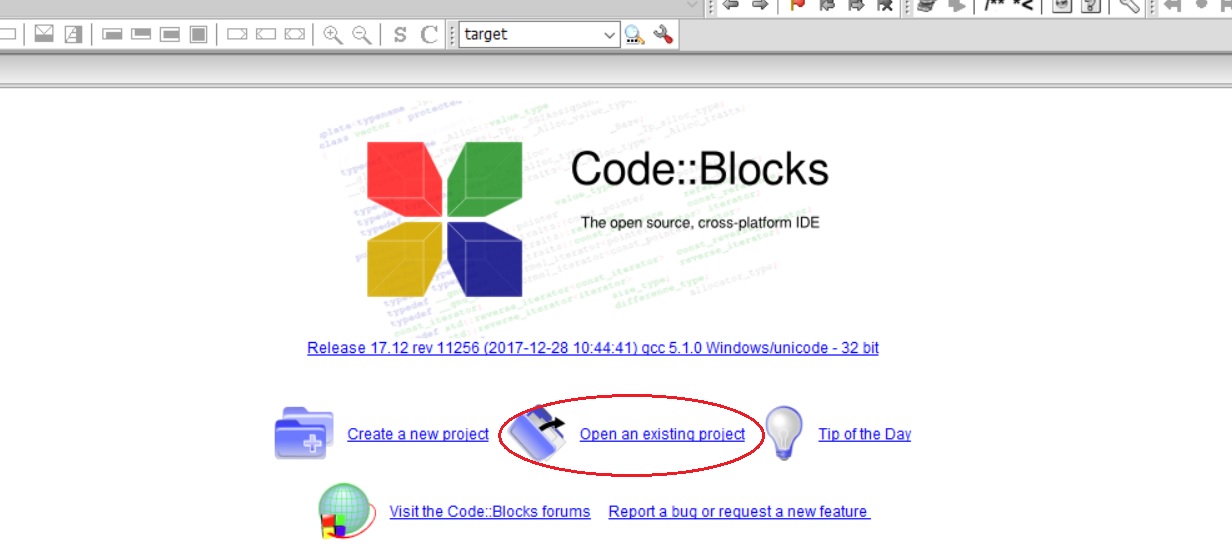
貳、執行流程

1.將壓縮檔內SicLexicalAnalysis資料夾解壓縮

一張含有 文字 的圖片

自動產生的描述

2.開啟Code Blocks，點選畫面正中央Open an existing project

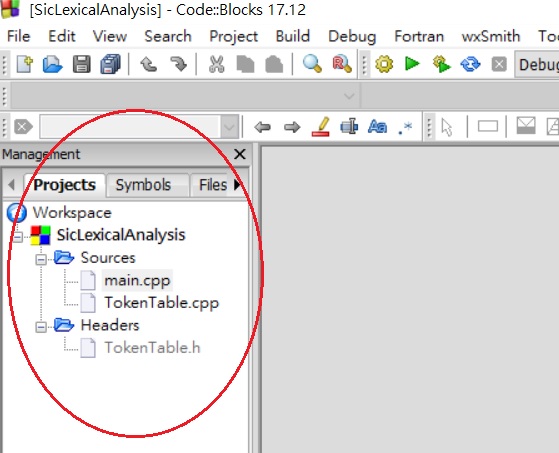


3.找到剛剛解壓縮的資料夾路徑，選取其內的SicLexicalAnalysis.cbp檔案，並按下開啟

一張含有 文字 的圖片

自動產生的描述

4.展開左方工作區即可看到project內.h/.cpp檔案



5.將欲讀取的input檔放進SicLexicalAnalysis資料夾內 (與main.cpp、SicLexicalAnalysis.cbp檔案同一路徑)

一張含有 文字 的圖片

自動產生的描述

6.回到Code Blocks，按下上方Build and run以執行程式

一張含有 文字 的圖片

自動產生的描述

7.於執行畫面輸入欲讀入檔案的檔名後，按下Enter

一張含有 文字 的圖片

自動產生的描述

8.畫面顯示分析完畢提示字後，輸入欲輸出檔案之檔名後，按下Enter

一張含有 文字 的圖片

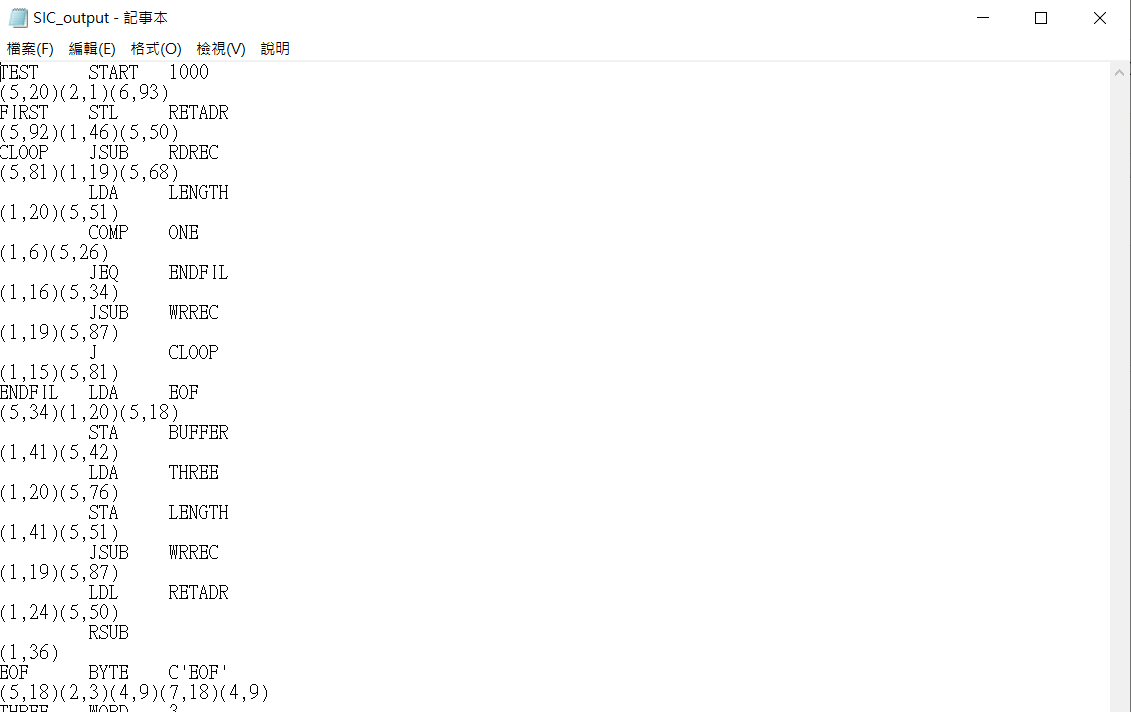
自動產生的描述

9.回到SicLexicalAnalysis資料夾，即可發現程式生成一個新的output檔案

一張含有 文字 的圖片

自動產生的描述

10.以記事本打開確認輸出結果



參、程式設計

一、功能

　　輸入：SIC組合語言程式碼

輸出：原始碼切完Token後的結果，輸出每個Token其所在Table編號以及其於Table內的位置

讀入並分析原始碼，將讀到字串切割成最小單位Token，並將Token分成七類，分別是指令(Instruction)、假指令(Pseudo)、暫存器(Register)、分隔符(Delimiter)、符號(Symbol)、數字(Integer/Real)、字串(String)，並分別以一個Table儲存同類Token，編號由以上類型順序編號1~7，其中編號1~4的Table為普通陣列容器，index由1開始存放，執行過程中不可變動；編號5~7的Table為Hash Table，容量限制為100，index由0開始，程式執行過程中可以插入。

二、流程

1.從原始碼讀取字元，加入至TKBUF，直到找到White-space、EOF或分隔符號。

2-1.若找到White-space或EOF，檢查TKBUF大小是否大於0，是則將TKBUF內容存入至Table，並將Token資訊記錄下來，存至LBUF；承上步，若找到字元是換行或EOF，則檢查LBUF大小是否大於0，是則輸出LBUF內容（一行原始碼的Tokens資訊）。

2-2.若找到quote字元，檢查其前個字元為X或C（在SIC中，X字元後面接quote表示十六進位數字、C字元後面接quote表示字串）；往後讀，並存入TKBUF，直到遇到下個quote字元，依照前面條件判斷TKBUF該存入編號6 (Integer/Real)或編號7 (String)的Table，並記錄Token資訊並存至LBUF。

2-3.若找到其他分隔符號，則檢查TKBUF大小是否大於0，是則將TKBUF內容存入至Table，紀錄Token資訊；再將找到的分隔符號比對Table，紀錄Token資訊存至LBUF。

3.若前個步驟找到的字元為分號（在SIC中，分號字元後為註解），則往後讀取字元，直到找到換行或EOF。

4.重複步驟1，直到前述過程中讀取到EOF，則程式結束。

取得Token資訊流程：

1.將TKBUF內容依序比對Table編號1~4，若匹配到相同字串，則回傳該Token所在Table編號以及該Token於該Table內的index，此兩個數字即為一個Token資訊。

2.若Table編號1~4內無匹配，則表示該Token為一個Symbol，於編號5的Table (Symbol)內嘗試找尋是否存在該Token，否則將TKBUF利用Hash Function計算Hash Value並存入該Hash Table，回傳編號5表示該Token種類以及該Token於Table內的index。

三、資料結構

1.Instruction、Pseudo、Register、Delimiter Table

2.Symbol、Integer/Real、String Table (Hash Table)

3.Token Buffer (TKBUF，讀取原始碼字元時，作為Token暫存區)

4.Line Buffer (LBUF，存放一行原始碼切割成Tokens的暫存區)

肆、未完成的功能

Lexical Analysis大致已完成。