

1111 compiler HW1: LR(1) parser

作業內容：

撰寫一個 LR(1) parser 的程式，使 Context-Free Grammar 轉換成 parsing table。並利用生成的 parsing table 判斷輸入的 Syntax 是否符合文法並列出過程。

Input：

1. 測資會放在一個資料夾“input”內，資料夾中會有三筆公開測資。

新增磁碟區 (D:) > 1111 compiler小考 > LR0 LR1 > LR(1) > input >				
名稱	修改日期	類型	大小	
1	2022/11/14 下午 06:20	檔案資料夾		
2	2022/11/14 下午 06:20	檔案資料夾		
3	2022/11/14 下午 06:20	檔案資料夾		

2. 每筆測資的資料夾會有兩個檔案
 - X_grammar.txt 是 Context-Free Grammar，一個檔案內僅會有一個 CFG。
 - X_testdata.txt 是判斷是否符合文法的所有輸入 Syntax，該檔案內會有多個 Syntax 要進行判斷。

新增磁碟區 (D:) > 1111 compiler小考 > LR0 LR1 > LR(1) > input > 1

名稱	修改日期	類型	大小
1_grammar.txt	2022/9/8 上午 02:12	文字文件	1 KB
1_testdata.txt	2022/9/8 上午 02:12	文字文件	1 KB

3. X_grammar.txt 中，第一行與第二行分別為 terminal 和 non_terminal，第三行開始為 Context-Free Grammar，Context-Free Grammar 可能會有或符號('|')。

```
1 terminal: a, b
2 nonterminal: S, A
3 S->AA
4 A->aA|b
```

4. 另有三組隱藏測資，格式與公開測資完全相同，皆為 input 資料夾，1/1_grammar，2/2_grammar...

程式碼輸出範例：

1. 將 CFG 轉換成 parsing table 的過程，並列出每一個 state 的内部如下圖(parsing table 可用 xls、txt...檔存，或是單純只顯示出來，只要能辨識就好):

```
["S'-.>.S", 'S->AA', 'A->aA', 'A->b']
////////// state //////////
0 : [[["S'-.>.S", ['$']], ['S->AA', ['$']], ['A->aA', ['a', 'b']], ['A->b', ['a', 'b']]]
1 : [[["S'-.>.S", ['$']], ['A->aA', ['$']], ['A->b', ['$']]]
2 : [[["S->A.A", ['$']], ['A->aA', ['$']], ['A->b', ['$']]]
3 : [[['A->a.A', ['a', 'b']], ['A->aA', ['a', 'b']], ['A->b', ['a', 'b']]]
4 : [[['A->b.', ['a', 'b']]]
5 : [[['S->AA.', ['$']]]
6 : [[['A->a.A', ['$']], ['A->aA', ['$']], ['A->b', ['$']]]
7 : [[['A->b.', ['$']]]
8 : [[['A->aA.', ['a', 'b']]]
9 : [[['A->aA.', ['$']]]
////////// parsing table //////////
+---+---+---+---+---+---+
| $ | a | b | S | A |
+---+---+---+---+---+---+
| 0 |   | s3 | s4 | 1 | 2 |
| 1 | Acc |   |   |   |   |
| 2 |   | s6 | s7 |   | 5 |
| 3 |   | s3 | s4 |   | 8 |
| 4 |   | r3 | r3 |   |   |
| 5 | r1 |   |   |   |   |
| 6 |   | s6 | s7 |   | 9 |
| 7 | r3 |   |   |   |   |
| 8 |   | r2 | r2 |   |   |
| 9 | r2 |   |   |   |   |
+---+---+---+---+---+---+
```

2. 判斷輸入的 Syntax 是否合法並判斷列出過程

```
parsing: aaaaaabb
result: Valid!
+---+---+---+---+---+---+
| stack | input | action |
+---+---+---+---+---+---+
| 0 | aaaaaabb$ | s3 |
| 0a3 | aaaaabb$ | s3 |
| 0a3a3 | aaaabb$ | s3 |
| 0a3a3a3 | aaabb$ | s3 |
| 0a3a3a3a3 | aabb$ | s3 |
| 0a3a3a3a3a3 | abb$ | s3 |
| 0a3a3a3a3a3a3 | bb$ | s4 |
| 0a3a3a3a3a3a3b4 | b$ | r3 |
| 0a3a3a3a3a3a3A8 | b$ | r2 |
| 0a3a3a3a3a3A8 | b$ | r2 |
| 0a3a3a3a3A8 | b$ | r2 |
| 0a3a3A8 | b$ | r2 |
| 0A2 | b$ | s7 |
| 0A2b7 | $ | r3 |
| 0A2A5 | $ | r1 |
| 0S1 | $ | Acc |
+---+---+---+---+---+---+

parsing: ab
result: Invalid!
+---+---+---+---+---+---+
| stack | input | action |
+---+---+---+---+---+---+
| 0 | ab$ | s3 |
| 0a3 | b$ | s4 |
| 0a3b4 | $ | X |
+---+---+---+---+---+---+
```

繳交檔案：

繳交檔案須含有：

1. 原始碼(即.cpp、.c、.java、.py 等檔案，請盡量不要傳送整個專案檔)。
2. readme.txt (opt.) 內部可寫程式碼資訊、編譯環境、套件需求...等。
3. **請勿**附上打包好的執行檔

最終繳交時間：

請將所有檔案壓縮成「學號_hw1.zip」或「學號_hw1.rar」並於 portal 公告截止日(12/13 晚間 11:59:59)之前上傳至 portal 作業區。

配分：

1. 本作業必須與助教現場 demo。
2. 公開測資全對給 40 分，有到場 demo 並能正常解釋程式碼者 40 分(得斟酌扣分)，現場測試隱藏測資全對 20 分(不會有特殊 case)。
3. 抄襲者不論 A 抄 B 或 B 抄 A 二者皆 0 分並依校規處理，抄襲網路上的程式碼與上述相同。
4. 作業遲交者(以 portal 時間戳記為憑)，本作業最終成績 * 60%。

Demo 注意事項：

1. 本作業必須與助教現場 demo (若因疫情或其他不可抗力因素再行討論)，未進行 demo 或未與助教連絡其他 demo 方式者，本次作業一律不給分。
2. 預約 demo 後若無法前來請提前至少兩個小時寄信給助教要求換時間，若未到也未事前要求助教更換時間視同未進行 demo (若有緊急狀況或不可抗力因素者除外)。
3. Demo 時，一律現場下載 portal 的檔案進行 demo，若 demo 時使用非 portal 的檔案或有新檔案一律視為遲交。
4. 本作業可用任意程式語言撰寫，但實驗室電腦僅有以下作業系統與以下程式語言之編譯環境。

作業系統: Windows 10、ubuntu 20.04、ubuntu 18.04.5 on Windows。

程式語言：

- C++ (ubuntu 使用 G++編譯，windows 使用 Visual Studio 2017 編譯)
- Python 3.8.2 (非 Anaconda 環境)
- Java 19

若需要其他環境或版本請自行攜帶筆電或提前告知並於 readme 寫上，以避免版本不同產生的錯誤。

如有上述未盡之事宜，將在 portal 上進行公告通知。

參考資料：

1. Textbook Chapter 4 (PPT ‘9 syntax analysis p5’)
2. <https://github.com/imlegend19/LR-0--Parsing>
3. <https://github.com/amirhossein-hkh/LR-Parser>
4. <https://github.com/kendalharland/LR0>
5. <https://github.com/AkashSharma93/LR-0--Parser>
6. https://github.com/linyacool/lexical_syntax_analysis
7. <https://github.com/LiuChangFreeman/C--Compiler>
8. <https://github.com/YanShulinjj/Sample-Compiler>

If you need **any assistance in English**, please contact TA.