

VÕ TIẾN

Thảo luận kiến thức CNTT trường BK về KHMT(CScience), KTMT(CEngineering)  
<https://www.facebook.com/groups/211867931379013>



Nguyên Lí Ngôn Ngữ Lập Trình (PPL)

---

PPL2 - HK242

Cuối Kỳ

---

Thảo luận kiến thức CNTT trường BK  
về KHMT(CScience), KTMT(CEngineering)  
<https://www.facebook.com/groups/211867931379013>

## Mục lục

1	Lý thuyết parser	2
2	Parser	3



# 1 Lý thuyết parser

Syntax bước này sẽ phân tích cú pháp vị trí của các *tokens* bắt ở phần trước lấy ra xử lý xem thứ tự trước sau có đúng hay không, chủ ngữ vị ngữ có vị trí phù hợp hay không

Để hiểu nhất trong quá trình sản xuất bánh mì thì các chế biến mới vào nặng bột sau đó hấp sau đó nữa là bán các bước này đầu vẩy nguyên liệu từ các bước trước đó.

- Cần *tokens* nào thì sẽ ưu cầu bước *lexer* lấy lên
- Sau đó kiểm tra vị trí từng *tokens*
- cuối cùng là sinh ra một cây AST từ các *tokens* trước đó có thể bỏ qua 1 số *tokens* không cần thiết
- Loại đầu *BNF* không cho phép xử dụng biểu thức chính quy
- Loại sau *EBNF* mở rộng của thặng trên nên cho xử dụng

Chú ý: vì phần này ảnh hưởng đến BTL2 nên các bạn không nên xử dụng các toán tử như \* và + vì tới phần sau code sẽ chậm đi khá hiểu nếu bạn không rành về lập trình hàm(function programming) ở môn LTNC vẫn sử dụng ? như bình thường

Một số loại hay dùng:

1. Loại một là các *Tokens ID* cách nhau bởi dấu *COMMA*, có thể rỗng

```
// cách 1
list_ID: list | ;
list: ID COMMA list | ID;
```

```
// cách 2
list_ID: list?;
list: ID COMMA list | ID;
```

2. Loại hai là các *Tokens ID* cách nhau bởi dấu *COMMA*, không thể rỗng

```
list_ID: ID COMMA list_ID | ID;
```

3. Loại ba là các *Tokens ID* không cách nhau bởi dấu gì, có thể rỗng

```
list_ID: ID list_ID | ;
```

4. Loại bốn là các *Tokens ID* không cách nhau bởi dấu gì, không thể rỗng

```
list_ID: ID list_ID | ID;
```

5. Loại năm phần *Expression* ví dụ bên dưới

6. loại sáu khai báo đối xứng số lượng id bên trái bằng với số lượng khởi tạo bên phải

```
variables: ID CM variables CM expression | ID all_types ASSIGNINIT expression;
```



## 2 Parser

1. Để liệt kê một tập hợp số nguyên, người ta sử dụng một danh sách các phần tử cách nhau bởi một dấu *CM*. Các phần tử đó có thể là một số nguyên *IL* hoặc một đoạn số nguyên. Đoạn số nguyên là hai số nguyên cách nhau bởi một dấu *TP*. Dạng EBNF của vế phải luật sinh tập hợp số nguyên kể trên là

- a)  $IL\ CM\ IL^* \mid (IL\ TP\ IL)^*$                       b)  $IL\ CM\ (IL\ TP\ IL)^* \mid (IL\ TP\ IL)\ CM\ IL^*$   
c)  $(IL \mid IL\ TP\ IL)\ (CM\ (IL \mid IL\ TP\ IL))^*$                       d)  $IL\ CM\ (IL \mid IL\ TP\ IL)^* \mid CM\ IL^+$

2. Hãy cho biết văn phạm nào sau đây bị nhập nhằng

- a)  $S \rightarrow \varepsilon \mid aSbS$                       b)  $S \rightarrow \varepsilon \mid aSa \mid bSb$   
c)  $S \rightarrow AB \mid BA, A \rightarrow \varepsilon \mid aA, B \rightarrow \varepsilon \mid bB$                       d)  $S \rightarrow \varepsilon \mid a \mid b \mid aSa \mid bSb$

3. Một danh sách *lcase* bao gồm các *case* nối tiếp nhau và có thể rỗng. Một *case* bao gồm từ khóa *CASE*, sau đó là một biểu thức *exp* và một dấu *CL* và kết thúc bởi danh sách các phát biểu *stmtlist*. Dạng EBNF của vế phải luật sinh *lcase* là

- a)  $CASE\ (exp\ CL\ stmtlist)^*?$                       b)  $(CASE\ exp^*\ CL\ stmtlist^*)?$   
c)  $(CASE\ exp\ CL\ stmtlist)^*$                       d)  $CASE\ exp^*\ CL\ stmtlist^*$

4. Cho một tập luật sinh như sau:

$S \rightarrow aSb \mid T$

$T \rightarrow cTd \mid \varepsilon$

Một quá trình dẫn xuất trái nhất là

- a)  $S \Rightarrow aSb \Rightarrow aaSbb \Rightarrow aacTdbb \Rightarrow aaccTddbb \Rightarrow aaccd dbb$   
b)  $S \Rightarrow aSb \Rightarrow aTb \Rightarrow acTdb \Rightarrow ac\varepsilon db$   
c) Các lựa chọn khác đều đúng  
d)  $S \Rightarrow aSb \Rightarrow aaSbb \Rightarrow aaaSbbb \Rightarrow aaaaSbbbb \Rightarrow aaaaTbbbb \Rightarrow aaaabbbb$

5. Vế phải luật sinh của *ui\_list* mô tả một danh sách gồm ít nhất một phần tử *UI* được viết dưới dạng BNF là

- a)  $UI\ ui\_list \mid$                       b)  $UI\ ui\_list$   
c)  $ui\_list \mid UI$                       d)  $UI\ ui\_list \mid UI$

6. Cho đoạn mã sau trong ANTLR

```
decl: ID decl_tail;
decl_tail: CM decl | CL ID CM;
```

Về phải nào sau đây phù hợp cho luật sinh *decl* để tương đương đoạn mã trên?

- a)  $(ID\ CM)^* (CL\ ID\ CM)?$                       b)  $(ID\ CM)^* CL\ ID\ CM$   
c)  $ID\ CM\ ID (CL\ ID)^* CM$                       d)  $ID (CM\ ID)^* CL\ ID\ CM$

7. Ký hiệu *list* mô tả một danh sách (có thể rỗng) các phần tử *a* được viết cách nhau bởi một dấu *C*. Văn phạm cho *list* được viết trên ANTLR như sau:

```
list: elist | (1);
elist: a C elist | (2);
```

(1) và (2) lần lượt là (biết rằng ký tự - đại diện cho rỗng)

- a)  $\_$  và  $a$                       b)  $C\ a$  và  $\_$                       c)  $\_$  và  $\_$                       d)  $\_$  và  $C$

8. Câu nào sau đây đúng về ngôn ngữ sinh ra bởi văn phạm phi ngữ cảnh sau?

$S \rightarrow AB, A \rightarrow aA \mid \varepsilon, B \rightarrow bB \mid \varepsilon$

- a) Ngôn ngữ này gồm tất cả các chuỗi trên ngôn ngữ với tập ký tự  $a, b$   
b) Ngôn ngữ này gồm tất cả các chuỗi dạng  $a^n b^h$ , trong đó  $m, n$  là số nguyên không âm.  
c) Ngôn ngữ này gồm chỉ chuỗi rỗng  
d) Ngôn ngữ này gồm tất cả các chuỗi dạng  $a^n b^h$ , trong đó  $n$  là số nguyên không âm.



9. Khi lập trình bằng Python, để đưa ký hiệu xuống hàng trực tiếp vào trong chuỗi bằng cách nhấn phím Enter thay vì dùng chuỗi escape `\n`, chuỗi phải được đặt trong:
- một cặp 3 dấu nháy kép
  - một cặp dấu nháy kép
  - một cặp dấu nháy đơn
  - phải dùng escape, không thể nhập trực tiếp
10. Để biểu diễn dạng lũy thừa  $a^b$  với  $a$  và  $b$  là các số nguyên dương trong Latex, người ta viết cơ số  $a$ , tiếp theo là dấu mũ, tiếp theo là phần cơ số  $b$ . Nếu cơ số  $b$  có từ hai chữ số trở lên phải được bao trong dấu ngoặc nhọn. Nếu cơ số  $b$  chỉ có một chữ số, việc sử dụng cặp ngoặc là tùy chọn. Trong ANTLR4, cho các fragment mô tả ký tự số là ***DIGIT***, dấu mũ là ***HAT***, cặp ngoặc nhọn mở và đóng lần lượt là ***LP*** và ***RP***. Phần mô tả nào sau đây thể hiện được dạng lũy thừa nói trên
- `DIGIT+ HAT (DIGIT | LP DIGIT DIGIT+ RP)`
  - `DIGIT+ HAT (LP (DIGIT | DIGIT+) RP)`
  - `DIGIT+ HAT DIGIT | LP DIGIT+ RP`
  - `DIGIT+ HAT (DIGIT | LP DIGIT+ RP)`
11. Trong ngôn ngữ X, lập trình viên có thể viết các hằng số nguyên không chính xác bằng cách thay thế các ký tự số chưa biết bằng dấu `#` với ký tự liền trước nó phải là ký tự số. Hằng số nguyên không chính xác phải được bắt đầu bằng ít nhất hai ký tự số. Ví dụ như `12 #345`, `123 #45#` là hợp lệ nhưng `123 ##45` là không hợp lệ. Trong ANTLR4, cho các fragment mô tả ký tự số là ***DIGIT***, dấu `#` là ***SHARP***. Phần mô tả nào sau đây thể hiện cho hằng số nguyên không chính xác trong ngôn ngữ X
- `DIGIT (DIGIT SHARP?)+`
  - `DIGIT (DIGIT SHARP?)*`
  - `DIGIT (DIGIT SHARP)+`
  - `DIGIT (DIGIT* SHARP)*`
12. Phát biểu đa gán cho phép gán nhiều tên ***ID*** (về trái, cách nhau bởi một dấu ***CM***) cho nhiều biểu thức ***exp*** (về phải, cách nhau bởi một dấu ***CM***) với về trái và về phải cách nhau bởi một dấu bằng EQ. Hãy dùng assignment để mô tả cho phát biểu đa gán và viết về phải của assignment sao cho số lượng tên và biểu thức phải bằng nhau và có ít nhất một tên và một biểu thức trong phát biểu đa gán:
- `(ID (CM ID)* EQ exp (CM exp)*`
  - `ID CM assignment CM exp | ID EQ exp`
  - `ID CM assignment CM exp | EQ`
  - `ID EQ exp CM assignment | ID EQ exp`
13. Cho đoạn mã sau trong ANTLR hãy chọn biểu thức chính quy đúng nhất
- ```
decl: M decl_tail | ;
decl_tail: N decl | PMN;
```
- `(MN)* MPMN`
  - `(MNM(PM)*N)?`
  - `(M(NM)*PMN)?`
  - `(MN)*(MPMN)?`
14. Cho đoạn mã sau trong ANTLR hãy chọn biểu thức chính quy đúng nhất
- ```
decl: M decl_tail | ;
decl_tail: N decl_tail | PMN;
```
- `(MN*PMN)?`
  - `(MNM(PM)*N)?`
  - `(M(NM)*PMN)?`
  - `(MN)*(MPMN)?`
15. Cho đoạn mã sau trong ANTLR hãy chọn biểu thức chính quy đúng nhất
- ```
decl: N decl_tail | N;
decl_tail: M decl_tail M | P;
```
- `N(M*PM*)` (với 2 \* có số lượng phần tử bằng nhau)
  - `(MNM(PM)*N)?`
  - `(M(NM)*PMN)?`
  - `(MN)*(MPMN)?`

Áp dụng mô tả sau cho các 3 câu tiếp theo:



Cho ngôn ngữ X được mô tả trong ANTLR4 như sau:

```
grammar X;
program: stat EOF | defi EOF;
stat: ID '=' expr ';' | expr ';';
defi: ID '(' ID (',' ID)* ')' '{' stat* '}' ;
expr: ID | INT | func | 'not' expr | expr 'and' expr | expr 'or' expr;
func: ID '(' expr (',' expr)* ')' ;
INT: [0-9]+ ;
ID: [a-zA-Z_][a-zA-Z_0-9]* ;
WS: [ \t\n\r\f] -> skip ;
```

16. Cho các đoạn mã sau:

- (a) x and y; f(1);
- (b) f(x, y) { a = 3 + foo; x and y; }
- (c) b = 1 and foo(1, 2);
- (d) f(x, y) { a = 3 or foo(x and y); }
- (e) f(x, y) { a = 3 or foo(); }

Số đoạn mã thỏa ngôn ngữ X là

- a) 1                                      b) 3                                      c) 4                                      d) 2
17. Với chuỗi nhập x and y; f(1);, bộ phân tích ngữ pháp theo mô tả trên sẽ tạo ra kết quả nào sau đây?
- a) Error on line 1 col 9: f                                      b) Error on line 1 col 11: 1
  - c) Error on line 1 col 7: ;                                      d) successful
18. Với chuỗi nhập f(x, y){a=foo(x and y) or b; }, cây phân tích cú pháp sinh ra bởi văn phạm trên có chiều cao là
- a) 10                                      b) 9                                      c) 8                                      d) 7
19. Cho một ngôn ngữ bao gồm các khai báo biến được mô tả trong ANTLR4 như sau:

```
program: decl EOF;
decl: INT ID (CM ID)* SM;
ID: [a-z]+; INT: 'int'; CM: ','; SM: ';';
WS: [ \n\r\f\t] -> skip;
```

Nếu một câu lệnh khai báo chứa các định danh trùng lặp (ví dụ: int x, y, x;), bộ phân tích cú pháp sẽ xử lý như thế nào?

- a) Nó sẽ phân tích thành công, xem xét các định danh trùng lặp là các biến riêng biệt.
  - b) Nó sẽ phát sinh lỗi cú pháp vì các định danh phải là duy nhất trong một câu lệnh khai báo.
  - c) Nó sẽ phát sinh lỗi cú pháp vì không có luật sinh phù hợp.
  - d) Nó sẽ phân tích thành công, lỗi ngữ nghĩa trong quá trình thực thi.
20. Câu nào sau đây đúng về ngôn ngữ sinh ra bởi văn phạm phi ngữ cảnh sau?

$$\begin{aligned} s &\rightarrow aCb \\ a &\rightarrow Aa \mid \varepsilon \\ b &\rightarrow Bb \mid B \end{aligned}$$

- a) Ngôn ngữ này gồm tất cả các chuỗi trên ngôn ngữ với tập ký tự {A, B, C}.
- b) Ngôn ngữ này gồm tất cả các chuỗi dạng  $A^mCB^n$ , trong đó  $m \geq 0, n \geq 1$ .
- c) Ngôn ngữ này gồm chỉ chuỗi rỗng.
- d) Ngôn ngữ này gồm tất cả các chuỗi dạng  $A^nCB^{n+1}$ , trong đó  $n$  là số nguyên không âm.



21. Một phát biểu có thể là phát biểu ghép hoặc phát biểu gán. Một phát biểu ghép được viết bởi

- Token LP rồi đến RP, hoặc
- Một danh sách không rỗng các phát biểu được đặt giữa LP và RP.

Gọi `stmt` và `assign` lần lượt mô tả cho một phát biểu và phát biểu gán, các ví dụ sau là các phát biểu hợp lệ: `assign`, `LP RP`, `LP assign LP RP RP ...`. Hãy chọn các vế phải phù hợp cho `stmt`?

- a) `LP RP | LP (assign | LP RP)+ RP | assign`
- b) `LP stmt* RP | assign`
- c) `LP RP | LP stmt RP | assign`
- d) `LP RP | LP stmt stmt RP | assign`

22. Cho `mlist` mô tả một danh sách (có thể rỗng) các biểu thức (`expr`) cách nhau bằng dấu `CM`. Văn phạm của `mlist` được viết như sau:

- a) `mlist: expr CM mlist | expr | ;`
- b) `mlist: (expr (CM expr)*)?;`
- c) `mlist: nmmlist | ;`
- d) `mlist: expr etail | ;`
- `nmmlist: expr CM nmmlist | expr ;`
- `etail: CM expr etail | ;`

Trong các cách viết văn phạm trên, có bao nhiêu cách viết ĐÚNG?

- a) 4
- b) 2
- c) 1
- d) 3

23. Cho các mô tả văn phạm sau viết trên ANTLR:

```
decl: M decl_tail | ;
decl_tail: N decl | P M N;
```

Hãy chọn vế phải phù hợp cho luật sinh `decl` để tương đương các mô tả trên?

- a)  $(M N)^* M P M N$
- b)  $(M N M (P M)^* N)?$
- c)  $(M (N M)^* P M N)?$
- d)  $(M N)^* (M P M N)?$

24. Cho các mô tả văn phạm sau viết trên ANTLR:

```
decl: M decl_tail | ;
decl_tail: N decl | P M N;
```

Áp dụng mô tả sau cho các 3 câu tiếp theo:

Cho ngôn ngữ X được mô tả trong ANTLR4 như sau:

```
grammar X;
```

```
ID: [a-zA-Z_][a-zA-Z0-9_]*;
ASSIGN: '=';
INT_LIT: [0-9]+;
WS: [ \n\r\f\t] -> skip;
CM: ',';
```

```
program: (list_assigni ';')+;
list_assigni: ID CM list_assigni CM INT_LIT | ID ASSIGN INT_LIT;
```

25. Cho các đoạn mã sau:

- (a) `a;`
- (b) `b, c = 1;`
- (c) `b, c = 1, 2;`
- (d) `b = c;`
- (e) `b, = ,c;`

Số đoạn mã thoả ngôn ngữ X là



a) 1

b) 3

c) 4

d) 2

26. Với chuỗi nhập  $b, c = 1, 3, 4;$ , bộ phân tích ngữ pháp theo mô tả trên sẽ tạo ra kết quả nào sau đây?

a) Error on line 1 col 0: b

b) Error on line 1 col 3: c

c) Error on line 1 col 11: ,

d) successful

27. Với chuỗi nhập  $a, e = 1, 3$ , cây phân tích cú pháp sinh ra bởi văn phạm trên có chiều cao là

a) 10

b) 9

c) 8

d) 7

28. Xét văn phạm phi ngữ cảnh sau:

$$S \rightarrow aSb \mid T$$

$$T \rightarrow cT \mid \varepsilon$$

Hãy chọn mô tả đúng nhất về ngôn ngữ sinh ra bởi văn phạm này:

a) Tất cả các chuỗi có dạng  $a^m c^n b^m$ , với  $m, n \geq 0$ .b) Tất cả các chuỗi có dạng  $a^m c^n b^m$ , với  $m \geq 0, n \geq 1$ .c) Tất cả các chuỗi có dạng  $a^m c^n b^m$ , với  $m, n \geq 1$ .d) Tất cả các chuỗi có dạng  $a^m b^m c^n$ , với  $m, n \geq 0$ .

29. Xét văn phạm phi ngữ cảnh sau:

$$S \rightarrow a(Sb)? \mid a$$

Hãy chọn mô tả đúng nhất về ngôn ngữ sinh ra bởi văn phạm này:

a) Tất cả các chuỗi có dạng  $a^n b^n$ , với  $n \geq 1$ .b) Tất cả các chuỗi có dạng  $a^n b$ , với  $n \geq 0$ .c) Tất cả các chuỗi có dạng  $a^n b^m$ , với  $n, m \geq 1$  và  $n \geq m$ .d) Tất cả các chuỗi có dạng  $a^n b^{n-1}$ , với  $n \geq 1$ 

30. Xét văn phạm phi ngữ cảnh sau:

$$S \rightarrow a(bS)? \mid a$$

Hãy chọn mô tả đúng nhất về ngôn ngữ sinh ra bởi văn phạm này:

a) Tất cả các chuỗi có dạng  $a^n b^n$ , với  $n \geq 1$ .b) Tất cả các chuỗi có dạng  $a^n b$ , với  $n \geq 0$ .c) Tất cả các chuỗi có dạng  $(ab)^n a|a$ , với  $n \geq 1$ d) Tất cả các chuỗi có dạng  $(ab)^n a|a$ , với  $n \geq 0$ 

31. Xét văn phạm phi ngữ cảnh sau:

$$S \rightarrow aSbS \mid \varepsilon$$

Hãy chọn mô tả đúng nhất về ngôn ngữ sinh ra bởi văn phạm này:

a) Tất cả các chuỗi có số lượng  $a$  và  $b$  bằng nhau, với các  $a$  xuất hiện trước  $b$ .b) Tất cả các chuỗi có số lượng  $a$  và  $b$  bằng nhau, nhưng thứ tự có thể bất kỳ.c) Tất cả các chuỗi có dạng  $a^n b^m$ , với  $n \geq m \geq 0$ .d) Tất cả các chuỗi có dạng  $a^n b^m$ , với  $m \geq n \geq 0$ .