TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**ĐỒ ÁN CUỐI KÌ MÔN NHẬP MÔN NGÔN NGỮ LẬP TRÌNH**

**Lexical analyzer and syntax analysis for PASC language**

*Người hướng dẫn*: **GV HÀ LÊ HOÀI TRUNG**

*Người thực hiện*: **NGUYỄN TRUNG TÍNH - 51603330**

**NGUYỄN TUÁN HUY - 51603144**

Lớp **: 16050302**

Khoá  **: 20**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2019**

TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**ĐỒ ÁN CUỐI KÌ MÔN NHẬP MÔN NGÔN NGỮ LẬP TRÌNH**

**Lexical analyzer and syntax analysis for PASC language**

*Người hướng dẫn*: **GV HÀ LÊ HOÀI TRUNG**

*Người thực hiện*: **NGUYỄN TRUNG TÍNH - 51603330**

**NGUYỄN TUÁN HUY - 51603144**

Lớp **: 16050302**

Khoá  **: 20**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2019**

LỜI CẢM ƠN

Đồ án này được hoàn thành dưới sự hỗ trợ tận tình của thầy Hà Lê Hoài Trung. Qua việc làm đề tài này chúng em đã học hỏi được những kinh nghiệm thầy truyền đạt lại và những kiến thức mới trong quá trình làm. Chúng em gửi lời cảm ơn sâu sắc tới thầy. Một lần nữa chúng em xin cảm ơn!

**ĐỒ ÁN ĐƯỢC HOÀN THÀNH**

**TẠI TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

Tôi xin cam đoan đây là sản phẩm đồ án của riêng chúng tôi và được sự hướng dẫn của Thầy Hà Lê Hoài Trung. Các nội dung nghiên cứu, kết quả trong đề tài này là trung thực và chưa công bố dưới bất kỳ hình thức nào trước đây. Những số liệu trong các bảng biểu phục vụ cho việc phân tích, nhận xét, đánh giá được chính tác giả thu thập từ các nguồn khác nhau có ghi rõ trong phần tài liệu tham khảo.

Ngoài ra, trong đồ án còn sử dụng một số nhận xét, đánh giá cũng như số liệu của các tác giả khác, cơ quan tổ chức khác đều có trích dẫn và chú thích nguồn gốc.

**Nếu phát hiện có bất kỳ sự gian lận nào tôi xin hoàn toàn chịu trách nhiệm về nội dung đồ án của mình.** Trường đại học Tôn Đức Thắng không liên quan đến những vi phạm tác quyền, bản quyền do tôi gây ra trong quá trình thực hiện (nếu có).

*TP. Hồ Chí Minh, ngày tháng năm*

*Tác giả*

*(ký tên và ghi rõ họ tên)*

*Nguyễn Trung Tính*

*Nguyễn Tuấn Huy*

PHẦN XÁC NHẬN VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN

**Phần xác nhận của GV hướng dẫn**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Tp. Hồ Chí Minh, ngày tháng năm

(kí và ghi họ tên)

**Phần đánh giá của GV chấm bài**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Tp. Hồ Chí Minh, ngày tháng năm

(kí và ghi họ tên)

TÓM TẮT

Trong đồ án, chúng tôi sẽ trình bày văn phạm phi ngữ cảnh và các sơ đồ trạng thái của những token. Có nghĩa là chúng tôi sẽ viết một công cụ phân tích ngữ nghĩa của một đoạn chương trình PASC, đưa ra những loại token đã được phát hiện cùng với đó là công cụ để sinh ra cây cú pháp, kiểm tra cú pháp của cây (Abstract Syntax Tree hay gọi tắt là AST). Và văn phạm phi ngữ cảnh sẽ giúp chúng ta hiểu hơn quá trình phân tích mà công cụ thực hiện mà mô hình hóa thành một cây cú pháp để biết được cấu trúc như thế nào.

MỤC LỤC

[LỜI CẢM ƠN i](#_Toc6263670)

[PHẦN XÁC NHẬN VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN iii](#_Toc6263671)

[TÓM TẮT iv](#_Toc6263672)

[MỤC LỤC 1](#_Toc6263673)

[DANH MỤC CÁC BẢNG BIỂU, HÌNH VẼ, ĐỒ THỊ 4](#_Toc6263674)

[CHƯƠNG 1 – VĂN PHẠM PHI NGỮ CẢNH VÀ LƯỢC ĐỒ TRẠNG THÁI 5](#_Toc6263675)

[1.1 Văn phạm phi ngữ cảnh 5](#_Toc6263676)

[1.2 Lược đồ trạng thái 8](#_Toc6263677)

[CHƯƠNG 2 – DEMO 11](#_Toc6263678)

[2.1 Phase 1 11](#_Toc6263679)

[2.1.1 TestCase-1: 11](#_Toc6263680)

[2.1.2 TestCase-2: 11](#_Toc6263682)

[2.1.3 TestCase-3: 11](#_Toc6263684)

[2.1.4 TestCase-4: 12](#_Toc6263686)

[2.1.5 TestCase-5: 12](#_Toc6263688)

[2.1.6 TestCase-6: 13](#_Toc6263690)

[2.2 Phase 2 13](#_Toc6263692)

[TÀI LIỆU THAM KHẢO 15](#_Toc6263693)

DANH MỤC CÁC BẢNG BIỂU, HÌNH VẼ, ĐỒ THỊ

**DANH MỤC HÌNH**

[Hình 1 Comparasion 6](#_Toc6263713)

[Hình 2 Identifier 6](#_Toc6263714)

[Hình 3 Number 7](#_Toc6263715)

[Hình 4 String 7](#_Toc6263716)

[Hình 5 Whitespace 8](#_Toc6263717)

[Hình 6 Symbol 8](#_Toc6263718)

[Hình 7: testcase 1 và output 9](#_Toc6263719)

[Hình 8: testcase 2 và output 9](#_Toc6263720)

[Hình 9: testcase 3 và output 9](#_Toc6263721)

[Hình 10: testcase 4 và output 10](#_Toc6263722)

[Hình 11: testcase 5 và output 10](#_Toc6263723)

[Hình 12: testcase 6 và output 11](#_Toc6263724)

[Hình 13: Code demo phase\_2 11](#_Toc6263725)

[Hình 14: Parse\_tree 12](#_Toc6263726)

CHƯƠNG 1 – VĂN PHẠM PHI NGỮ CẢNH VÀ LƯỢC ĐỒ TRẠNG THÁI

Sau đây để hiểu rõ hơn về cấu trúc để phân tích ngữ nghĩa và cú pháp của chương trình được viết bằng ngôn ngữ Pascal, chúng tôi sẽ trình bày 2 phần:

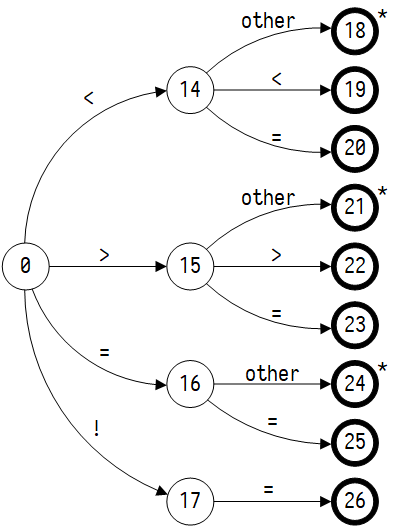
1.1 Văn phạm phi ngữ cảnh

Ký hiệu bắt đầu cho chương trình là chữ *program.*

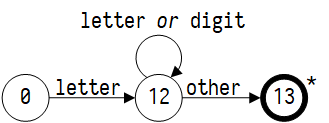
program: PROGRAMnumber expr SEMInumber declare block  
//Expresstion:  
expr : expr operator expr  
 | expr assign  
 | paren  
 | constvalue  
 | expr COMMAnumber expr  
 | declare  
 | ifStatement  
 | loopStatement  
 | identifier  
 | constnumber  
 | string  
 | forStatement  
 | repeat

//Block  
block::= BEGINnumber expr+ ENDnumber;  
  
//Loop-Statment  
loopStatement: := WHILEnumber condition expr+ #while  
 | LOOPnumber expr+ ENDLOOPnumber #loop ;  
  
//If-Statement  
ifStatement: := IFnumber condition THENnumber expr+ (ELSEnumber (block|expr+)) ENDIFnumber  
 |IFnumber condition THENnumber expr+ ENDIFnumber;  
  
//Condition  
condition: := expr compare=(EQnumber|LEnumber|GEnumber|NEnumber) expr  
 | condition ANDnumber condition  
 | condition ORnumber condition;  
  
//Operator  
operator: := (PLUSnumber|MINUSnumber|TIMESnumber|DIVnumber);  
  
//Assign  
assign: := COLEQnumber expr SEMInumber  
 | expr COLEQnumber constnumber;  
  
//Const-number  
constnumber: := ICONTSnumber;  
  
//const-value  
constvalue: := CONSTnumber expr SEMInumber;  
  
//Paren  
paren: := LPAREN expr RPAREN;  
  
//Identifier  
identifier: := IDnumber;  
  
//declare-value  
declare: := VARnumber expr SEMInumber;  
  
//String  
string: := '(\*' expr '\*)';  
  
//for  
forStatement: := FORnumber COLEQnumber expr TOnumber constnumber 'do' expr+  
 | FORnumber COLEQnumber expr DOWNTOnumber constnumber 'do' expr+;  
  
//repeat  
repeat: := REPEATnumber expr+ UNTILnumber condition SEMInumber;  
  
// Whitespace  
WS : := [ \t\n\r]+ -> channel(HIDDEN) ;  
  
// Keywords  
VARnumber : 'var' ;  
ANDnumber : 'and' ;  
ARRAYnumber : 'array' ;  
BEGINnumber : 'begin' ;  
CONSTnumber : 'const' ;  
DIVnumber : 'div' ;  
DOWNTOnumber : 'downto' ;  
ELSEnumber : 'else' ;  
ELSIFnumber : 'elsif' ;  
ENDnumber : 'end' ;  
ENDIFnumber : 'endif' ;  
ENDLOOPnumber : 'endloop' ;  
ENDRECnumber : 'endrec' ;  
EXITnumber : 'exit' ;  
FORnumber : 'for' ;  
FORWARDnumber : 'forward' ;  
FUNCTIONnumber : 'function' ;  
IFnumber : 'if' ;  
ISnumber : 'is' ;  
LOOPnumber : 'loop' ;  
NOTnumber : 'not' ;  
OFnumber : 'of' ;  
ORnumber : 'or' ;  
PROCEDUREnumber : 'procedure' ;  
PROGRAMnumber : 'program' ;  
RECORDnumber : 'record' ;  
REPEATnumber : 'repeat' ;  
RETURNnumber : 'return' ;  
THENnumber : 'then' ;  
TOnumber : 'to' ;  
TYPEnumber : 'type' ;  
UNTILnumber : 'util' ;  
WHILEnumber : 'while' ;  
// Literals  
ICONTSnumber : [0-9]+;  
// Operators  
PLUSnumber : '+' ;  
MINUSnumber : '-' ;  
TIMESnumber : '\*' ;  
COLEQnumber : ':=' ;  
LPARENnumber : '(' ;  
RPARENnumber : ')' ;  
DOTnumber : '.' ;  
SEMInumber : ';' ;  
COLONnumber : ':' ;  
COMMAnumber : ',' ;  
LTnumber : '<' ;  
GTnumber : '>' ;  
EQnumber : '==' ;  
LEnumber : '<=';  
GEnumber : '>=' ;  
NEnumber : '<>' ;  
// Identifiers  
IDnumber : := [A-Za-z0-9]+ ;

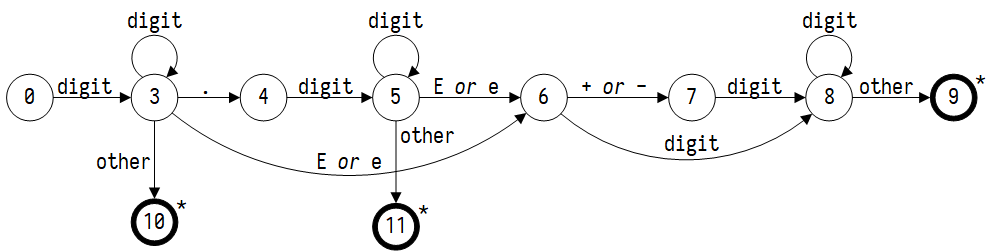
1.2 Lược đồ trạng thái



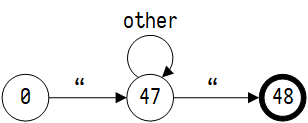
Hình 1 Comparasion



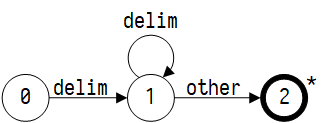
Hình 2 Identifier



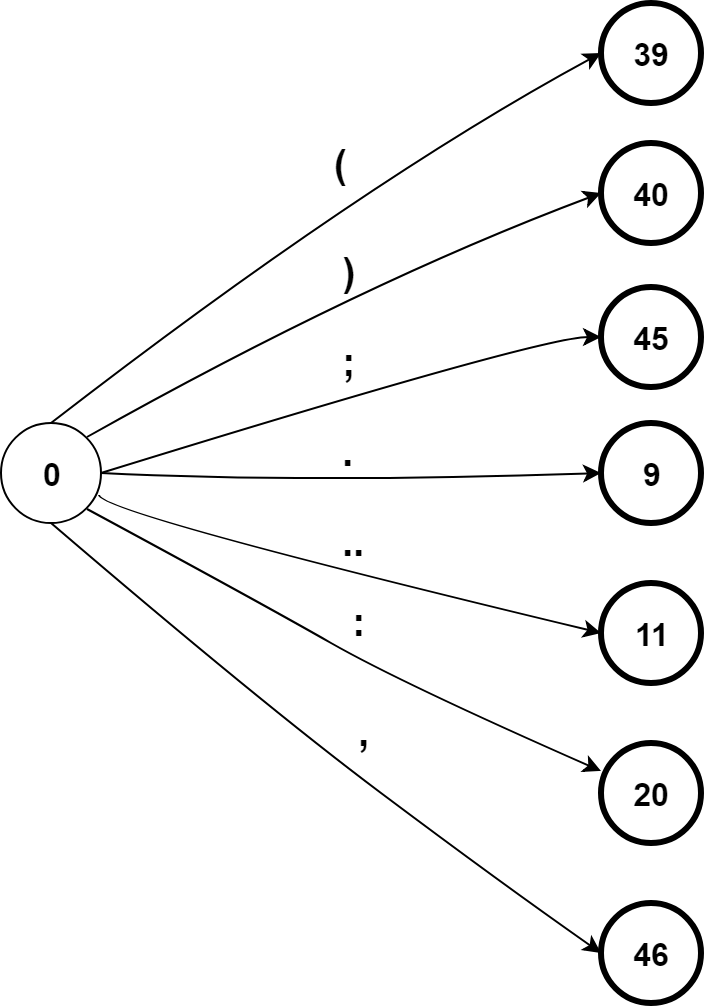
Hình 3 Number



Hình 4 String



Hình 5 Whitespace

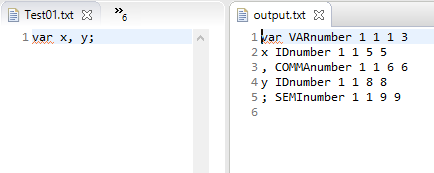


Hình 6 Symbol

CHƯƠNG 2 – DEMO

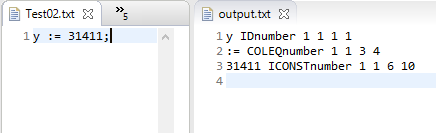
2.1 Phase 1

2.1.1 TestCase-1:



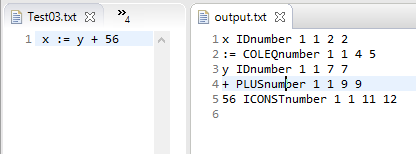
Hình 7: testcase 1 và output

2.1.2 TestCase-2:



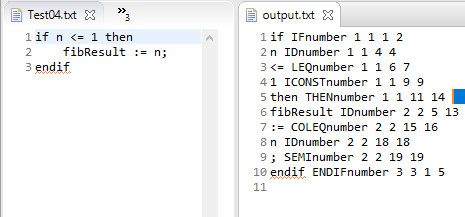
Hình 8: testcase 2 và output

2.1.3 TestCase-3:



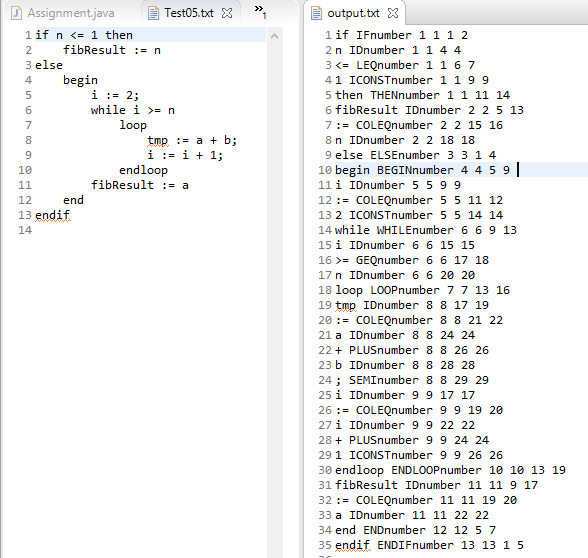
Hình 9: testcase 3 và output

2.1.4 TestCase-4:



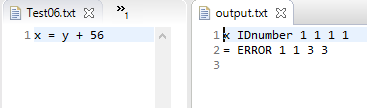
Hình 10: testcase 4 và output

2.1.5 TestCase-5:



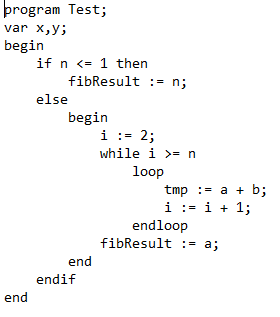
Hình 11: testcase 5 và output

2.1.6 TestCase-6:

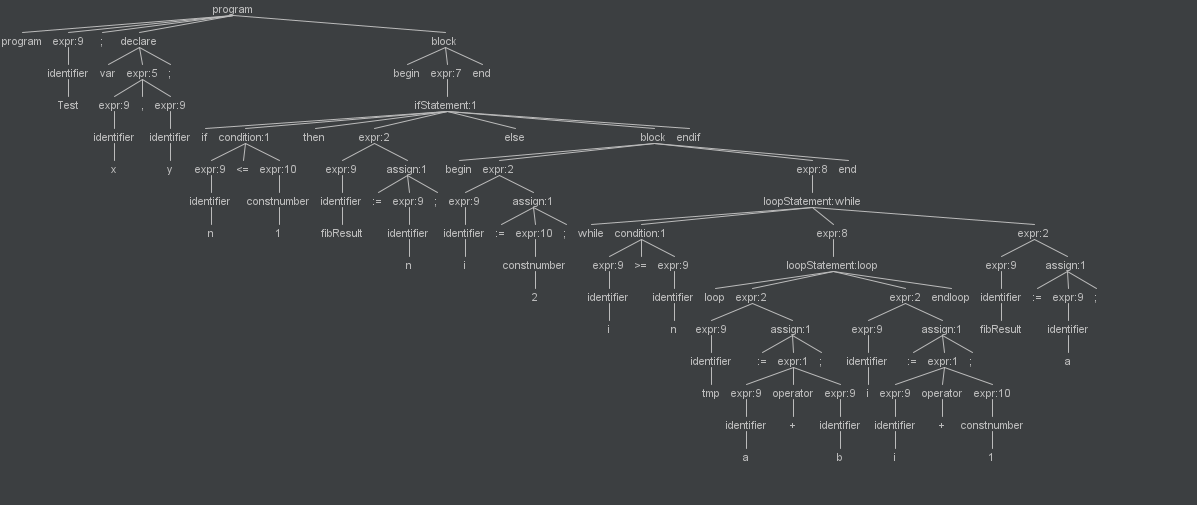


Hình 12: testcase 6 và output

2.2 Phase 2



Hình 13: Code demo phase\_2



Hình 14: Parse\_tree

TÀI LIỆU THAM KHẢO

1. <https://dzone.com/articles/building-a-compiler-for-your-own-language-from-the-1>
2. <https://tomassetti.me/parse-tree-abstract-syntax-tree/>