Syntax Analysis

Symbol Table

A Symbol Table is a container for maintaining information extracted from a source file. Using a Symbol Table is a simpler alternative to using a Syntax Tree. At the most basic level a Symbol Table is just an associative map (key → value). Beyond a simple associative map a Symbol Table organizes information about defined constructs in a programming language and provides search mechanizes relevant to constructing a Compiler.

The Symbol Table for your Compiler will be populated with data during Syntax Analysis. You are not complete with Syntax Analysis until you have populated your Symbol Table.

See the Symbol Table Examples document.

Syntax Analysis

During Syntax Analysis your Compiler receives a sequence of Tokens from the Lexical Analyzer (Scanner) and verifies that each of the Tokens can be generated by the Grammar for the Source Language.

Various methods exist for parsing a Source file we will focus on one very generalized method: Recursive-Descent Parsing

Recursive-Descent Parsing

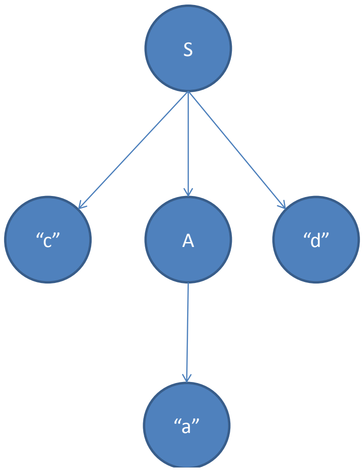
Recursive-Descent Parsing is a Top-Down parsing technique that may involve backtracking.

Top-Down Parsing produces a Parse Tree from the Top Down, meaning from the Root to Leaves.

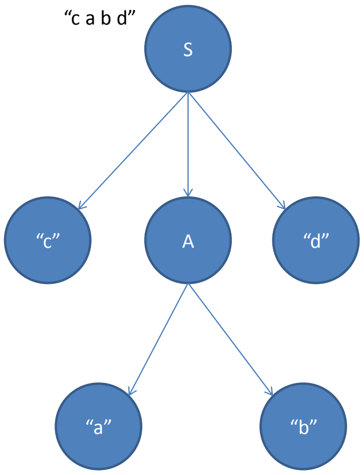
Backtracking is a search strategy used in searching Trees that involves returning to a previously visited node in a Tree to search an alternative path (child) that originates at the nodes.

S ::= “c” A “d”
A ::= “a” “b” | “a”

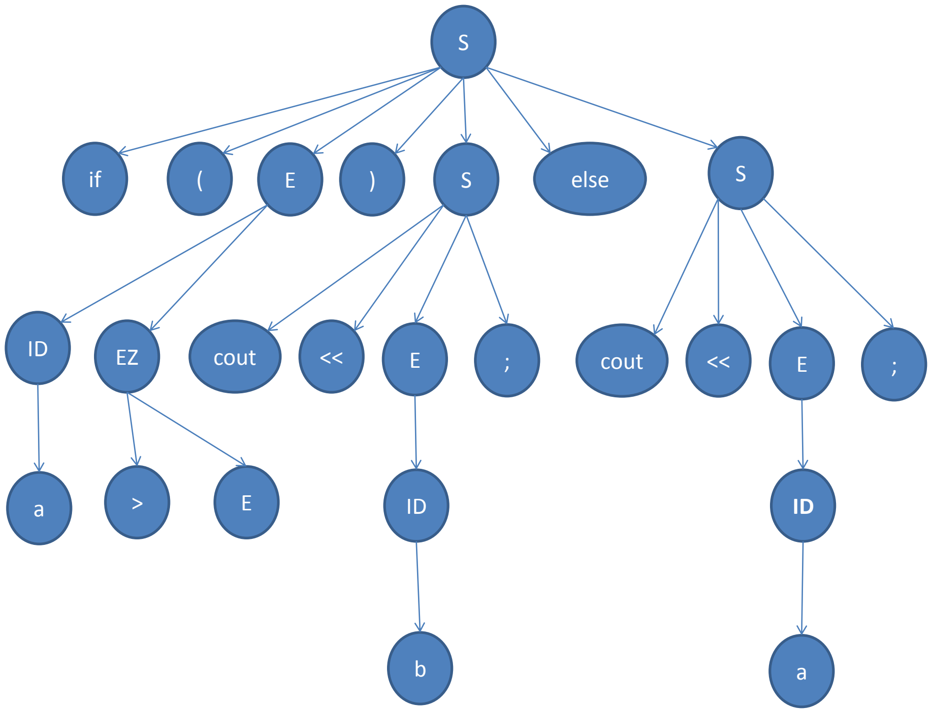
“c a d”



“c a b d”



Non-terminals are Uppercase
E ::= ID [EZ]
EZ ::= > E | < E
S ::= if (E) S [else S] | cout <<
Input: if (a > b) cout << b ; else cout << a;



The Six Forms for Recursive Descent Parsing

Form 1: Sequence of terminals

NT → a b c

```
Void NT() {
     if (scanner.type() != Token.a.type)
          genError();
     scanner.next();
     if (scanner.type() != Token.b.type)
          genError();
     scanner.next();
     if (scanner.type() != Token.c.type)
          genError();
     scanner.next();
}
```

```
Matching on lexemes vs. Token type
if (scanner.lexeme() != Token.a.lexeme)
     genError();
```

Form 2: Sequence of Non-terminals

NT → A B C

```
Void NT() {
     A();
     B();
     C();
}
```

Form 3: Mix of Terminals & Non-terminals

NT → a X Y b Z

```
Void NT() {
     if(scanner.type() != Token.a.type)
          genError();
     scanner.next();
     X();
     Y();
     if(scanner.type() != Token.b.type)
          genError();
     scanner.next();
     Z();
}
```

Form 4: Something is Optional

NT → X [a Y] Z

```
Void NT() {
      X();

      if(scanner.type() == Token.a.type) {
            scanner.next();
            Y();
      }
      Z();
}
```

For this to work the first non-terminal of Z must not be terminal “a”.

If it is NOT then you must look ahead by more than one Token:

```
      scanner.peek().lexeme();
      scanner.peek().type();
```

Form 5: There is a choice

Case 5.1 NT → a X | b Y | c Z

```
Void NT() {
      switch(scanner.type()) {
      case Token.a.type:
            scanner.next();
            X();
            break;
      case Token.b.type:
            scanner.next();
            Y();
            break;
      case Token.c.type:
            scanner.next();
            Z();
            break;
      default:
            genError();
      }
}
```

Case 5.2 NT → X | Y | Z

We must know the first terminal symbol of each Non-terminal X, Y and Z to encode this
Otherwise we must support backtracking.

```
Void NT() {
     switch(scanner.type()) {
     case Token.X.first.type:
          // Don't proceed to scanner.next();
          X();
          break;
     case Token.Y.first.type:
          // Don't proceed to scanner.next();
          Y();
          break;
     case Token.Z.first.type:
          // Don't proceed to scanner.next();
          Z();
          break;
     default:
          genError();
     }
}
```

Form 6: Zero or more Occurrences

NT → { a N }

```
Void NT() {
     while(scanner.type == Token.a.type) {
          scanner.next();
          N();
     }
}
```

Partial(incomplete) example from Kxi expression

Hint: Work from the bottom up starting with the statement "`x = y;`"

```
expression::=
      "(" expression ")" [ expressionz ]
     | "true" [ expressionz ]
     | "false" [ expressionz ]
     | "null" [ expressionz ]
     | numeric_literal [ expressionz ]
     | character_literal [ expressionz ]
     | identifier [ fn_arr_member ] [ member_refz ] [ expressionz ]
     ;
```

Don't copy my code; understand it then write your own

```
void Expression() {
     if (scanner.lexeme() == "(" ) {
          scanner.next();
          Expression();
          if (scanner.lexeme() == ")" ) {
               scanner.next();
               if(scanner.type() == Token.isAexpressionZ ) {
                    ExpressionZ();
               }
          }
          else syntaxError(scanner.lexeme(), ")" );
     }
     else if (scanner.lexeme() == "true" ) {
               if(scanner.type() == Token.isAexpressionZ ) {
                    ExpressionZ();
               }
     }
     else if (scanner.lexeme() == "false" ) {
               if(scanner.type() == Token.isAexpressionZ ) {
                    ExpressionZ();
               }
     } else if (scanner.lexeme() == "null" ) {
               if(scanner.type() == Token.isaExpressionZ ) {
                    ExpressionZ();
               }
     } else if (scanner.type() == Token.isaNumbericLiteral ) {
               if(scanner.type() == Token.isaExpressionZ ) {
                    ExpressionZ();
               }
     } else if (scanner.type() == Token.isaCharLiteral ) {
               if(scanner.type() == Token.isaExpressionZ ) {
                    ExpressionZ();
               }
     } else if (scanner.type() == Token.isaIdentifier ) {
          // This is not complete
     } else syntaxError(scanner.lememe(), Token.isaExpression );
}
```