Compiler Design
编 译 器 构 造 实 验

Lab 7： Project-2

张献伟

xianweiz.github.io

DCS292, 3/31/2022

Project 2: What?

- 文档描述：https://github.com/arcsysu/SYsU-lang/tree/main/parser

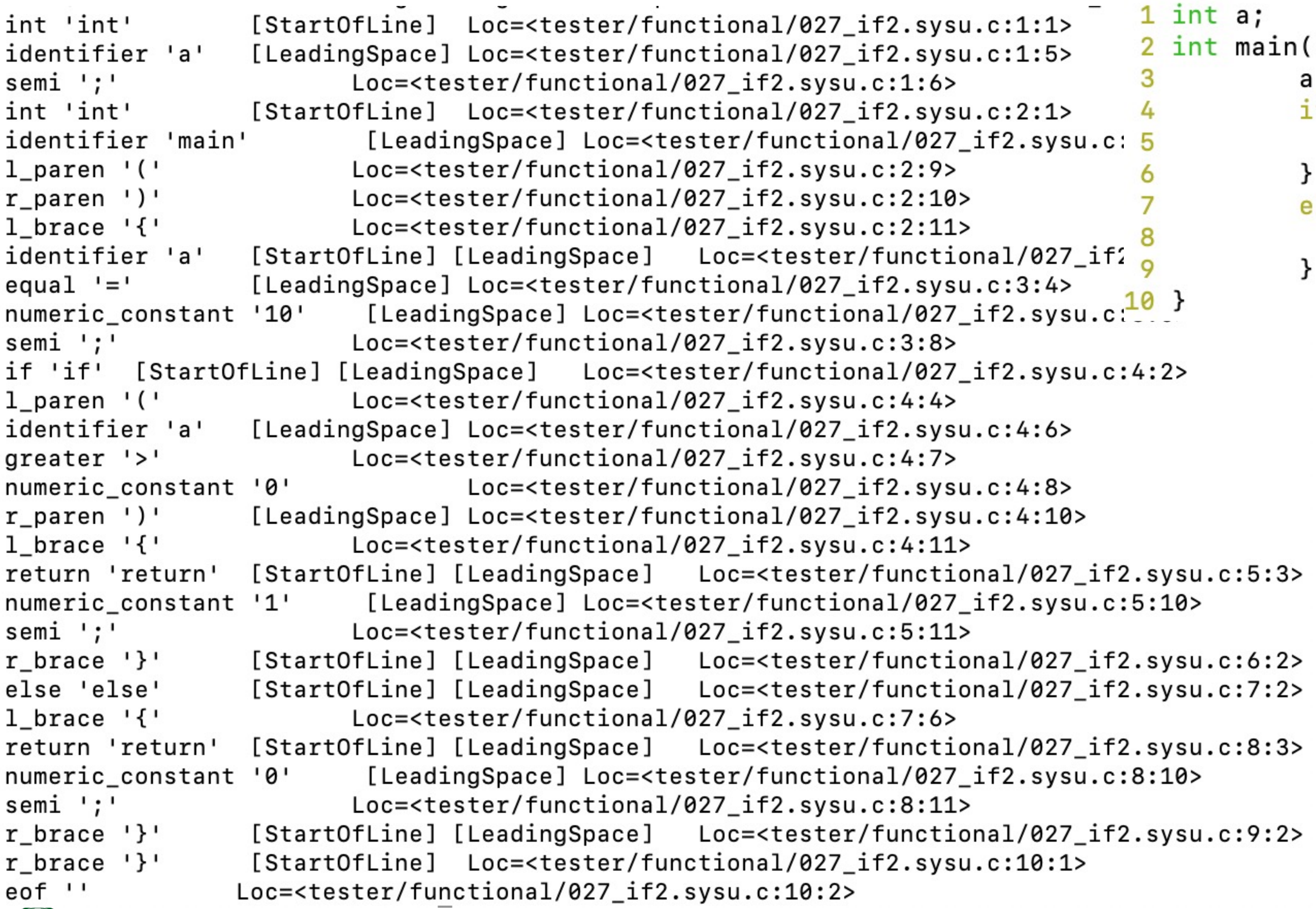
- 基于YACC/Bison实现一个语法分析器
  - 输入：token序列（由Project 1或Clang提供）
  - 输出：语法树（类似Clang AST）
- 总体流程
  - 引入Project1的lexer.l（可能需要简单修改）
  - 理解SYsU语言语法，构建上下文无关文法（CFG）规则
  - 使用YACC/Bison表示CFG文法
  - 提供语义动作，逐步构建分析树
- 截止时间
  - 4/28/2022

Project 2: How?

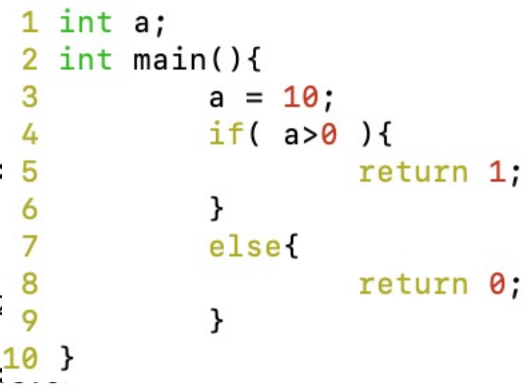
- 实现
  - $vim parser/parser.y
- 编译
  - $cmake --build ~/sysu/build -t install
    - 输出：~/sysu/build/parser
- 运行
  - $(export PATH=~/sysu/bin:$PATH \
    CPATH=~/sysu/include:$CPATH \
    LD_LIBRARY_PATH=~/sysu/lib:$LD_LIBRARY_PATH && sysu-preprocessor tester/functional/000_main.sysu.c | <THE_LEXER>| sysu-parser)
    - Clang提供token：<THE_LEXER> = clang -cc1 -dump-tokens 2>&1
    - Project1提供token： <THE_LEXER> = sysu-lexer

Clang Tokens

- $clang -cc1 -dump-tokens tester/functional/027_if2.sysu.c



```
int 'int'        [StartOfLine]  Loc=<tester/functional/027_if2.sysu.c:1:1>
identifier 'a'   [LeadingSpace] Loc=<tester/functional/027_if2.sysu.c:1:5>
semi ';'                 Loc=<tester/functional/027_if2.sysu.c:1:6>
int 'int'        [StartOfLine]  Loc=<tester/functional/027_if2.sysu.c:2:1>
identifier 'main'        [LeadingSpace] Loc=<tester/functional/027_if2.sysu.c:
l_paren '('              Loc=<tester/functional/027_if2.sysu.c:2:9>
r_paren ')'              Loc=<tester/functional/027_if2.sysu.c:2:10>
l_brace '{'              Loc=<tester/functional/027_if2.sysu.c:2:11>
identifier 'a'   [StartOfLine] [LeadingSpace]   Loc=<tester/functional/027_if2
equal '='        [LeadingSpace] Loc=<tester/functional/027_if2.sysu.c:3:4>
numeric_constant '10'    [LeadingSpace] Loc=<tester/functional/027_if2.sysu.c:
semi ';'                 Loc=<tester/functional/027_if2.sysu.c:3:8>
if 'if'  [StartOfLine] [LeadingSpace]   Loc=<tester/functional/027_if2.sysu.c:4:2>
l_paren '('              Loc=<tester/functional/027_if2.sysu.c:4:4>
identifier 'a'   [LeadingSpace] Loc=<tester/functional/027_if2.sysu.c:4:6>
greater '>'              Loc=<tester/functional/027_if2.sysu.c:4:7>
numeric_constant '0'             Loc=<tester/functional/027_if2.sysu.c:4:8>
r_paren ')'      [LeadingSpace] Loc=<tester/functional/027_if2.sysu.c:4:10>
l_brace '{'              Loc=<tester/functional/027_if2.sysu.c:4:11>
return 'return'  [StartOfLine] [LeadingSpace]   Loc=<tester/functional/027_if2.sysu.c:5:3>
numeric_constant '1'     [LeadingSpace] Loc=<tester/functional/027_if2.sysu.c:5:10>
semi ';'                 Loc=<tester/functional/027_if2.sysu.c:5:11>
r_brace '}'      [StartOfLine] [LeadingSpace]   Loc=<tester/functional/027_if2.sysu.c:6:2>
else 'else'      [StartOfLine] [LeadingSpace]   Loc=<tester/functional/027_if2.sysu.c:7:2>
l_brace '{'              Loc=<tester/functional/027_if2.sysu.c:7:6>
return 'return'  [StartOfLine] [LeadingSpace]   Loc=<tester/functional/027_if2.sysu.c:8:3>
numeric_constant '0'     [LeadingSpace] Loc=<tester/functional/027_if2.sysu.c:8:10>
semi ';'                 Loc=<tester/functional/027_if2.sysu.c:8:11>
r_brace '}'      [StartOfLine] [LeadingSpace]   Loc=<tester/functional/027_if2.sysu.c:9:2>
r_brace '}'      [StartOfLine]  Loc=<tester/functional/027_if2.sysu.c:10:1>
eof ''          Loc=<tester/functional/027_if2.sysu.c:10:2>
```



```
int a;
int main(){
        a = 10;
        if( a>0 ){
                return 1;
        }
        else{
                return 0;
        }
}
```

Clang AST

- $clang -Xclang -ast-dump -fsyntax-only tester/functional/027_if2.sysu.c

```c
 1 int a;
 2 int main(){
 3         a = 10;
 4         if( a>0 ){
 5                 return 1;
 6         }
 7         else{
 8                 return 0;
 9         }
10 }
```

The toplevel declaration in a translation unit is always the translation unit declaration

a variable declaration or definition

a function declaration or definition

```
TranslationUnitDecl 0x1d2654a8 <<invalid sloc>> <invalid sloc>
            ... cutting out internal declarations of clang ...
|-VarDecl 0x307fff10 <tester/functional/027_if2.sysu.c:1:1, col:5> col:5 used a 'int'
`-FunctionDecl 0x30800018 <line:2:1, line:10:1> line:2:5 main 'int ()'
  `-CompoundStmt 0x30800248 <col:11, line:10:1>
    |-BinaryOperator 0x308000f8 <line:3:2, col:6> 'int' '='
    | |-DeclRefExpr 0x308000b8 <col:2> 'int' lvalue Var 0x307fff10 'a' 'int'
    | `-IntegerLiteral 0x308000d8 <col:6> 'int' 10
    `-IfStmt 0x30800220 <line:4:2, line:9:2> has_else
      |-BinaryOperator 0x30800170 <line:4:6, col:8> 'int' '>'
      | |-ImplicitCastExpr 0x30800158 <col:6> 'int' <LValueToRValue>
      | | `-DeclRefExpr 0x30800118 <col:6> 'int' lvalue Var 0x307fff10 'a' 'int'
      | `-IntegerLiteral 0x30800138 <col:8> 'int' 0
      |-CompoundStmt 0x308001c0 <col:11, line:6:2>
      | `-ReturnStmt 0x308001b0 <line:5:3, col:10>
      |   `-IntegerLiteral 0x30800190 <col:10> 'int' 1
      `-CompoundStmt 0x30800208 <line:7:6, line:9:2>
        `-ReturnStmt 0x308001f8 <line:8:3, col:10>
          `-IntegerLiteral 0x308001d8 <col:10> 'int' 0
```

{}

a = 10

if-else

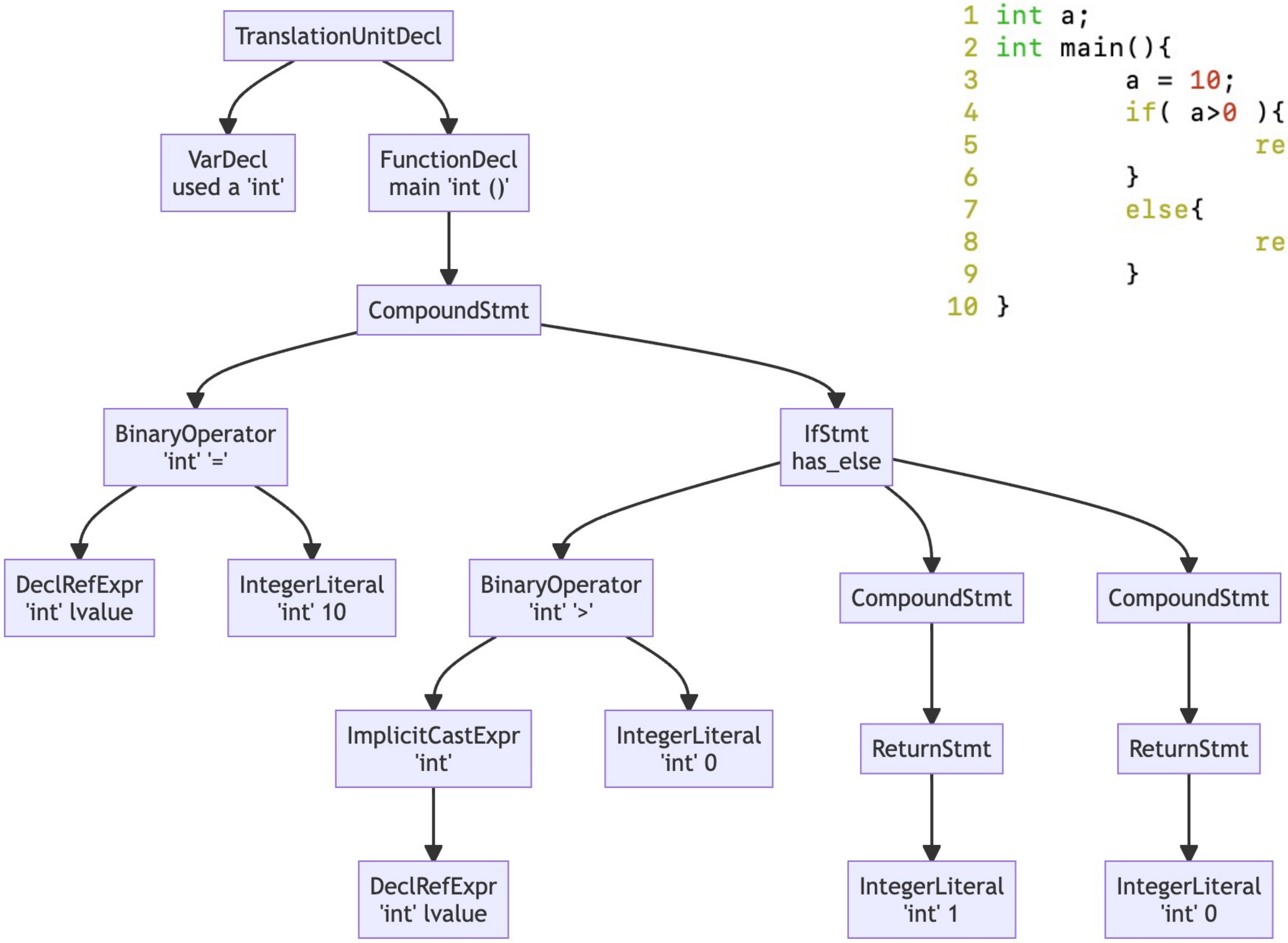
a>0

return 1

return 0

https://clang.llvm.org/docs/IntroductionToTheClangAST.html

Clang AST (cont.)



```c
int a;
int main(){
        a = 10;
        if( a>0 ){
                return 1;
        }
        else{
                return 0;
        }
}
```

Example

- $clang -Xclang -ast-dump -fsyntax-only tester/functional/000_main.sysu.c

```
1 int main(){
2     return 3;
3 }
```

```
TranslationUnitDecl 0x460b4a8 <<invalid sloc>> <invalid sloc>
                 ... cutting out internal declarations of clang ...
`-FunctionDecl 0x46aaf58 <tester/functional/000_main.sysu.c:1:1, line:3:1> line:1:5 main 'int ()'
  `-CompoundStmt 0x46ab070 <col:11, line:3:1>
    `-ReturnStmt 0x46ab060 <line:2:5, col:12>
      `-IntegerLiteral 0x46ab040 <col:12> 'int' 3
```

```
1 int a;
2 int main(){
3     a = 10;
4         if( a>0 ){
5                 return 1;
6         }
7         else{
8                 return 0;
9         }
10    return 3;
11 }
```

```
TranslationUnitDecl 0x1ab2b798 <<invalid sloc>> <invalid sloc>
                 ... cutting out internal declarations of clang ...
|-VarDecl 0x1abcb4b0 <tester/functional/000_main.sysu.c:1:1, col:5> col:5 used a 'int'
`-FunctionDecl 0x1abcb5b8 <line:2:1, line:11:1> line:2:5 main 'int ()'
  `-CompoundStmt 0x1abcb818 <col:11, line:11:1>
    |-BinaryOperator 0x1abcb698 <line:3:5, col:9> 'int' '='
    | |-DeclRefExpr 0x1abcb658 <col:5> 'int' lvalue Var 0x1abcb4b0 'a' 'int'
    | `-IntegerLiteral 0x1abcb678 <col:9> 'int' 10
    |-IfStmt 0x1abcb7c0 <line:4:2, line:9:2> has_else
    | |-BinaryOperator 0x1abcb710 <line:4:6, col:8> 'int' '>'
    | | |-ImplicitCastExpr 0x1abcb6f8 <col:6> 'int' <LValueToRValue>
    | | | `-DeclRefExpr 0x1abcb6b8 <col:6> 'int' lvalue Var 0x1abcb4b0 'a' 'int'
    | | `-IntegerLiteral 0x1abcb6d8 <col:8> 'int' 0
    | |-CompoundStmt 0x1abcb760 <col:11, line:6:2>
    | | `-ReturnStmt 0x1abcb750 <line:5:3, col:10>
    | |   `-IntegerLiteral 0x1abcb730 <col:10> 'int' 1
    | `-CompoundStmt 0x1abcb7a8 <line:7:6, line:9:2>
    |   `-ReturnStmt 0x1abcb798 <line:8:3, col:10>
    |     `-IntegerLiteral 0x1abcb778 <col:10> 'int' 0
    `-ReturnStmt 0x1abcb808 <line:10:5, col:12>
      `-IntegerLiteral 0x1abcb7e8 <col:12> 'int' 3
```

Example: int a;

```
int main(){
    return 3;
}
```

```
int a;
int main(){
    return 3;
}
```

VarDecl → int id;

VarDecl → Type Vars;
Type → int | float | double | …;
Vars → Vars VarDef | VarDef
VarDef → id ‘=‘ Initval | id
Initval → val

```
CompUnit: xwVarDef FuncDef {
    // global variable + function
    llvm::errs() << " -- xwVarDef FuncDef\n";
    auto inner2 = stak.back();
    stak.pop_back();
    auto inner1 = stak.back();
    stak.pop_back();
    stak.push_back(llvm::json::Object{{"kind", "TranslationUnitDecl"},
                                      {"inner", llvm::json::Array{inner1, inner2}}});
 }
 | xwVarDef {
    // global variable only
    llvm::errs() << " -- xwVarDef\n";
    auto inner = stak.back();
    stak.pop_back();
    stak.push_back(llvm::json::Object{{"kind", "TranslationUnitDecl"},
                                      {"inner", llvm::json::Array{inner}}});
 }
 | FuncDef {
    // global function only
    llvm::errs() << " -- FuncDef\n";
    auto inner = stak.back();
    stak.pop_back();
    stak.push_back(llvm::json::Object{{"kind", "TranslationUnitDecl"},
                                      {"inner", llvm::json::Array{inner}}});
 }
 | %empty // neither

xwVarDef: T_INT Ident T_SEMI {
    llvm::errs() << " -- VarDecl\n";
    auto name = stak.back().getAsObject();
    assert(name != nullptr);
    assert(name->get("value") != nullptr);
    stak.pop_back();
    stak.push_back(llvm::json::Object{{"kind", "VarDecl"},
                                      {"name", *(name->get("value"))}});
 }
```

Example: a = 10;

```c
int main(){
    return 3;
}
```

```c
int a;
int main(){
    return 3;
}
```

```c
int a;
int main(){
    a = 10;
    return 3;
}
```

```
BlockItem: xwStmt {
  auto inner = stak.back();
  stak.pop_back();
  stak.push_back(llvm::json::Object{{"kind", "CompoundStmt"},
                                    {"inner", llvm::json::Array{inner}}});
}

BlockItem: BlockItem xwStmt {
  auto inner = stak.back();
  stak.pop_back();
  auto fa = stak.back();
  fa.getAsObject()->get("inner")->getAsArray()->push_back(inner);
  stak.pop_back();
  stak.push_back(fa);
}

xwStmt: xwBinaryOperator
      | xwIfStmt
      | RetStmt

xwBinaryOperator: xwBinaryOperatorExp T_SEMI {
    llvm::errs() << " -- xwBinaryOperatorExp\n";
}

xwBinaryOperatorExp: Ident xwOp Exp {
  auto exp = stak.back();
  stak.pop_back();
  auto ident = stak.back();
  stak.pop_back();
  stak.push_back(llvm::json::Object{{"kind", "BinaryOperator"},
                                    {"inner", llvm::json::Array{ident,exp}}});
}

xwOp: T_EQUAL
    | T_GREATER
```

Example: if-else;

```
int main(){
    return 3;
}
```

```
int a;
int main(){
    return 3;
}
```

```
int a;
int main(){
    a = 10;
    return 3;
}
```

```
int a;
int main(){
    a = 10;
        if( a>0 ){
                return 1;
        }
        else{
                return 0;
        }
    return 3;
}
```

```
xwStmt: xwBinaryOperator
        | xwIfStmt
        | RetStmt

xwBinaryOperator: xwBinaryOperatorExp T_SEMI {
    llvm::errs() << " -- xwBinaryOperatorExp\n";
}

xwBinaryOperatorExp: Ident xwOp Exp {
  auto exp = stak.back();
  stak.pop_back();
  auto ident = stak.back();
  stak.pop_back();
  stak.push_back(llvm::json::Object{{"kind", "BinaryOperator"},
                                  {"inner", llvm::json::Array{ident,exp}}});
}

xwOp: T_EQUAL
    | T_GREATER

xwIfStmt: T_IF T_L_PAREN xwBinaryOperatorExp T_R_PAREN Block T_ELSE Block {
  llvm::errs() << " -- IfStmt\n";
  auto inner3 = stak.back();
  stak.pop_back();
  auto inner2 = stak.back();
  stak.pop_back();
  auto inner1 = stak.back();
  stak.pop_back();
  stak.push_back(llvm::json::Object{{"kind", "IfStmt"},
                        {"inner", llvm::json::Array{inner1, inner2, inner3}}});
}
    | T_IF T_L_PAREN xwBinaryOperatorExp T_R_PAREN Block {}
```

Example: Parse Tree

```
int main(){
    return 3;
}
```

yylex()

①

```
{
  "value": "main"
}
{
  "kind": "IntegerLiteral",
  "value": "3"
}
```

RetStmt: T_RETURN Exp T_SEMI {

②

```
{
  "value": "main"
}
{
  "inner": [
    {
      "kind": "IntegerLiteral",
      "value": "3"
    }
  ],
  "kind": "ReturnStmt"
}
```

BlockItem: xwStmt {

③

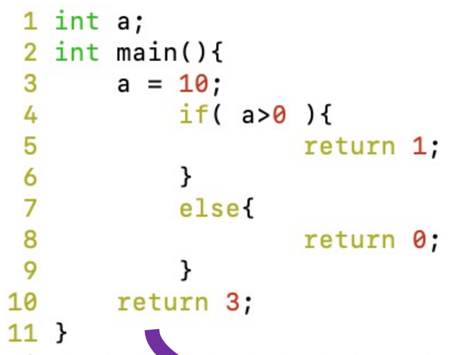
```
{
  "value": "main"
}
{
  "inner": [
    {
      "inner": [
        {
          "kind": "IntegerLiteral",
          "value": "3"
        }
      ],
      "kind": "ReturnStmt"
    }
  ],
  "kind": "CompoundStmt"
}
```

FuncDef: T_INT Ident T_L_PAREN T_R_PAREN Block {

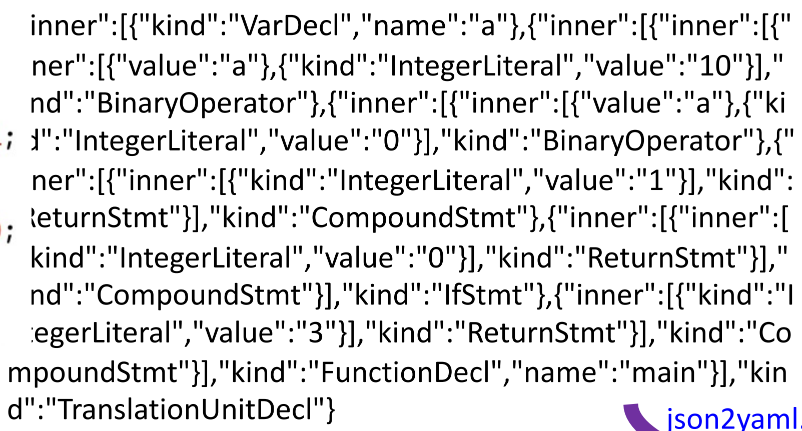
④

```
{
  "inner": [
    {
      "inner": [
        {
          "inner": [
            {
              "kind": "IntegerLiteral",
              "value": "3"
            }
          ],
          "kind": "ReturnStmt"
        }
      ],
      "kind": "CompoundStmt"
    }
  ],
  "kind": "FunctionDecl",
  "name": "main"
}
```

Example: Parse Tree (cont.)

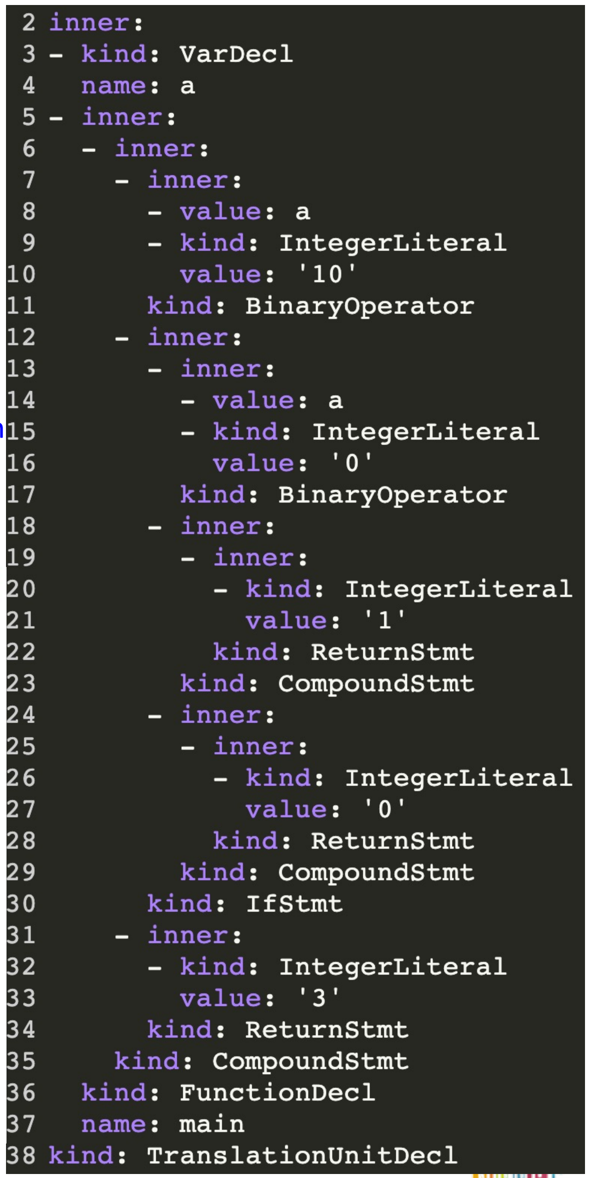


```c
int a;
int main(){
  a = 10;
    if( a>0 ){
        return 1;
    }
    else{
        return 0;
    }
  return 3;
}
```

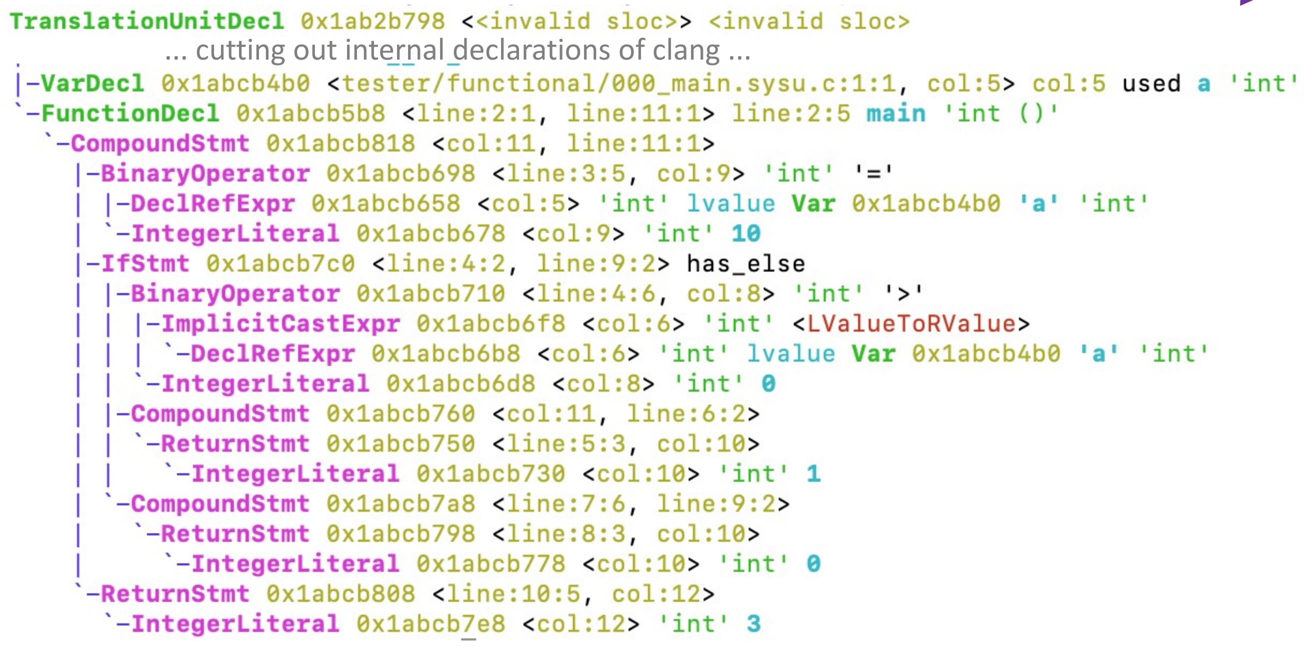


inner":[{"kind":"VarDecl","name":"a"},{"inner":[{"inner":[{"
ner":[{"value":"a"},{"kind":"IntegerLiteral","value":"10"}],"
nd":"BinaryOperator"},{"inner":[{"inner":[{"value":"a"},{"ki
d":"IntegerLiteral","value":"0"}],"kind":"BinaryOperator"},{"
ner":[{"inner":[{"kind":"IntegerLiteral","value":"1"}],"kind":
ReturnStmt"}],"kind":"CompoundStmt"},{"inner":[{"inner":[
kind":"IntegerLiteral","value":"0"}],"kind":"ReturnStmt"}],"
nd":"CompoundStmt"}],"kind":"IfStmt"},{"inner":[{"kind":"I
tegerLiteral","value":"3"}],"kind":"ReturnStmt"}],"kind":"Co
mpoundStmt"}],"kind":"FunctionDecl","name":"main"}],"kin
d":"TranslationUnitDecl"}

json2yaml.com



```yaml
inner:
- kind: VarDecl
  name: a
- inner:
  - inner:
    - inner:
      - value: a
      - kind: IntegerLiteral
        value: '10'
      kind: BinaryOperator
    - inner:
      - inner:
        - value: a
        - kind: IntegerLiteral
          value: '0'
        kind: BinaryOperator
      - inner:
        - inner:
          - kind: IntegerLiteral
            value: '1'
          kind: ReturnStmt
        kind: CompoundStmt
      - inner:
        - inner:
          - kind: IntegerLiteral
            value: '0'
          kind: ReturnStmt
        kind: CompoundStmt
      kind: IfStmt
    - inner:
      - kind: IntegerLiteral
        value: '3'
      kind: ReturnStmt
    kind: CompoundStmt
  kind: FunctionDecl
  name: main
kind: TranslationUnitDecl
```



```
TranslationUnitDecl 0x1ab2b798 <<invalid sloc>> <invalid sloc>
```

... cutting out internal declarations of clang ...

```
|-VarDecl 0x1abcb4b0 <tester/functional/000_main.sysu.c:1:1, col:5> col:5 used a 'int'
`-FunctionDecl 0x1abcb5b8 <line:2:1, line:11:1> line:2:5 main 'int ()'
  `-CompoundStmt 0x1abcb818 <col:11, line:11:1>
    |-BinaryOperator 0x1abcb698 <line:3:5, col:9> 'int' '='
    | |-DeclRefExpr 0x1abcb658 <col:5> 'int' lvalue Var 0x1abcb4b0 'a' 'int'
    | `-IntegerLiteral 0x1abcb678 <col:9> 'int' 10
    |-IfStmt 0x1abcb7c0 <line:4:2, line:9:2> has_else
    | |-BinaryOperator 0x1abcb710 <line:4:6, col:8> 'int' '>'
    | | |-ImplicitCastExpr 0x1abcb6f8 <col:6> 'int' <LValueToRValue>
    | | | `-DeclRefExpr 0x1abcb6b8 <col:6> 'int' lvalue Var 0x1abcb4b0 'a' 'int'
    | | `-IntegerLiteral 0x1abcb6d8 <col:8> 'int' 0
    | |-CompoundStmt 0x1abcb760 <col:11, line:6:2>
    | | `-ReturnStmt 0x1abcb750 <line:5:3, col:10>
    | |   `-IntegerLiteral 0x1abcb730 <col:10> 'int' 1
    | `-CompoundStmt 0x1abcb7a8 <line:7:6, line:9:2>
    |   `-ReturnStmt 0x1abcb798 <line:8:3, col:10>
    |     `-IntegerLiteral 0x1abcb778 <col:10> 'int' 0
    `-ReturnStmt 0x1abcb808 <line:10:5, col:12>
      `-IntegerLiteral 0x1abcb7e8 <col:12> 'int' 3
```

其他

- Parser细节(文法、状态等)
  - $bison -v parser.y
    - 输出：./parser.output
- 文法规则参考
  - https://buaa-se-compiling.github.io/miniSysY-tutorial/
  - https://github.com/Komorebi660/SysYF-Compiler/blob/master/grammar/SysYFParser.yy
- Jason to XML
  - https://json2yaml.com/
- Clang/LLVM Tutorial
  - Introduction to Clang AST, https://clang.llvm.org/docs/IntroductionToTheClangAST.html
  - https://www.cs.rochester.edu/u/criswell/asplos19/ASPLOS19-LLVM-Tutorial.pdf
- Bison
  - Introduction to Bison, https://web.stanford.edu/class/archive/cs/cs143/cs143.1128/handouts/120%20Introducing%20bison.pdf
  - Compiler construction using Flex and Bison, http://www.admb-project.org/tools/flex/compiler.pdf
  - Bison, https://www.gnu.org/software/bison/manual/bison.pdf