

编译原理—词法分析器实验报告

09016414 罗崑洪

实验目的

根据编译原理课程知识，编写一个词法分析器，以此深入理解词法分析的规则。

实验内容

1. 选择一个你熟悉的程序设计语言，找到它的规范（reference or standard）。在规范中找到其词法的BNF或正规式描述。
2. 选择该语言的一个子集（能够构成一个mini的语言，该语言至少能够进行函数调用、控制流语句（分支或循环）、简单的运算和赋值操作。）给出该mini语言的词法的正规文法或正规式。
3. 将该正规文法或正规式转换为NFA。给出转换过程。
4. 将上述NFA转换为DFA，并最小化。给出具体过程。
5. 根据最小化后的DFA写出词法分析程序。将该词法分析程序打印出来。
6. 用mini语言写出若干个小程序作为测试用例。用上述词法分析程序对这些小程序进行词法分析，并将测试用例和分析结果打印出来。

实验假设

1. 识别语言的正则表达式描述

选用C语言的子集作为词法分析器所识别的语言，参考C99标准，并为了方便起见对词法进行了重新整理。词法的正则表达式可以用lex语言描述如下：

```
1  /* operators */
2  // 赋值运算符
3  "=" {return(=)}
4  "+=" {return(+=)}
5  "-=" {return(-=)}
6  "*=" {return(*=)}
7  "/=" {return(/=)}
8  "%=" {return(%=)}
9  "<<=" {return(<<=)}
10 ">>=" {return(>>=)}
11 // 算数运算符
12 "+" {return(+)}
13 "-" {return(-)}
14 "*" {return(*)}
15 "/" {return(/)}
16 "%" {return(%)}
17 "<<" {return(<<)}
18 ">>" {return(>>)}
19 // 单目运算符
20 "++" {return(++)}
```

```

21  "--" {return(--)}
22  "~" {return(~)}
23  // 位运算符
24  "&" {return(&)}
25  "|" {return(|)}
26  // 逻辑运算符
27  "&&" {return(&&)}
28  "||" {return(||)}
29  "!" {return(!)}
30  // 关系运算符
31  "==" {return(eq)}
32  "<" {return(lt)}
33  ">" {return(gt)}
34  "<=" {return(lte)}
35  ">=" {return(gte)}
36  /* 标点符号、界符 */
37  ":" {return(:)}
38  "?" {return(?)}
39  "." {return(.)}
40  "," {return(,)}
41  ";" {return(;)}
42  "\" {return(\"\\\"") }
43  "'" {return("'\"") }
44  "{" {return(LP)}
45  "}" {return(RP)}
46  "(" {return(SLP)}
47  ")" {return(SRP)}
48  "[" {return(SLB)}
49  "]" {return(SRB)}
50  /* 标识符 */
51  "[_a-zA-Z][_a-zA-Z0-9]*" {return(id)}
52  /* 常量 */
53  "[0-9]+" {return(integer)}
54  "[0-9]+(.[0-9]+)?" {return(real)}

```

2. 关键字约定

除了以上正则表达式外，C99标准中还规定了一些关键字。我们使用符号表来保存这些关键字，词法分析中，遇到标识符后，词法分析器会先查找该标识符是否存在于符号表的关键字集合中，以此判断其是用户定义的标识符或是保留字。

支持的关键字如下：

```

1 // type specifier
2 int char bool void
3 // storage class specifier
4 typedef static
5 // type qualifier
6 const
7 // statements
8 if else switch case default for
9 do while break continue return
10 // others
11 struct enum class public private
12 this new delete friend true false

```

逐步建立有穷自动机

由于我们对于实验假设的语言具有一定的规模，不难想象一开始构建的NFA状态数会相当庞大（事实也是如此，只要想想26个字母、10个数字和其他标点符号就能大概估算其状态数），以至于难以用手算的方式进行建立和化简。考虑到下学期的课程设计中要实现lex，本着发挥动手能力，践行“实验精神”的初衷，我们决定使用算法来完成有穷自动机的逐步建立。

因此，在下文中，我们会给出建立和化简中得到的状态转换表，但不会给出图形化的状态转换图（因为状态数量过于庞大）。当然，我们会对具体实现思路进行详细说明。

正则表达式的预处理

输入的正则表达式往往是人容易理解的形式，但对于程序而言，要处理这些表达式还需要经过一些步骤。

支持扩展的语法

在实验假设中，我们用到了扩展的正则表达式语法，形如 `[0-9]`，这表示0到9间的任意字符。为了简化后续程序的复杂度，我们可以把它转化为未经扩展的等价形式。这一轮处理后的效果如下所示。

```

1 相关代码见 Regex.cpp 中 void Regex::preprocess()函数
2 输入：. [_a-zA-Z]
3 输出：. ( _ | a | b | c | A | B | C )

```

增加显式的连接符号

实际上，连接运算也是正则表达式中运算的一种，但它常常容易被忽视，这将导致之后在建立NFA时遇到问题。为了避免这种情况，我们需要对正则表达式显式增加连接符号。

这一轮处理后的效果如下所示，我们规定 `#` 表示连接操作。

```

1 相关代码见 Regex.cpp 中 void Regex::preprocess()函数
2 输入：ab*c | (d|ef)+
3 输出：a#b*#c | (d|e#f)+

```

中缀正则表达式转后缀正则表达式

为了让程序构造NFA时简单一些，需要把中缀表达式转换为后缀表达式，这样我们可以对表达式的优先级不加区分。

算法思想如下：

```
1 // Regex.cpp
2 /**
3 | 操作符      | (      | #      | \      | )      |
4 | ----- | ---- | ---- | ---- | ---- |
5 | 栈外优先级 | 6      | 4      | 2      | 1      |
6 | 栈内优先级 | 1      | 5      | 3      | 6      |
7 */
8 string Regex::to_postfix()
9 {
10     /* pattern只有一个字符，则直接返回 */
11     if (pattern.size() == 1) {
12         postfix = pattern;
13         return postfix;
14     }
15
16     /* 预处理，仅执行一次 */
17     preprocess();
18
19     stack<char> op; // 操作符栈
20     op.push('$'); // 哨兵
21     pattern += '$'; // 哨兵
22
23     for (int i = 0, n = pattern.size(); i < n && !op.empty(); i++) {
24         char current = pattern[i];
25         /* 转义字符处理 */
26         if (current == '\\') {
27             if (i + 1 < n && RegexOperator::includes(pattern[i + 1])) {
28                 postfix += '\\';
29                 postfix += pattern[i + 1];
30                 i += 2;
31             }
32         }
33         /* 操作符处理 */
34         else if (RegexOperator::includes(current)) {
35             /* 单目运算符，直接输出 */
36             if (RegexOperator::typeof(current) == 1) {
37                 postfix += current;
38                 i++;
39                 continue;
40             }
41             /* 双目运算符，根据栈内栈外优先级执行算法 */
42             char top = op.top();
43             if (RegexOperator::in_comming_priority(current) >
44                 RegexOperator::in_stack_priority(top)) {
45                 op.push(current);
46                 i++;
47             }
48             else if (RegexOperator::in_comming_priority(current) <
49                 RegexOperator::in_stack_priority(top)) {
50                 postfix += top;
51                 op.pop();
52             }
53             else {
54                 postfix += current;
55                 op.push(current);
56                 i++;
57             }
58         }
59     }
60     postfix += op.top();
61     op.pop();
62     return postfix;
63 }
```

```

50         }
51         else if (RegexOperator::in_comming_priority(current) ==
RegexOperator::in_stack_priority(top)) {
52             op.pop();
53             if (top == '(')
54                 i++;
55         }
56     }
57     /* 操作数直接输出 */
58     else {
59         postfix += current;
60         i++;
61     }
62 }
63
64 return postfix;
65 }

```

这属于数据结构中的知识，在这里不加以展开。

由正则表达式集合建立NFA

正则表达式转NFA

我们可以使用 McNaughton-Yamada-Thompson 算法从正则表达式构造NFA，而注意到实验假设中我们对正则表达式进行了扩展（加入了 `?` 和 `+` 运算符），因此延续该算法的思想，构造过程如下：

对于字母表中的表达式 a ，构造如下的NFA：

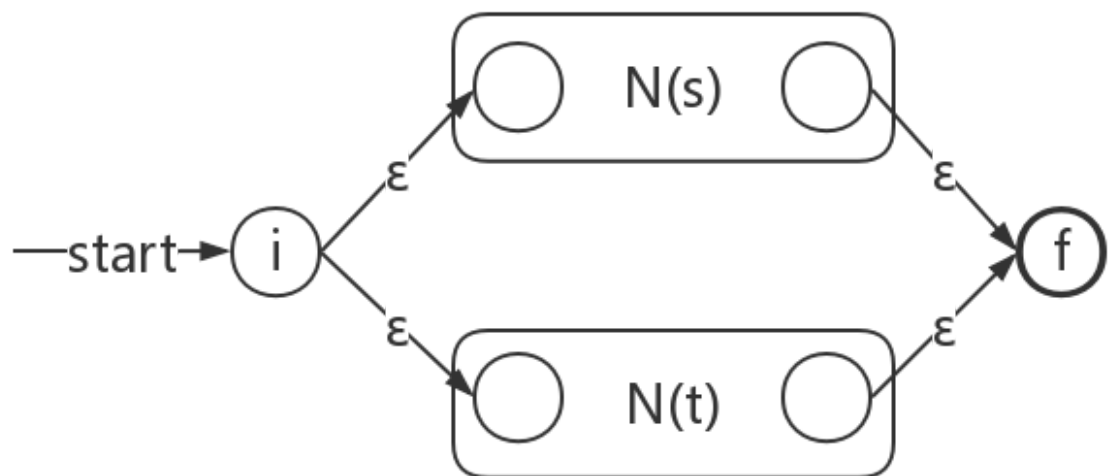


对于表达式 ϵ ，构造如下的NFA：



下面开始递推地构造NFA。假设正则表达式 s 和 t 的NFA分别为 $N(s)$ 和 $N(t)$ ，则新的NFA可那如下方法构造：

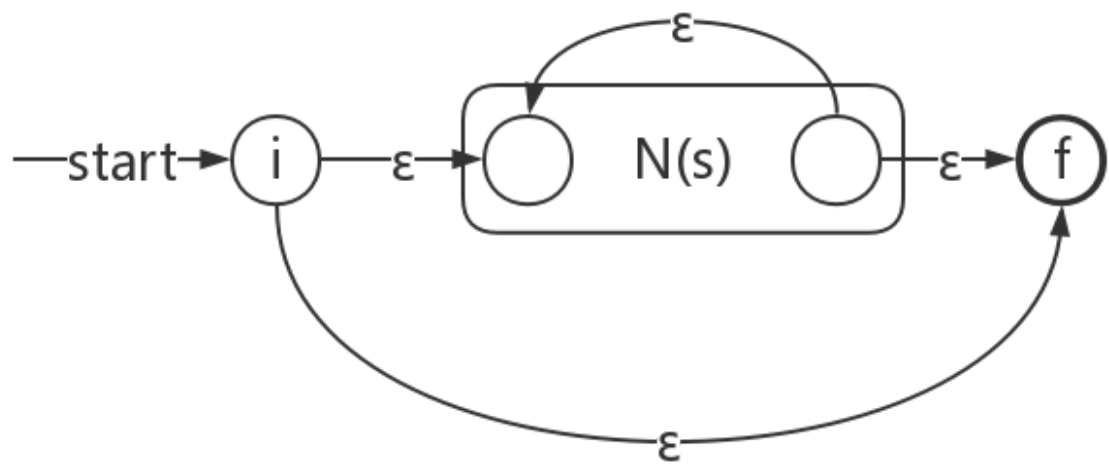
1. $r = s|t$



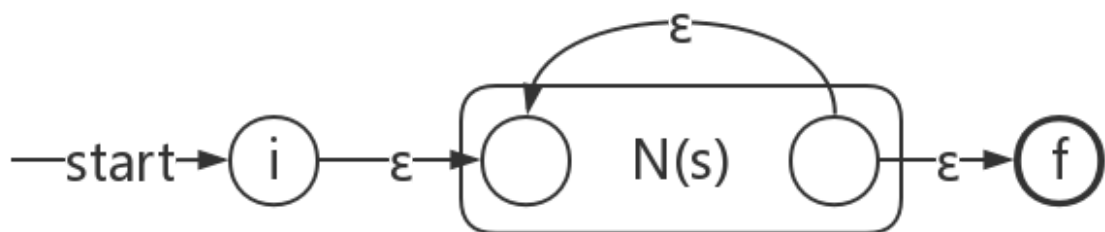
2. $r = st$



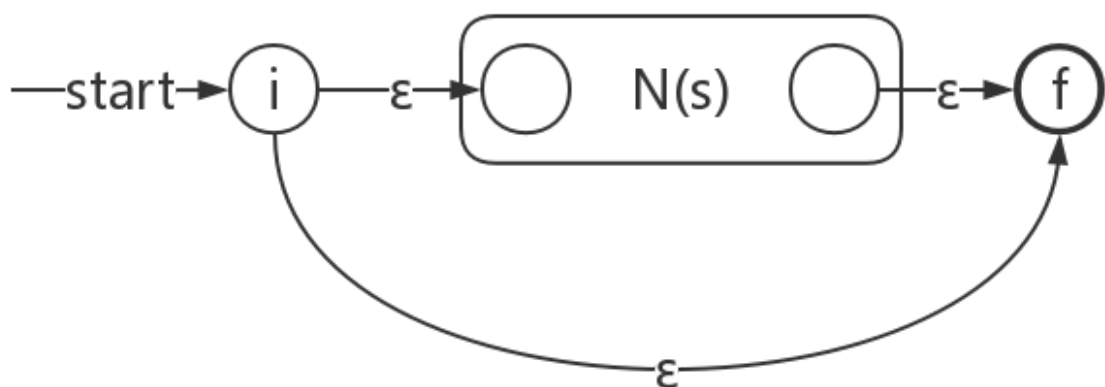
3. $r = s^*$



4. $r = s^+$



5. $r=s$?



6. $r=(s)$

这种情况不需要构建新的NFA。

根据上述思想，我们可以根据单个正则表达式建立NFA。相关代码可以见 `NFA.cpp` 中的 `void NFA::create_from_regex(const string & postfix_pattern, const string & action)` 函数。注意，由于之后建立状态转换表的时候还需要用到终结符集合以及接收状态的行为 `action`，因此这个函数还负责记录该NFA的字母表（即终结符集）以及NFA所识别的语言 `action`。

正则表达式集合转NFA

我们可以对每个正则表达式建立一个NFA，最终用一个新的开始节点连接这些NFA的开始节点，并把它们的终结节点连接到新的终结节点，得到一个完整的NFA。相关代码见 `NFA.cpp` 中的 `void NFA::create_from_regexes(const string & postfix_pattern, const string & action)` 函数。

NFA转DFA

我们可以根据子集构造法来构造DFA，该算法在龙书上有详细介绍，这里对算法本身不做过多讨论，而讨论其实现方法。由于这属于优化阶段，因此算法定义在 `optimizer` 类中。

要实现这个算法，需要实现几个重要的函数（具体代码见 `optimizer.cpp`）：

1. `move(T, a)`：返回能够从T中某状态s出发，通过标号为a的转换到达的NFA状态的集合。该算法通过对状态机的广度优先搜索实现。注意到接下来求 ϵ 闭包时，返回的集合仅仅比 `move(T, ϵ)` 多了集合T中的元素，因此我们可以直接在这里实现 ϵ 闭包，根据传入参数 `a` 的值判断返回的集合中是否要加上集合T中的元素。

```
1 // Optimizer.cpp
2 set<FANode*> Optimizer::move(set<FANode*> & T, int a)
3 {
4     set<FANode*> a_closure_T;
5
6     stack<FANode*> stack;
7     if (T.size() == 0) {
8         // 空集访问任何元素也都是空集
9         return a_closure_T;
10    }
11    for (set<FANode*>::iterator it = T.begin(); it != T.end(); it++) {
12        stack.push(*it);
13        if (a == FANode::EPSILON)
14            a_closure_T.insert(*it);
15    }
16
17    while (!stack.empty()) {
18        FANode* t = stack.top();
19        stack.pop();
20        vector<pair<int, FANode*>>* t_trans = t->get_transition_table();
21
22        for (auto u : *t_trans) {
23            if (u.first == a && a_closure_T.find(u.second) ==
a_closure_T.end()) {
24                a_closure_T.insert(u.second);
25                stack.push(u.second);
26            }
27        }
28    }
29    return a_closure_T;
30 }
```

2. `epsilon_closure(T)`：返回能够从T中某NFA状态s开始只通过 ϵ 转换到达的NFA状态集合。这里我们直接用 `move` 函数实现。

```
1 // Optimizer.cpp
2 set<FANode*> Optimizer::epsilon_closure(set<FANode*> & T)
3 {
4     return move(T, FANode::EPSILON);
5 }
```

3. `epsilon_closure(s)`：返回通过状态s开始只通过 ϵ 转换到达的NFA状态集合。注意这个函数只会在算法的第一步用到，因此没必要额外实现。我们可以向 `epsilon_closure(T)` 函数传入一个只包含初始状态的状态集合即可达到我们的目的。

实现了上述操作后，我们可以实现子集构造法。与书中不同的是，我们在实现时并不真的需要对在状态转换表中完成记录的状态进行标记，原因是因为新加入的状态总被加入了DFA状态集合列表 `d_states` 的末端，而迭代在索引到达列表尾部结束，这意味着当没有新状态产生后，由于 `d_states` 的大小不再增加，因此下标到达末尾相当于对所有状态都进行过标记了。为了减少数据维度，我们使用 `d_actions` 队列来记录相应下标状态的行为。对于开始状态，给其一个特殊的行为 `start` 作为标记。

```
1 // Optimizer.cpp
2 void Optimizer::build_dfa(NFA & nfa)
3 {
4     FANode * s0 = nfa.get_start_node();
5     set<FANode*> t;
6     t.insert(s0);
7     set<FANode*> t_closure = epsilon_closure(t);
8     d_states.push_back(t_closure);
9     d_actions.push_back("start");
10    alphabet = nfa.get_alphabet();
11    for (int i = 0; i < d_states.size(); i++) {
12        for (int a : alphabet) {
13            set<FANode*> a_closure = move(d_states[i], a);
14            set<FANode*> u = epsilon_closure(a_closure);
15            int u_index = get_index(u);
16            if (u_index == -1) {
17                d_states.push_back(u);
18                d_actions.push_back(get_action(u));
19                u_index = get_index(u);
20            }
21            d_trans[pair<int, int>(i, a)] = u_index;
22        }
23    }
24 }
```

最小化DFA

使用划分的方法对DFA状态进行最小化，该算法有三个阶段：

1. 初始化分组，和书中实现有些区别的是，我们把最初的分组划分为：
 - 非接受状态
 - 不同的接受状态（即采取不同 `action`，或者说识别不同正则表达式的状态）
 - 死状态

注意到我们在构造DFA的时候可能出现了死状态（我们标记死状态的 `action` 为 `phai`），这是容易理解的。死状态并不影响我们之后的操作，并且维护了算法操作的统一性，因此不需要清理。当然，死状态的出现会使得最终的状态转换表是稠密的，因为非法的输入最终都将指向死状态，这对于空间利用上可能不太友好。如果出现了死状态，有且仅有一个，具体地，死状态是一个空集（size为0的set），由于使用了 `map`，因此算法会认为所有的空集都是一样的，执行时不会有多个不同的空集出现在DFA的状态集合中。

2. 迭代并产生最终的划分，这也将产生最小DFA的状态转换表
3. 确定每个状态的action

篇幅有限，具体代码可见 `optimizer.cpp` 中的 `void optimizer::minimize_dfa()` 函数。

模拟最小DFA的执行

虽然最小DFA也是状态机，可以用图的数据结构进行模拟，但对于状态较多的DFA，构建图的过程是比较耗时耗力的，而且状态之间除了要转移，还要考虑可能的“退回上一状态操作”。实际上，在之前执行算法后得到的最小DFA状态转换表以及各个状态对应的action就已经足以描述一个DFA，根据状态转换表写出转移函数 `move` 即可实现状态转移。

```
1 // DFA.cpp
2 int DFA::move(int state, int step)
3 {
4     if (trans.count(pair<int, int>(state, step)) == 1)
5         return trans[pair<int, int>(state, step)];
6     return -1;
7 }
```

在实现 `move` 函数后，就可以模拟DFA的执行。执行中要考虑对空白字符的过滤，同时要考虑报错信息。因此在处理程序代码时，还要记录当前的行号、行内偏移等信息，以便进行报错。

此外，根据书上的思想，要实现最长匹配的原则，例如不能识别到 `bool` 就返回保留字 `bool`，因为这可能是一个名为 `boolean` 的 `identifier`。因此设置两个索引，`lexeme_begin` 和 `peek_forward`，前者记录词素起始位置，后者则一直向前“偷窥”，直到遇到了最长匹配的词素，程序进行返回。

出错的情况

1. 输出完前一个token后，进行下一轮处理时，发现DFA的开始状态对于读头输入无法走通。
2. 已经读到程序结尾，但状态是非接收状态。

符号表维护

实际上，符号表是词法分析器和语法分析去共同维护的。在当前版本中，符号表暂时作为词法分析器的内部成员，且功能并不完全，仅仅是表达一下动机。

在执行中，我们如果识别了id，那么对于这个id，我们先查询其是否已经出现在符号表中。如果是，说明这个id是保留字，输出的token将其类型说明为该保留字，而非用户定义的id。

当然，对于真正的符号表，还需要记录id的更多信息，并且对于用户定义的id也要在符号表中进行记录，当前的词法分析器里没有完成这一功能。

此外，虽然可以对保留字也设置相关正则表达式，以使得DFA可以和识别其他模式一样识别这些保留字，但这会增加许多额外状态，大幅降低DFA的构建过程，而且违背建立符号表的思想。经过实验，如果我们把这些保留字作为正则规则加入DFA，那么实验假设的语言的最小DFA需要花一分多钟，而是用符号表可以把时间降到30秒左右。

对于符号表的建立过程，我们放在了DFA的构造函数中。下面的代码片段中 `scan` 函数给出了DFA模拟的过程，篇幅有限，部分代码以 `//...` 形式省略。完整代码见 `DFA.cpp`。

```
1 // DFA.cpp
2 DFA::DFA(map<pair<int, int>, int> trans, vector<string> actions)
3 {
4     // ...
5     /* keywords */
6     // type specifier
```

```

7     symbol_table.push_back("int");
8     symbol_table.push_back("char");
9     symbol_table.push_back("bool");
10    // ...
11    // storage class specifier
12    symbol_table.push_back("typedef");
13    symbol_table.push_back("static");
14    // type qualifier
15    symbol_table.push_back("const");
16    // statements
17    symbol_table.push_back("if");
18    symbol_table.push_back("else");
19    symbol_table.push_back("switch");
20    symbol_table.push_back("case");
21    // ...
22 }
23
24 void DFA::scan(const string & code)
25 {
26     // ...
27     while (true) {
28         // 遇到空白符
29         while (遇到空白符) {
30             if (code[peek_forward] == '\n') {
31                 // 换行符, 行号加一, 行内偏移量归零
32                 cnt_line++;
33                 cnt_offset = 0;
34             }
35             else {
36                 // 碰到空格或tab
37                 cnt_offset++;
38             }
39             peek_forward++;
40         }
41         // 到达源代码末尾
42         if (peek_forward >= N)
43             return;
44         cnt_offset++; // 找到了一个词素起始点, 行内字符序号加1, 记录
45         lexeme_begin = peek_forward;
46         int cur_state = start_state;
47         // 最长匹配原则
48         do {
49             cur_state = move(cur_state, (int)code[peek_forward]);
50             if (cur_state == -1) {
51                 // 错误处理
52             }
53             peek_forward++;
54             cnt_offset++;
55         } while (peek_forward < N && actions[cur_state] == "");
56         // 读到程序结尾, 但是状态是非接受状态
57         if (cur_state != -1 && actions[cur_state] == "") {
58             // 错误处理
59         }

```

```

60 // 从该接受状态开始，继续走过连续的接收状态，直至走到一个非终结状态或达到死状态phai
61 if (cur_state != -1 && actions[cur_state] != "") {
62     // 用mark来标记最后到达的接收状态（最长匹配）
63     int mark = cur_state;
64     while (peek_forward < N) {
65         cur_state = move(cur_state, (int)code[peek_forward]);
66         if (是非终结状态或达到死状态phai)
67             break;
68         mark = cur_state;
69         cnt_offset++;
70         peek_forward++;
71     }
72     // 处理符号表
73     string attribute = actions[mark];
74     string value = code.substr(lexeme_begin, peek_forward - lexeme_begin);
75     if (attribute == "id") {
76         if (在符号表里出现) {
77             // id识别为保留字
78             continue;
79         }
80     }
81     // 输出token
82 }
83 }
84 }

```

测试

输入样例

接下来我们对程序进行测试，假定输入的样例程序如下所示。这个程序没有具体意义，但尽量覆盖了实验假定语言中可能出现的情况。

```

1 // code.cpp
2 class Helper {
3 private:
4     Helper();
5     ~Helper();
6 public:
7     static struct tool {
8         enum color {
9             RED,
10            BLUE,
11            GREEN
12        };
13        enum sex {
14            MALE,
15            FEMALE
16        };
17    } tool;
18    static bool greater_or_equal_than(int a, int b) {
19        if((a >= b) && ((a > b) || (a == b)))

```

```

20         return true;
21     return false;
22 }
23 };
24
25
26
27 int main() {
28     typedef double myDouble;
29     myDouble myNum = 10.5;
30     char c = 'a';
31     const int hisNum = (int) c;
32     int herNum = myNum - hisNum;
33     bool flag = true;
34     do {
35         myNum %= 22;
36     } while (!flag);
37     for(int i = 0; i < 100; i++) {
38         if(herNum == myNum) {
39             break;
40         }
41         else if(Helper::greater_or_equal_than(herNum, myNum)) {
42             herNum--;
43             continue;
44         }
45         else {
46             herNum++;
47             continue;
48         }
49         myNum = (myNum != hisNum ? ++myNum : --myNum);
50     }
51     switch(herNum) {
52         case 10:
53             herNum *= myNum;
54             break;
55         case 20:
56             herNum = Helper::tool.RED;
57         default:
58             herNum -= myNum;
59     }
60     return 0;
61 }

```

输出结果

词法分析器给出的输出结果如下，由此可以验证其执行的正确性：

```

1 < 0      class    class >
2 < 1      id       Helper >
3 < 2      LP       { >
4 < 3      private  private >
5 < 4      :        : >
6 < 5      id       Helper >

```

```

7 < 6 SLP ( >
8 < 7 SRP ) >
9 < 8 ; ; >
10 < 9 ~ ~ >
11 < 10 id Helper >
12 < 11 SLP ( >
13 < 12 SRP ) >
14 < 13 ; ; >
15 < 14 public public >
16 < 15 : : >
17 < 16 static static >
18 < 17 struct struct >
19 < 18 id tool >
20 < 19 LP { >
21 < 20 enum enum >
22 < 21 id color >
23 < 22 LP { >
24 < 23 id RED >
25 < 24 , , >
26 < 25 id BLUE >
27 < 26 , , >
28 < 27 id GREEN >
29 < 28 RP } >
30 < 29 ; ; >
31 < 30 enum enum >
32 < 31 id sex >
33 < 32 LP { >
34 < 33 id MALE >
35 < 34 , , >
36 < 35 id FEMALE >
37 < 36 RP } >
38 < 37 ; ; >
39 < 38 RP } >
40 < 39 id tool >
41 < 40 ; ; >
42 < 41 static static >
43 < 42 bool bool >
44 < 43 id greater_or_equal_than >
45 < 44 SLP ( >
46 < 45 int int >
47 < 46 id a >
48 < 47 , , >
49 < 48 int int >
50 < 49 id b >
51 < 50 SRP ) >
52 < 51 LP { >
53 < 52 if if >
54 < 53 SLP ( >
55 < 54 SLP ( >
56 < 55 id a >
57 < 56 gte >= >
58 < 57 id b >
59 < 58 SRP ) >

```

```

60 < 59  &&    && >
61 < 60  SLP    ( >
62 < 61  SLP    ( >
63 < 62  id     a >
64 < 63  gt     > >
65 < 64  id     b >
66 < 65  SRP    ) >
67 < 66  ||     || >
68 < 67  SLP    ( >
69 < 68  id     a >
70 < 69  eq     == >
71 < 70  id     b >
72 < 71  SRP    ) >
73 < 72  SRP    ) >
74 < 73  SRP    ) >
75 < 74  return return >
76 < 75  true   true >
77 < 76  ;      ; >
78 < 77  return return >
79 < 78  false  false >
80 < 79  ;      ; >
81 < 80  RP     } >
82 < 81  RP     } >
83 < 82  ;      ; >
84 < 83  int    int >
85 < 84  id     main >
86 < 85  SLP    ( >
87 < 86  SRP    ) >
88 < 87  LP     { >
89 < 88  typedef typedef >
90 < 89  id     double >
91 < 90  id     myDouble >
92 < 91  ;      ; >
93 < 92  id     myDouble >
94 < 93  id     myNum >
95 < 94  =      = >
96 < 95  real   10.5 >
97 < 96  ;      ; >
98 < 97  char   char >
99 < 98  id     c >
100 < 99  =      = >
101 < 100 '      ' >
102 < 101 id     a >
103 < 102 '      ' >
104 < 103 ;      ; >
105 < 104 const  const >
106 < 105 int    int >
107 < 106 id     hisNum >
108 < 107 =      = >
109 < 108 SLP    ( >
110 < 109 int    int >
111 < 110 SRP    ) >
112 < 111 id     c >

```

```

113 < 112 ; ; >
114 < 113 int int >
115 < 114 id herNum >
116 < 115 = = >
117 < 116 id myNum >
118 < 117 - - >
119 < 118 id hisNum >
120 < 119 ; ; >
121 < 120 bool bool >
122 < 121 id flag >
123 < 122 = = >
124 < 123 true true >
125 < 124 ; ; >
126 < 125 do do >
127 < 126 LP { >
128 < 127 id myNum >
129 < 128 %= %= >
130 < 129 integer 22 >
131 < 130 ; ; >
132 < 131 RP } >
133 < 132 while while >
134 < 133 SLP ( >
135 < 134 ! ! >
136 < 135 id flag >
137 < 136 SRP ) >
138 < 137 ; ; >
139 < 138 for for >
140 < 139 SLP ( >
141 < 140 int int >
142 < 141 id i >
143 < 142 = = >
144 < 143 integer 0 >
145 < 144 ; ; >
146 < 145 id i >
147 < 146 lt < >
148 < 147 integer 100 >
149 < 148 ; ; >
150 < 149 id i >
151 < 150 ++ ++ >
152 < 151 SRP ) >
153 < 152 LP { >
154 < 153 if if >
155 < 154 SLP ( >
156 < 155 id herNum >
157 < 156 eq == >
158 < 157 id myNum >
159 < 158 SRP ) >
160 < 159 LP { >
161 < 160 break break >
162 < 161 ; ; >
163 < 162 RP } >
164 < 163 else else >
165 < 164 if if >

```



```

166 < 165 SLP ( >
167 < 166 id Helper >
168 < 167 : : >
169 < 168 : : >
170 < 169 id greater_or_equal_than >
171 < 170 SLP ( >
172 < 171 id herNum >
173 < 172 , , >
174 < 173 id myNum >
175 < 174 SRP ) >
176 < 175 SRP ) >
177 < 176 LP { >
178 < 177 id herNum >
179 < 178 -- -- >
180 < 179 ; ; >
181 < 180 continue continue >
182 < 181 ; ; >
183 < 182 RP } >
184 < 183 else else >
185 < 184 LP { >
186 < 185 id herNum >
187 < 186 ++ ++ >
188 < 187 ; ; >
189 < 188 continue continue >
190 < 189 ; ; >
191 < 190 RP } >
192 < 191 id myNum >
193 < 192 = = >
194 < 193 SLP ( >
195 < 194 id myNum >
196 < 195 ne != >
197 < 196 id hisNum >
198 < 197 ? ? >
199 < 198 ++ ++ >
200 < 199 id myNum >
201 < 200 : : >
202 < 201 -- -- >
203 < 202 id myNum >
204 < 203 SRP ) >
205 < 204 ; ; >
206 < 205 RP } >
207 < 206 switch switch >
208 < 207 SLP ( >
209 < 208 id herNum >
210 < 209 SRP ) >
211 < 210 LP { >
212 < 211 case case >
213 < 212 integer 10 >
214 < 213 : : >
215 < 214 id herNum >
216 < 215 *= *= >
217 < 216 id myNum >
218 < 217 ; ; >

```

```

219 < 218 break break >
220 < 219 ; ; >
221 < 220 case case >
222 < 221 integer 20 >
223 < 222 : : >
224 < 223 id herNum >
225 < 224 = = >
226 < 225 id Helper >
227 < 226 : : >
228 < 227 : : >
229 < 228 id tool >
230 < 229 . . >
231 < 230 id RED >
232 < 231 ; ; >
233 < 232 default default >
234 < 233 : : >
235 < 234 id herNum >
236 < 235 -= -= >
237 < 236 id myNum >
238 < 237 ; ; >
239 < 238 RP } >
240 < 239 return return >
241 < 240 integer 0 >
242 < 241 ; ; >
243 < 242 RP } >

```

调试信息

我们设置了一些辅助函数，用于打印构建过程中的信息，如状态转换表。

下图是DFA的状态转换表，这个DFA有180个状态。`index` 为状态编号，`action` 为状态的行为（如180号状态接受串 `>>=`），而后的集合表示与其对应的NFA中的状态集合。下图中可以发现根据算法转换正则表达式得到的NFA的状态数量高达1000多个！集合之后的“字母表-状态号”对表示了状态转换的行为。可以发现很多输入都将转换到89状态，这是一个死状态，表明非法的输入。

```
162 1146 1165 1171 1150 1178 1179 1181 1183 } !: 89 " : 89 %: 89 &: 89 ' : 89 (: 89 ): 89 *: 89 +: 89 ,: 89 -: 89 .: 89
/: 89 0: 169 1: 170 2: 171 3: 172 4: 173 5: 174 6: 175 7: 176 8: 177 9: 178 :: 89 :: 89 <: 89 =: 89 >: 89 ? : 89 A: 89 B:
89 C: 89 D: 89 E: 89 F: 89 G: 89 H: 89 I: 89 J: 89 K: 89 L: 89 M: 89 N: 89 O: 89 P: 89 Q: 89 R: 89 S: 89 T: 89 U: 89 V:
89 W: 89 X: 89 Y: 89 Z: 89 [: 89 ]: 89 _ : 89 a: 89 b: 89 c: 89 d: 89 e: 89 f: 89 g: 89 h: 89 i: 89 j: 89 k: 89 l: 89 m:
89 n: 89 o: 89 p: 89 q: 89 r: 89 s: 89 t: 89 u: 89 v: 89 w: 89 x: 89 y: 89 z: 89 { : 89 | : 89 } : 89 ~ : 89
index: 176 action:real { 1166 1170 1164 1168 1172 1148 1152 1144 1154 1158 1142 1160 1156 1174 1169 1175 1176 1
162 1146 1171 1150 1178 1179 1181 1183 } !: 89 " : 89 %: 89 &: 89 ' : 89 (: 89 ): 89 *: 89 +: 89 ,: 89 -: 89 .: 89
/: 89 0: 169 1: 170 2: 171 3: 172 4: 173 5: 174 6: 175 7: 176 8: 177 9: 178 :: 89 :: 89 <: 89 =: 89 >: 89 ? : 89 A: 89 B:
89 C: 89 D: 89 E: 89 F: 89 G: 89 H: 89 I: 89 J: 89 K: 89 L: 89 M: 89 N: 89 O: 89 P: 89 Q: 89 R: 89 S: 89 T: 89 U: 89 V:
89 W: 89 X: 89 Y: 89 Z: 89 [: 89 ]: 89 _ : 89 a: 89 b: 89 c: 89 d: 89 e: 89 f: 89 g: 89 h: 89 i: 89 j: 89 k: 89 l: 89 m:
89 n: 89 o: 89 p: 89 q: 89 r: 89 s: 89 t: 89 u: 89 v: 89 w: 89 x: 89 y: 89 z: 89 { : 89 | : 89 } : 89 ~ : 89
index: 177 action:real { 1166 1170 1164 1168 1172 1148 1152 1144 1154 1158 1142 1160 1156 1173 1174 1175 1176 1
162 1146 1150 1178 1179 1181 1183 } !: 89 " : 89 %: 89 &: 89 ' : 89 (: 89 ): 89 *: 89 +: 89 ,: 89 -: 89 .: 89 /: 89 0:
169 1: 170 2: 171 3: 172 4: 173 5: 174 6: 175 7: 176 8: 177 9: 178 :: 89 :: 89 <: 89 =: 89 >: 89 ? : 89 A: 89 B: 89 C: 8
9 D: 89 E: 89 F: 89 G: 89 H: 89 I: 89 J: 89 K: 89 L: 89 M: 89 N: 89 O: 89 P: 89 Q: 89 R: 89 S: 89 T: 89 U: 89 V: 89 W: 8
9 X: 89 Y: 89 Z: 89 [: 89 ]: 89 _ : 89 a: 89 b: 89 c: 89 d: 89 e: 89 f: 89 g: 89 h: 89 i: 89 j: 89 k: 89 l: 89 m: 89 n: 8
9 o: 89 p: 89 q: 89 r: 89 s: 89 t: 89 u: 89 v: 89 w: 89 x: 89 y: 89 z: 89 { : 89 | : 89 } : 89 ~ : 89
index: 178 action:real { 1166 1170 1164 1168 1172 1148 1152 1144 1154 1158 1142 1160 1156 1174 1176 1162 1146 1
150 1178 1179 1177 1181 1183 } !: 89 " : 89 %: 89 &: 89 ' : 89 (: 89 ): 89 *: 89 +: 89 ,: 89 -: 89 .: 89 /: 89 0: 169 1:
170 2: 171 3: 172 4: 173 5: 174 6: 175 7: 176 8: 177 9: 178 :: 89 :: 89 <: 89 =: 89 >: 89 ? : 89 A: 89 B: 89 C: 89 D: 89
E: 89 F: 89 G: 89 H: 89 I: 89 J: 89 K: 89 L: 89 M: 89 N: 89 O: 89 P: 89 Q: 89 R: 89 S: 89 T: 89 U: 89 V: 89 W: 89 X: 89
Y: 89 Z: 89 [: 89 ]: 89 _ : 89 a: 89 b: 89 c: 89 d: 89 e: 89 f: 89 g: 89 h: 89 i: 89 j: 89 k: 89 l: 89 m: 89 n: 89 o: 89
p: 89 q: 89 r: 89 s: 89 t: 89 u: 89 v: 89 w: 89 x: 89 y: 89 z: 89 { : 89 | : 89 } : 89 ~ : 89
index: 179 action:<=<= { 43 } !: 89 " : 89 %: 89 &: 89 ' : 89 (: 89 ): 89 *: 89 +: 89 ,: 89 -: 89 .: 89 /: 89 0:
89 1: 89 2: 89 3: 89 4: 89 5: 89 6: 89 7: 89 8: 89 9: 89 :: 89 :: 89 <: 89 =: 89 >: 89 ? : 89 A: 89 B: 89 C: 89 D: 89 E:
89 F: 89 G: 89 H: 89 I: 89 J: 89 K: 89 L: 89 M: 89 N: 89 O: 89 P: 89 Q: 89 R: 89 S: 89 T: 89 U: 89 V: 89 W: 89 X: 89 Y:
89 Z: 89 [: 89 ]: 89 _ : 89 a: 89 b: 89 c: 89 d: 89 e: 89 f: 89 g: 89 h: 89 i: 89 j: 89 k: 89 l: 89 m: 89 n: 89 o: 89 p:
89 q: 89 r: 89 s: 89 t: 89 u: 89 v: 89 w: 89 x: 89 y: 89 z: 89 { : 89 | : 89 } : 89 ~ : 89
index: 180 action:>=>= { 51 } !: 89 " : 89 %: 89 &: 89 ' : 89 (: 89 ): 89 *: 89 +: 89 ,: 89 -: 89 .: 89 /: 89 0:
89 1: 89 2: 89 3: 89 4: 89 5: 89 6: 89 7: 89 8: 89 9: 89 :: 89 :: 89 <: 89 =: 89 >: 89 ? : 89 A: 89 B: 89 C: 89 D: 89 E:
89 F: 89 G: 89 H: 89 I: 89 J: 89 K: 89 L: 89 M: 89 N: 89 O: 89 P: 89 Q: 89 R: 89 S: 89 T: 89 U: 89 V: 89 W: 89 X: 89 Y:
89 Z: 89 [: 89 ]: 89 _ : 89 a: 89 b: 89 c: 89 d: 89 e: 89 f: 89 g: 89 h: 89 i: 89 j: 89 k: 89 l: 89 m: 89 n: 89 o: 89 p:
89 q: 89 r: 89 s: 89 t: 89 u: 89 v: 89 w: 89 x: 89 y: 89 z: 89 { : 89 | : 89 } : 89 ~ : 89
```

类似地，我们可以打印最小DFA的状态转换表。最小DFA只有46个状态，状态数被大大减少。然而，这一结论已经证明了一开始我们的假设，即如果手动构建NFA，然后再转为最小DFA，这个过程对人而言是相当困难甚至无法实现的。因此，我们不给出最终DFA的图结构形式的图示，仅仅给出其状态转换表，请谅解。

```
=: 28 >: 28 ? : 28 A: 28 B: 28 C: 28 D: 28 E: 28 F: 28 G: 28 H: 28 I: 28 J: 28 K: 28 L: 28
M: 28 N: 28 O: 28 P: 28 Q: 28 R: 28 S: 28 T: 28 U: 28 V: 28 W: 28 X: 28 Y: 28 Z: 28 [: 28
]: 28 _ : 28 a: 28 b: 28 c: 28 d: 28 e: 28 f: 28 g: 28 h: 28 i: 28 j: 28 k: 28 l: 28 m: 28
n: 28 o: 28 p: 28 q: 28 r: 28 s: 28 t: 28 u: 28 v: 28 w: 28 x: 28 y: 28 z: 28 { : 28 | : 28
}: 28 ~ : 28
group: 43 action: >> !: 28 " : 28 %: 28 &: 28 ' : 28 (: 28 ): 28 *: 28 +: 28 ,: 28 -: 28
.: 28 /: 28 0: 28 1: 28 2: 28 3: 28 4: 28 5: 28 6: 28 7: 28 8: 28 9: 28 :: 28 :: 28 <: 28
=: 46 >: 28 ? : 28 A: 28 B: 28 C: 28 D: 28 E: 28 F: 28 G: 28 H: 28 I: 28 J: 28 K: 28 L: 28
M: 28 N: 28 O: 28 P: 28 Q: 28 R: 28 S: 28 T: 28 U: 28 V: 28 W: 28 X: 28 Y: 28 Z: 28 [: 28
]: 28 _ : 28 a: 28 b: 28 c: 28 d: 28 e: 28 f: 28 g: 28 h: 28 i: 28 j: 28 k: 28 l: 28 m: 28
n: 28 o: 28 p: 28 q: 28 r: 28 s: 28 t: 28 u: 28 v: 28 w: 28 x: 28 y: 28 z: 28 { : 28 | : 28
}: 28 ~ : 28
group: 44 action: || !: 28 " : 28 %: 28 &: 28 ' : 28 (: 28 ): 28 *: 28 +: 28 ,: 28 -: 28
.: 28 /: 28 0: 28 1: 28 2: 28 3: 28 4: 28 5: 28 6: 28 7: 28 8: 28 9: 28 :: 28 :: 28 <: 28
=: 28 >: 28 ? : 28 A: 28 B: 28 C: 28 D: 28 E: 28 F: 28 G: 28 H: 28 I: 28 J: 28 K: 28 L: 28
M: 28 N: 28 O: 28 P: 28 Q: 28 R: 28 S: 28 T: 28 U: 28 V: 28 W: 28 X: 28 Y: 28 Z: 28 [: 28
]: 28 _ : 28 a: 28 b: 28 c: 28 d: 28 e: 28 f: 28 g: 28 h: 28 i: 28 j: 28 k: 28 l: 28 m: 28
n: 28 o: 28 p: 28 q: 28 r: 28 s: 28 t: 28 u: 28 v: 28 w: 28 x: 28 y: 28 z: 28 { : 28 | : 28
}: 28 ~ : 28
group: 45 action: <<= !: 28 " : 28 %: 28 &: 28 ' : 28 (: 28 ): 28 *: 28 +: 28 ,: 28 -: 28
.: 28 /: 28 0: 28 1: 28 2: 28 3: 28 4: 28 5: 28 6: 28 7: 28 8: 28 9: 28 :: 28 :: 28 <: 28
=: 28 >: 28 ? : 28 A: 28 B: 28 C: 28 D: 28 E: 28 F: 28 G: 28 H: 28 I: 28 J: 28 K: 28 L: 28
M: 28 N: 28 O: 28 P: 28 Q: 28 R: 28 S: 28 T: 28 U: 28 V: 28 W: 28 X: 28 Y: 28 Z: 28 [: 28
]: 28 _ : 28 a: 28 b: 28 c: 28 d: 28 e: 28 f: 28 g: 28 h: 28 i: 28 j: 28 k: 28 l: 28 m: 28
n: 28 o: 28 p: 28 q: 28 r: 28 s: 28 t: 28 u: 28 v: 28 w: 28 x: 28 y: 28 z: 28 { : 28 | : 28
}: 28 ~ : 28
group: 46 action: >>= !: 28 " : 28 %: 28 &: 28 ' : 28 (: 28 ): 28 *: 28 +: 28 ,: 28 -: 28
.: 28 /: 28 0: 28 1: 28 2: 28 3: 28 4: 28 5: 28 6: 28 7: 28 8: 28 9: 28 :: 28 :: 28 <: 28
=: 28 >: 28 ? : 28 A: 28 B: 28 C: 28 D: 28 E: 28 F: 28 G: 28 H: 28 I: 28 J: 28 K: 28 L: 28
M: 28 N: 28 O: 28 P: 28 Q: 28 R: 28 S: 28 T: 28 U: 28 V: 28 W: 28 X: 28 Y: 28 Z: 28 [: 28
]: 28 _ : 28 a: 28 b: 28 c: 28 d: 28 e: 28 f: 28 g: 28 h: 28 i: 28 j: 28 k: 28 l: 28 m: 28
n: 28 o: 28 p: 28 q: 28 r: 28 s: 28 t: 28 u: 28 v: 28 w: 28 x: 28 y: 28 z: 28 { : 28 | : 28
}: 28 ~ : 28
```

状态转换表

状态转换表较大，可以在附录中查看。

如何执行我们的Lexer程序

我们的程序不难直接双击 `Lexer.exe` 运行，这是一个控制台程序，需要打开 `cmd` 传入参数后才可执行。我们在 `Lexer.exe` 所在目录下打开 `cmd`，之后按如下规则输入指令，其中`debug`是可选项：

```
1 | Lexer.exe <path-to-code> [debug]
```

例如：

```
C:\Users\Serica\Archive\大三上\编译原理\MiniC_Compiler\Lexer\x64\Release>Lexer.exe C:\Users\Serica\Archive\大三上\编译原理\MiniC_Compiler\Lexer\test\code.cpp
```

第一个参数为要进行词法分析的程序源码路径，由于时间有限，我们还没有将文件系统的更多操作考虑在内，因此请输入绝对路径（可以直接把代码文件拖到 `cmd` 中，`cmd` 会生成该文件的绝对路径）。同样，程序的词法分析结果不会写入到文件里，而仅仅是打印在控制台上。程序执行时会先在控制台上打印要进行词法分析的源代码，随后会打印DFA建立时各个步骤花费的时间。**在我们的电脑上耗时间如下，可见NFA转DFA的过程耗时最多，原因是显然的。**因为根据之前的结果，NFA的状态高达上千个，而要把它们转换为只有一百多个状态的DFA，是十分费力的。

```
1 | Building Min DFA now... It can take minutes, please be patient and wait...
2 | ...Regular expressions loaded, start to build NFA from them.
3 | ...NFA already built, takes 0.001s, now start to build DFA from the NFA.
4 | ...DFA already built, takes 38.201s, now start to minimize the DFA.
5 | ...Min DFA already built, takes 0.001s, ready to do lexical analysis.
```

打印调试信息

在路径之后增加参数 `debug` 即可。

```
C:\Users\Serica\Archive\大三上\编译原理\MiniC_Compiler\Lexer\x64\Release>Lexer.exe C:\Users\Serica\Archive\大三上\编译原理\MiniC_Compiler\Lexer\test\code.cpp debug
```

Future Work以及软件工程相关思想

Future Work（会体现在下学期的课程设计中）

1. 完善文件系统，即输入输出到文件
2. 状态表也将以文件形式存储，这样只需要生成一次状态转换表，以后使用时只需要读入状态转换表即可，无须重复生成。
3. 尝试优化性能（目前的瓶颈在于NFA状态过多，导致生成DFA时过慢）
4. 增强鲁棒性和报错能力
5. 引入更多数据结构，完善代码结构组织（如目前的action应该用枚举类而不是字符串表示，又如token是否应该进行封装等）
6. 将符号表抽离出词法分析器中。符号表应该是词法分析器和语法分析器共同维护的。

软件工程思想

词法分析的过程是经过一系列步骤的，各个步骤间要实现松耦合，而单个步骤内的实现应该是高内聚的。我们在写代码时为各个类编写了测试文件，在源代码中以 `Test` 开头，如 `TestRegex.cpp`、`TestNFA.cpp` 等等。在进行下一步开发时，总是要保证代码可以通过这些测试。由于时间有限，程序的架构并不是十分合理有序，且暂时没有绘制类图，之后会花时间重构。

对于具体实现部分，我们使用了一些数据结构，如转换表使用 `map<pair<int, int>, int> trans` 这一形式，里面的 `pair<int, int>` 表示采取的一个 `step`，前一个表示当前状态号，后一个表示读入的符号，而 `int` 则是这一步骤将转换到的状态。因此如果状态1有一条值为 `a` 的边指向状态2，那么有 `trans[pair<int, int>(1, 'a')] = 2`。

在构建时，我们还要单独构建各个状态的行为，体现在 `vector<string> actions` 中，建立的顺序是有序的，所以和状态号是一一对应的。当然，我们做了许多思考和工作保证 `actions` 里的下标和状态号一一对应，为了使得以后开发方便，我们可以把这些都绑在一起，使得今后不需要考虑顺序信息。

附录：最小DFA的状态转换表

group: 0 action: start !: 1 ": 2 %: 3 &: 4 ': 5 (: 6): 7 *: 8 +: 9 ,: 10 -: 11 .: 12 /: 13 0: 14 1: 14 2: 14 3: 14 4: 14 5: 14 6: 14 7: 14 8: 14 9: 14 :: 15 ;: 16 <: 17 =: 18 >: 19 ?: 20 A: 21 B: 21 C: 21 D: 21 E: 21 F: 21 G: 21 H: 21 I: 21 J: 21 K: 21 L: 21 M: 21 N: 21 O: 21 P: 21 Q: 21 R: 21 S: 21 T: 21 U: 21 V: 21 W: 21 X: 21 Y: 21 Z: 21 [: 22]: 23 _: 21 a: 21 b: 21 c: 21 d: 21 e: 21 f: 21 g: 21 h: 21 i: 21 j: 21 k: 21 l: 21 m: 21 n: 21 o: 21 p: 21 q: 21 r: 21 s: 21 t: 21 u: 21 v: 21 w: 21 x: 21 y: 21 z: 21 { 24 |: 25 } 26 ~: 27

\n: 1 action: !: 28 ": 28 %: 28 &: 28 ': 28 (: 28): 28 *: 28 +: 28 ,: 28 -: 28 .: 28 /: 28 0: 28 1: 28 2: 28 3: 28 4: 28 5: 28 6: 28 7: 28 8: 28 9: 28 :: 28 ;: 28 <: 28 =: 29 >: 28 ?: 28 A: 28 B: 28 C: 28 D: 28 E: 28 F: 28 G: 28 H: 28 I: 28 J: 28 K: 28 L: 28 M: 28 N: 28 O: 28 P: 28 Q: 28 R: 28 S: 28 T: 28 U: 28 V: 28 W: 28 X: 28 Y: 28 Z: 28 [: 28]: 28 _: 28 a: 28 b: 28 c: 28 d: 28 e: 28 f: 28 g: 28 h: 28 i: 28 j: 28 k: 28 l: 28 m: 28 n: 28 o: 28 p: 28 q: 28 r: 28 s: 28 t: 28 u: 28 v: 28 w: 28 x: 28 y: 28 z: 28 { 28 |: 28 } 28 ~: 28

group: 2 action: " !: 28 ": 28 %: 28 &: 28 ': 28 (: 28): 28 *: 28 +: 28 ,: 28 -: 28 .: 28 /: 28 0: 28 1: 28 2: 28 3: 28 4: 28 5: 28 6: 28 7: 28 8: 28 9: 28 :: 28 ;: 28 <: 28 =: 28 >: 28 ?: 28 A: 28 B: 28 C: 28 D: 28 E: 28 F: 28 G: 28 H: 28 I: 28 J: 28 K: 28 L: 28 M: 28 N: 28 O: 28 P: 28 Q: 28 R: 28 S: 28 T: 28 U: 28 V: 28 W: 28 X: 28 Y: 28 Z: 28 [: 28]: 28 _: 28 a: 28 b: 28 c: 28 d: 28 e: 28 f: 28 g: 28 h: 28 i: 28 j: 28 k: 28 l: 28 m: 28 n: 28 o: 28 p: 28 q: 28 r: 28 s: 28 t: 28 u: 28 v: 28 w: 28 x: 28 y: 28 z: 28 { 28 |: 28 } 28 ~: 28

group: 3 action: % !: 28 ": 28 %: 28 &: 28 ': 28 (: 28): 28 *: 28 +: 28 ,: 28 -: 28 .: 28 /: 28 0: 28 1: 28 2: 28 3: 28 4: 28 5: 28 6: 28 7: 28 8: 28 9: 28 :: 28 ;: 28 <: 28 =: 30 >: 28 ?: 28 A: 28 B: 28 C: 28 D: 28 E: 28 F: 28 G: 28 H: 28 I: 28 J: 28 K: 28 L: 28 M: 28 N: 28 O: 28 P: 28 Q: 28 R: 28 S: 28 T: 28 U: 28 V: 28 W: 28 X: 28 Y: 28 Z: 28 [: 28]: 28 _: 28 a: 28 b: 28 c: 28 d: 28 e: 28 f: 28 g: 28 h: 28 i: 28 j: 28 k: 28 l: 28 m: 28 n: 28 o: 28 p: 28 q: 28 r: 28 s: 28 t: 28 u: 28 v: 28 w: 28 x: 28 y: 28 z: 28 { 28 |: 28 } 28 ~: 28

group: 4 action: & !: 28 ": 28 %: 28 &: 31 ': 28 (: 28): 28 *: 28 +: 28 ,: 28 -: 28 .: 28 /: 28 0: 28 1: 28 2: 28 3: 28 4: 28 5: 28 6: 28 7: 28 8: 28 9: 28 :: 28 ;: 28 <: 28 =: 28 >: 28 ?: 28 A: 28 B: 28 C: 28 D: 28 E: 28 F: 28 G: 28 H: 28 I: 28 J: 28 K: 28 L: 28 M: 28 N: 28 O: 28 P: 28 Q: 28 R: 28 S: 28 T: 28 U: 28 V: 28 W: 28 X: 28 Y: 28 Z: 28 [: 28]: 28 _: 28 a: 28 b: 28 c: 28 d: 28 e: 28 f: 28 g: 28 h: 28 i: 28 j: 28 k: 28 l: 28 m: 28 n: 28 o: 28 p: 28 q: 28 r: 28 s: 28 t: 28 u: 28 v: 28 w: 28 x: 28 y: 28 z: 28 { 28 |: 28 } 28 ~: 28

group: 5 action: ' !: 28 ": 28 %: 28 &: 28 ': 28 (: 28): 28 *: 28 +: 28 ,: 28 -: 28 .: 28 /: 28 0: 28 1: 28 2: 28 3: 28 4: 28 5: 28 6: 28 7: 28 8: 28 9: 28 :: 28 ;: 28 <: 28 =: 28 >: 28 ?: 28 A: 28 B: 28 C: 28 D: 28 E: 28 F: 28 G: 28 H: 28 I: 28 J: 28 K: 28 L: 28 M: 28 N: 28 O: 28 P: 28 Q: 28 R: 28 S: 28 T: 28 U: 28 V: 28 W: 28 X: 28 Y: 28 Z: 28 [: 28]: 28 _: 28 a: 28 b: 28 c: 28 d: 28 e: 28 f: 28 g: 28 h: 28 i: 28 j: 28 k: 28 l: 28 m: 28 n: 28 o: 28 p: 28 q: 28 r: 28 s: 28 t: 28 u: 28 v: 28 w: 28 x: 28 y: 28 z: 28 { 28 |: 28 } 28 ~: 28

group: 6 action: SLP !: 28 ": 28 %: 28 &: 28 ': 28 (: 28): 28 *: 28 +: 28 ,: 28 -: 28 .: 28 /: 28 0: 28 1: 28 2: 28 3: 28 4: 28 5: 28 6: 28 7: 28 8: 28 9: 28 :: 28 ;: 28 <: 28 =: 28 >: 28 ?: 28 A: 28 B: 28 C: 28 D: 28 E: 28 F: 28 G: 28 H: 28 I: 28 J: 28 K: 28 L: 28 M: 28 N: 28 O: 28 P: 28 Q: 28 R: 28 S: 28 T: 28 U: 28 V: 28 W: 28 X: 28 Y: 28 Z: 28 [: 28]: 28 _: 28 a: 28 b: 28 c: 28 d: 28 e: 28 f: 28 g: 28 h: 28 i: 28 j: 28 k: 28 l: 28 m: 28 n: 28 o: 28 p: 28 q: 28 r: 28 s: 28 t:

group: 7 action: SRP !: 28 ": 28 #: 28 &: 28 ': 28 (: 28): 28 *: 28 +: 28 ,: 28 -: 28 .: 28 /: 28 0: 28 1: 28 2: 28 3: 28 4: 28 5: 28 6: 28 7: 28 8: 28 9: 28 :: 28 ;: 28 <: 28 =: 28 >: 28 ?: 28 A: 28 B: 28 C: 28 D: 28 E: 28 F: 28 G: 28 H: 28 I: 28 J: 28 K: 28 L: 28 M: 28 N: 28 O: 28 P: 28 Q: 28 R: 28 S: 28 T: 28 U: 28 V: 28 W: 28 X: 28 Y: 28 Z: 28 [: 28]: 28 _: 28 a: 28 b: 28 c: 28 d: 28 e: 28 f: 28 g: 28 h: 28 i: 28 j: 28 k: 28 l: 28 m: 28 n: 28 o: 28 p: 28 q: 28 r: 28 s: 28 t: 28 u: 28 v: 28 w: 28 x: 28 y: 28 z: 28 {: 28 |: 28 }: 28 ~: 28

group: 9 action: + !: 28 ": 28 %: 28 &: 28 ': 28 (: 28): 28 *: 28 +: 33 ,: 28 -: 28 .: 28 /: 28 0: 28 1: 28 2: 28 3: 28 4: 28 5: 28 6: 28 7: 28 8: 28 9: 28 :: 28 ;: 28 <: 28 =: 34 >: 28 ?: 28 A: 28 B: 28 C: 28 D: 28 E: 28 F: 28 G: 28 H: 28 I: 28 J: 28 K: 28 L: 28 M: 28 N: 28 O: 28 P: 28 Q: 28 R: 28 S: 28 T: 28 U: 28 V: 28 W: 28 X: 28 Y: 28 Z: 28 [: 28]: 28 _: 28 a: 28 b: 28 c: 28 d: 28 e: 28 f: 28 g: 28 h: 28 i: 28 j: 28 k: 28 l: 28 m: 28 n: 28 o: 28 p: 28 q: 28 r: 28 s: 28 t: 28 u: 28 v: 28 w: 28 x: 28 y: 28 z: 28 {: 28 |: 28 }: 28 ~: 28

group: 11 action: - !: 28 " : 28 %: 28 &: 28 ' : 28 (: 28) : 28 *: 28 +: 28 ,: 28 -: 35 .: 28 /: 28 0: 28 1: 28 2: 28 3: 28 4: 28 5: 28 6: 28 7: 28 8: 28 9: 28 :: 28 ;: 28 <: 28 =: 36 >: 28 ?: 28 A: 28 B: 28 C: 28 D: 28 E: 28 F: 28 G: 28 H: 28 I: 28 J: 28 K: 28 L: 28 M: 28 N: 28 O: 28 P: 28 Q: 28 R: 28 S: 28 T: 28 U: 28 V: 28 W: 28 X: 28 Y: 28 Z: 28 [: 28]: 28 _: 28 a: 28 b: 28 c: 28 d: 28 e: 28 f: 28 g: 28 h: 28 i: 28 j: 28 k: 28 l: 28 m: 28 n: 28 o: 28 p: 28 q: 28 r: 28 s: 28 t: 28 u: 28 v: 28 w: 28 x: 28 y: 28 z: 28 { : 28 | : 28 } : 28 ~: 28

group: 13 action: / !: 28 ": 28 %: 28 &: 28 ': 28 (: 28): 28 *: 28 +: 28 ,: 28 -: 28 .: 28 /: 28 0: 28 1: 28 2: 28 3: 28 4: 28 5: 28 6: 28 7: 28 8: 28 9: 28 :: 28 ;: 28 <: 28 =: 37 >: 28 ? : 28 A: 28 B: 28 C: 28 D: 28 E: 28 F: 28 G: 28 H: 28 I: 28 J: 28 K: 28 L: 28 M: 28 N: 28 O: 28 P: 28 Q: 28 R: 28 S: 28 T: 28 U: 28 V: 28 W: 28 X: 28 Y: 28 Z: 28 [: 28]: 28 _ : 28 a: 28 b: 28 c: 28 d: 28 e: 28 f: 28 g: 28 h: 28 i: 28 j: 28 k: 28 l: 28 m: 28 n: 28 o: 28 p: 28 q: 28 r: 28 s: 28 t: 28 u: 28 v: 28 w: 28 x: 28 y: 28 z: 28 { : 28 | : 28 } : 28 ~: 28

group: 14 action: integer !: 28 ": 28 %: 28 &: 28 ': 28 (: 28): 28 *: 28 +: 28 ,: 28 -: 28 .: 38 /: 28 0: 14 1: 14 2: 14 3: 14 4: 14 5: 14 6: 14 7: 14 8: 14 9: 14 :: 28 ;: 28 <: 28 =: 28 >: 28 ?: 28 A: 28 B: 28 C: 28 D: 28 E: 28 F: 28 G: 28 H: 28 I: 28 J: 28 K: 28 L: 28 M: 28 N: 28 O: 28 P: 28 Q: 28 R: 28 S: 28 T: 28 U: 28 V: 28 W: 28 X: 28 Y: 28 Z: 28 [: 28]: 28 _: 28 a: 28 b: 28 c: 28 d: 28 e: 28 f: 28 g: 28 h: 28 i: 28 j: 28 k: 28 l: 28 m: 28 n: 28 o: 28 p: 28 q: 28 r: 28 s: 28 t: 28 u: 28 v: 28 w: 28 x: 28 y: 28 z: 28 {: 28 |: 28 }: 28 ~: 28

group: 41 action: eq !: 28 ": 28 %: 28 &: 28 ': 28 (: 28): 28 *: 28 +: 28 ,: 28 -: 28 .: 28 /: 28 0: 28 1: 28 2: 28 3: 28 4: 28 5: 28 6: 28 7: 28 8: 28 9: 28 :: 28 ;: 28 <: 28 =: 28 >: 28 ?: 28 A: 28 B: 28 C: 28 D: 28 E: 28 F: 28 G: 28 H: 28 I: 28 J: 28 K: 28 L: 28 M: 28 N: 28 O: 28 P: 28 Q: 28 R: 28 S: 28 T: 28 U: 28 V: 28 W: 28 X: 28 Y: 28 Z: 28 [: 28]: 28 _: 28 a: 28 b: 28 c: 28 d: 28 e: 28 f: 28 g: 28 h: 28 i: 28 j: 28 k: 28 l: 28 m: 28 n: 28 o: 28 p: 28 q: 28 r: 28 s: 28 t: 28 u: 28 v: 28 w: 28 x: 28 y: 28 z: 28 {: 28 |: 28 }: 28 ~: 28

group: 43 action: >> !: 28 "; 28 %: 28 &; 28 ': 28 (: 28): 28 *: 28 +; 28 ,; 28 -: 28 .: 28 /: 28 0: 28 1: 28 2: 28 3: 28 4: 28 5: 28 6: 28 7: 28 8: 28 9: 28 :: 28 ;; 28 <; 28 =; 46 >; 28 ?; 28 A: 28 B: 28 C: 28 D: 28 E: 28 F: 28 G: 28 H: 28 I: 28 J: 28 K: 28 L: 28 M: 28 N: 28 O: 28 P: 28 Q: 28 R: 28 S: 28 T: 28 U: 28 V: 28 W: 28 X: 28 Y: 28 Z: 28 [: 28]; 28 _; 28 a: 28 b: 28 c: 28 d: 28 e: 28 f: 28 g: 28 h: 28 i: 28 j: 28 k: 28 l: 28 m: 28 n: 28 o: 28 p: 28 q: 28 r: 28 s: 28 t: 28 u: 28 v: 28 w: 28 x: 28 y: 28 z: 28 {; 28 |: 28 }; 28 ~; 28

group: 45 action: <= !: 28 ": 28 %: 28 &: 28 ': 28 (: 28): 28 *: 28 +: 28 ,: 28 -: 28 .: 28 /: 28 0: 28 1: 28 2: 28 3: 28 4: 28 5: 28 6: 28 7: 28 8: 28 9: 28 :: 28 ;: 28 <: 28 =: 28 >: 28 ?: 28 A: 28 B: 28 C: 28 D: 28 E: 28 F: 28 G: 28 H: 28 I: 28 J: 28 K: 28 L: 28 M: 28 N: 28 O: 28 P: 28 Q: 28 R: 28 S: 28 T: 28 U: 28 V: 28 W: 28 X: 28 Y: 28 Z: 28 [: 28]: 28 _: 28 a: 28 b: 28 c: 28 d: 28 e: 28 f: 28 g: 28 h: 28 i: 28 j: 28 k: 28 l: 28 m: 28 n: 28 o: 28 p: 28 q: 28 r: 28 s: 28 t: 28 u: 28 v: 28 w: 28 x: 28 y: 28 z: 28 {: 28 |: 28 }: 28 ~: 28

group: 46 action: >>=: 28 "; 28 %: 28 &: 28 ': 28 (: 28): 28 *: 28 +: 28 ,: 28 -: 28 .: 28 /: 28 0: 28 1: 28 2: 28 3: 28 4: 28 5: 28 6: 28 7: 28 8: 28 9: 28 :: 28 ;: 28 <: 28 =: 28 >: 28 ?: 28 A: 28 B: 28 C: 28 D: 28 E: 28 F: 28 G: 28 H: 28 I: 28 J: 28 K: 28 L: 28 M: 28 N: 28 O: 28 P: 28 Q: 28 R: 28 S: 28 T: 28 U: 28 V: 28 W: 28 X: 28 Y: 28 Z: 28 [: 28]: 28 _: 28 a: 28 b: 28 c: 28 d: 28 e: 28 f: 28 g: 28 h: 28 i: 28 j: 28 k: 28 l: 28 m: 28 n: 28 o: 28 p: 28 q: 28 r: 28 s: 28 t: 28 u: 28 v: 28 w: 28 x: 28 y: 28 z: 28 {: 28 |: 28 }: 28 ~: 28