Atititi 为什么需要手写词法解析器

[1.1. 手写的优点：代码可读，对源代码中的各种错误给出友好的提示信息，用户体验高， 1](#_Toc9275)

[1.2. 使用状态表比较简单，dfa比较麻烦 1](#_Toc26270)

[1.3. 然后给了你代码框架（这里以nested case statement 为例）： 1](#_Toc31887)

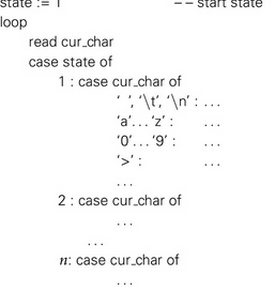
[1.4. 源码实现 2](#_Toc8285)

## 手写的优点：代码可读，对源代码中的各种错误给出友好的提示信息，用户体验高，

产品级别的编译器需要对源代码中的各种错误给出友好的提示信息。而DFA在这方面无法满足需要

## 使用状态表比较简单，dfa比较麻烦

## 然后给了你代码框架（这里以nested case statement 为例）：



然后告诉你最外层case覆盖DFA的所有状态，内层每个case覆盖这个状态的所有转换。特别清晰简单有木有！！！编程时候直接填空就行了！

## 源码实现

**public** List<Token> **getTokens**(String codeStr) {

List<Token> **li** = Lists.*newArrayList*();

code\_char\_arr = codeStr.toCharArray();

**while** (**true**) {

Object **tk**;

**try** {

tk = nextTokens();

} **catch** (TokenEndEx **e**) {

**break**;

}

**if** (tk **instanceof** Token)

li.add((Token) tk);

**else** **if** (tk **instanceof** List)

li.addAll((Collection<? **extends** Token>) tk);

**else**

**throw** **new** RuntimeException("token type err,curchar:" + **this**.cur\_char + ",colidx:" + **this**.gColumn);

}

**return** li;

}

**public** Object **nextTokens**() **throws** TokenEndEx {

// code\_char\_arr = code.toCharArray();

gColumn++;

**if** (gColumn > code\_char\_arr.length - 1)

**throw** **new** TokenEndEx(**new** String(code\_char\_arr));

cur\_char = code\_char\_arr[gColumn];

// cur\_char=cur\_char;

**if** (**this**.curTokenTxt.equals("1598"))

System.***out***.println("dbg");

**if** (**this**.gColumn == 30)

System.***out***.println("dbg");

// get next char,,then change stat

// jude cur char and cur stat...then if or not chage stat

**switch** (cur\_char) {

**case** '(':

**return** BrkStartEvt();

// break;

**case** ')':

**return** brkEndEvt();

**case** '\'':

**return** sQuoEvt();

**case** '\"':

**return** dbQuoEvt();

**case** ':':

**return** colonEvt();

**case** ',':

**return** commaEvt();

**default**:

**return** normalCharEvt();

// break;

}

}

**private** Object **BrkStartEvt**() {

**char** **c** = **this**.cur\_char;

**if** (c == '(' && !**this**.curStat.equals("strStart")) { // && cur stta=ini

List<Token> **li** = Lists.*newArrayList*();

Token **tk** = **new** Token(**this**.curTokenTxt).setType("var");

li.add(tk);

Token **tk2** = **new** Token("(").setType("op");

li.add(tk2);

**this**.curTokenTxt = "";

**this**.curStat = "brkStart";

**return** li;

}

**throw** **new** RuntimeException("BrkStartEvt");

}

作者:: 绰号:老哇的爪子 （ 全名：：Attilax Akbar Al Rapanui 阿提拉克斯 阿克巴 阿尔 拉帕努伊 ）

汉字名：艾提拉（艾龙），   EMAIL:1466519819@qq.com

转载请注明来源： http://blog.csdn.net/attilax

Atiend