中国科学技术大学

实验报告

编译原理实验报告

作者姓名: 康F

学科专业: 信息安全

导师姓名: 李卫海 副教授

完成时间: 二〇一八年四月二十六日

University of Science and Technology of China

A dissertation for compilation principle

Compilation Principle Lab Report

Author: Sheng Kang

Speciality: Information Security

Supervisor: A/Prof. Weihai Li

Finished time: April 26, 2018

中文内容摘要

本文为编译原理大作业实验报告。主要实现了一个简单的类 LaTeX 语言编译器，可以识别上标、下标、积分、求和和空格等符号，并输出一个 HTML 文件，可以在浏览器中看到相应公式。

本文从实验目的，词法，文法，制导和收获总结几个方面总结了本次实验。报告先简述了实验目的，之后通过设计确定有限自动机来处理输入字符串，进行文法分析。通过改写词法，使得可以通过 LL(1) 文法来进行处理，之后采用了自顶向下和自下向上结合的方法来进行语法制导的翻译，因为需要处理的属性中，既有综合属性，也有继承属性。最后将理论所得予以编程实现，并总结了收获。

关键词：编译原理；LaTeX ；词法；文法；制导

Abstract

This article is an experiment report for the compilation principle. A simple LaTeX-like language compiler is mainly implemented that recognizes superscripts, subscripts, integrals, sums, and spaces, and outputs an HTML file, which can be seen in the browser.

This article summarizes the experiment from the aspects of experiment purpose, lexical, grammatical, guidance and summary. The report first outlines the purpose of the experiment, after which the finite automata are designed to process the input string for grammar analysis. By rewriting the lexical word, it can be handled by the LL(1) grammar. Later, the syntax-directed translation is performed using a combination of top-down and bottom-up approaches because the attributes to be processed include both comprehensive and inherited attributes. Finally, the theoretical gains were programmed and summarized.

Key Words: compilation principle; LaTeX; lexical; grammatical; guidance

目　录

第一章　实验目的

第一节　词法

通过实现词法分析器，理解编译器处理输入字符串的过程。

第二节　文法

- 掌握消除左递归的方法
- 掌握 LL(1) 文法成立的基本条件
- 掌握 LL(1) 文法工作的原理

第三节　制导

- 掌握综合属性和继承属性的区别
- 掌握文法分析与制导的区别和联系
- 掌握 L 属性定义的自上而下计算的方法
- 掌握编写和调试复杂程序的方法

第二章　词法

第一节　分析

编译器在词法分析阶段要做的工作为把源程序的字符流翻译成词法记号流。具体到本次实验，需要做的工作是识别如表2.1中十三种不同的输入：

$	_	^	{	}	\int	
\sum	\blank	id	num	(	)	_^

表 2.1　需要识别的字符

为了储存识别结果，将每个符号对应为一个整数，成功识别后存入数组即可。为了输出识别结果，建立一个二维字符串数组，在对应位置储存不同符号的字符串表示。

第二节　程序运行情况

以 test1 和 test4 为例，匹配成功的情况如下：

```
1  test1: $154^{5}$
2  $    num    ^    {    num    }    $
3  $    154    ^    {    5    }    $
4
5  test4: $a^{b^{c^{2}}d}e$
6  $    id    ^    {    id    ^    {    id    ^
       {    num    }    }    id    }    id    $
7  $    a    ^    {    b    ^    {    c    ^
       {    2    }    }    d    }    e    $
```

以 test10 为例，匹配失败的情况如下：

```
1 test10: $a^{b^{c^{2.1}}d}$
2 Unknown token!
3 last matched: $a^{b^{c^{2
```

第三章　文法

第一节　分析

编译器通过词法把字符流翻译为记号流后，需要进一步读取词法分析器提供的记号流，检查它是否能由源语言的文法产生，输出分析树的某种表示。

由于给定的文法含有多个左递归，无法直接使用 LL(1) 文法进行分析，需要先对文法进行转换，消除左递归。之后，先构建分析表，然后采用非递归的预测分析的方法，对记号流进行分析。最后输出分析过程并储存分析结果。

第二节　消除左递归

（1）原文法

原文法见表3.1。

表 3.1　原文法

S →B
B →BB
B → B_^{B}{B}
B → B^{B}
B → B_{B}
B → \int{B}{B}{B}
B → \sum{B}{B}{B}
B → id|num|\blank|(B)

（2）最终文法

对原文法进行修改[1]，消除左递归，增加一个非终结符 D。最终的文法见表3.2。

表 3.2　消除左递归后的文法

S →B
B → D | \int{B}{B}{B}D | \sum{B}{B}{B}D | id D | num D | \blank D | (B)D
D → _^{B}D | ^{B}D | _{B}D | BD | ϵ

第三节　确定符号集

为了构造预测分析表，同时确定该文法是否为 LL(1) 文法，需要计算该文法中所有非终止符的开始符号集合和终止符号集合。因为 D 的产生式包含 ϵ，故需要计算 D 的终止符号集。开始符号集和终止符号集分别见表3.3和表3.4。

表 3.3　开始符号集

first(S) = { $ }
first(B) ={ \int, \sum, id, num, \blank, (}
first(D) ={ \int, \sum, id, num, \blank, (, ϵ , _, ^, _^ }

表 3.4　终止符号集

follow(D) = { $, ϵ , (,), }, \int, \sum, \blank, id, num}

第四节　构造分析表

依据之前修改后的文法及开始符号集和终止符号集，构造 LL(1) 文法的分析表如表3.5：

表 3.5　文法分析表

输入符号	非终结符		
	S	B	D
$	S →B		D → ϵ
_			D →_{B}D
^			D →^{B}D
{			
}			D →ϵ
\int		B → \int{B}{B}{B}D	D →BD
\sum		B → \sum{B}{B}{B}D	D →BD
\blank		B → \blank D	D →BD
id		B → id D	D →BD
num		B → num D	D →BD
(		B → (B)D	D →BD
)			D → ϵ
_^			D →_^{B}{B}D
ϵ			D →ϵ

第五节　代码运行情况

一、文法分析成功

先展示文法分析成功的情况，以 test7 为例：

```
test7: $(a_{4})^{2}$
$     (     id    _     {     num   }     )     ^
   {     num   }     $
$     (     a     _     {     4     }     )     ^
   {     2     }     $

S → $ B $
$ B $        $ ( id _ { num } ) ^ { num } $
B → ( B ) D
$ D ) B (          ( id _ { num } ) ^ { num } $
B → id D
$ D ) D id         id _ { num } ) ^ { num } $
D → _ { B } D
$ D ) D } B { _          _ { num } ) ^ { num } $
$ D ) D } B {      { num } ) ^ { num } $
B → num D
$ D ) D } D num          num } ) ^ { num } $
D → ε
$ D ) D } ε        } ) ^ { num } $
$ D ) D }          } ) ^ { num } $
D → ε
$ D ) ε            ) ^ { num } $
$ D )        ) ^ { num } $
D → ^ { B } D
```

```
$ D } B { ^        ^ { num } $
$ D } B {          { num } $
B → num D
$ D } D num        num } $
D → ε
$ D } ε            } $
$ D }        } $
D → ε
$ ε          $
$            $
Congratulations! Match succeed!
```

二、文法分析出错

再展示文法分析出错的情况，以 test9 为例：

```
test9: $a^{b^{c^{2}d}}$
$     id    ^     {     id    ^     {     id    ^
   {     num   }     id    }     $
$     a     ^     {     b     ^     {     c     ^
   {     2     }     d     }     $

S → $ B $
$ B $        $ id ^ { id ^ { id ^ { num } id } $
B → id D
$ D id       id ^ { id ^ { id ^ { num } id } $
D → ^ { B } D
$ D } B { ^        ^ { id ^ { id ^ { num } id } $
$ D } B {          { id ^ { id ^ { num } id } $
```

```
B → id D
$ D } D id          id ^ { id ^ { num } id } $
D → ^ { B } D
$ D } D } B { ^           ^ { id ^ { num } id } $
$ D } D } B {       { id ^ { num } id } $
B → id D
$ D } D } D id      id ^ { num } id } $
D → ^ { B } D
$ D } D } D } B { ^       ^ { num } id } $
$ D } D } D } B {         { num } id } $
B → num D
$ D } D } D } D num       num } id } $
D → ε
$ D } D } D } ε           } id } $
$ D } D } D }       } id } $
D → B D
B → id D
$ D } D } D D id          id } $
D → ε
$ D } D } D ε       } $
D → ε
$ D } D } ε         } $
$ D } D }           } $
D → ε
$ D } ε             $
Match error!
```

第四章　制导

第一节　分析

因为文法分析采用的是自上而下的分析方法，因此需要使用 L 属性的自上而下的计算。其中需要确定的属性有 top、left、font-style、font-size，其中，font-style 与是否为 id 有关，这个属性在词法分析时即可确定，top、font-size 为继承属性，在文法分析时，每次一个非终结符匹配一个产生式时，通过固定的规则向下传导，并将属性保存在栈中对应的元素里。每当成功匹配一个输入串中的元素时，将栈中对应元素的属性转存入输入串的元素中。

而 left 属性，是一个综合属性，需要自底向上的推导，因此需要构建一个临时栈。将文法分析时栈中弹出的元素存入这个栈中，之后开始反向推导，从栈顶再次弹出元素并存入对应的非终结符中。经过这次处理，将可以成功分析出每个非终结符和终结符的长度，这样在自顶向下推导时，可以提前确定好其绝对位置。

第二节　运行情况

以 test1 为例，制导翻译后输出的 HTML 文件为：

```
1 <html>
2 <head>
3 <META content="text/html;␣charset=gb2312">
4 </head>
5 <body><div style="position:␣absolute;␣top:175px;␣
    left:500px;"><span style="font-size:50px;␣
    font-style:oblique;␣line-
6 height:100%;">154</span></div>
```

```
7  <div style="position:␣absolute;␣top:160px;␣
       left:575px;"><span style="font-size:30px;␣
       font-style:oblique;␣line-
8  height:100%;">5</span></div>
9  </body>
10 </html>
```

test1 至 test7 的运行情况分别如图4.1至图4.7所示。

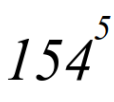


图 4.1 test1

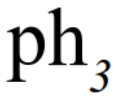


图 4.2 test2

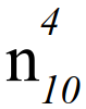


图 4.3 test3

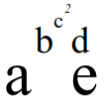


图 4.4 test4

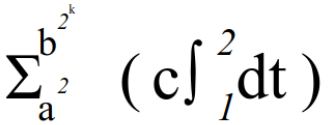
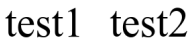


图 4.5 test5



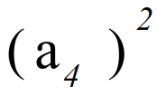


图 4.6 test6

图 4.7 test7

test8, test9, test10 未通过文法分析，故没有进行制导。

第五章　总结

第一节　本文主要工作

本文分析了如何对给定的文法进行词法分析，文法分析和制导，并予以编程实现。词法分析将输入符号区分为 13 中类型，如果有字符不属于该类型之一，则报错。文法分析中，先对给定的文法稍加修改，消除左递归，并通过 LL(1) 文法实现了文法分析。最后在制导翻译时，对于继承属性，在文法分析时加入确定继承属性的代码，之后通过一个堆栈，从后向前传递综合属性，并又做了一次文法分析，确定了综合属性。最后，总结了文章的工作，找出了优点与不足并描述了感悟。

第二节　优点与不足

一、优点

1. 可以在不修改源代码的情况下自动读入最多 99 个 test 文件。

2. 词法分析时，将 _^ 视为一个符号，避免了文法分析时提左因子。

3. 重写文法消除左递归时，只引入了一个非终结符，简化了文法分析流程。

4. 制导翻译时，使用堆栈来从底向上传递属性，简化了代码架构。

5. 撰写报告使用 LaTeX 模板，并参照了毕业论文规范。

二、不足

1. 制导翻译时，进行了两边文法分析，加大了代码量并且有部分重复冗余代码。

2. 在词法分析和文法分析阶段，仅能识别错误并报错，没有实现对错误的恢复。

3. 使用数组而非结构体来存储属性，导致代码的可读性和可维护性降低。

第三节　感想与收获

在本次实验中，我最大的收获是深入理解了编译器的词法分析、文法分析和制导翻译的流程，理解了编译器的工作过程。手工区别字符串和数字很简单，然而具体到代码中，从条件判断，到结果存储，都是对词法分析计算机实现的全新理解。文法在消除左递归后变的很简单，使用矩阵存储控制跳转的条件，当看到计算机完全能够按照自己的程序正确处理输出和匹配时获得了极大的成就感。做到制导时，因为需要对文法分析的程序进行修改，在 DEBUG 时，需要多次手动推导文法分析及制导过程，因此在完成制导后，多次推导已经使我完全掌握了本次实验的理论内容。在分析应该如何处理综合属性和继承属性的过程中，我又理解了如何使用递归或语法树的方法来处理综合属性。

除此之外，我的编程水平得到了较大的提升。本次实验虽然只有 600 多行代码，但主要函数为文法分析和制导翻译综合起来的函数，该函数贡献了过半的代码量。在编写和调试过程中必须时刻掌握整个函数，同时对各个细节也必须了然于胸。

本次实验还提高了我分析问题和解决问题的能力，从刚开始的不知所措，到逐步实现词法、文法时的欣喜，处理制导时又因长时间找不出程序出错原因而几度想放弃。在实现过程中多次修改了既定的程序、数据结构，从而需要很快相出既定计划中，哪一部分需要做修改而哪一部分需要保持不变。多次的尝试使我很好的掌握了程序中哪些点容易出错，而哪些点应当在多个方向进行尝试。

第四节　未来工作

要想更好的完善程序，首先是需要进行错误处理, 例如处理自动匹配不全的括号，或是忽略未知的标点等。其次还需要改进代码，消除现有代码中的冗余。此外还可以尝试其他 LaTeX 排版符号。

参 考 文 献

[1] 陈意云，张昱. 编译原理. 3 版. 北京: 高等教育出版社, 2014: 440.

致　谢

在研究学习期间，我有幸得到了一位老师和两位助教的教导，他们是：我的导师，中国科大李卫海副教授，中国科大信息安全系研究生李莉和崔州平。三位深厚的学术功底，严谨的工作态度和敏锐的科学洞察力使我受益良多。衷心感谢他们一学期来给予我的悉心教导和热情帮助。

科大的姚青松同学在实验过程中与我进行了多次讨论，使我能更好的掌握实验的目的、要求、方法，在此深表谢意。

撰写本篇报告时使用了中科大毕业论文模板[①]，再此一并向模板的创建者和维护者致谢。

[①] https://github.com/ustctug/ustcthesis