



第1章 编译技术概述

信息与软件工程学院

邓伏虎



本章内容

- 1.1 程序设计语言
- 1.2 编译的作用与分类
- 1.3 编译器的结构
- 1.4 编译技术的应用



1.1 程序设计语言

地址	源程序	注解
	ORG 2000H	; 程序机器码从 2000H 单元开始存放。
LABEL0	EQU 2100H	; 将地址 2100H 赋给标号 LABEL0
LABEL1	EQU 2101H	; 将地址 2101H 赋给标号 LABEL1。
LABEL2	EQU 2102H	; 将地址 2102H 赋给标号 LABEL2。
2000	MOV DPTR, #LABEL1	; 加数地址 2101H 赋给 DPTR
2003	MOVBX A, @DPTR	; 取出加数 38 送入累加器 A
2004	MOV B, A	; 转存到 B 寄存器中
2006	MOV DPTR, #LABEL0	; 被加数地址 2100H 赋给 DPTR
2009	MOVBX A, @DPTR	; 取出被加数 55 送入累加器 A
200A	ADD A, B	; 55H+38H 得 8DH 送 A
200C	DA A	; 调整为十进数 93
200D	MOV DPTR, #LABEL2	; 和的地址 2102H 赋给 DPTR
2010	MOVBX @DPTR, A	; 将 93 送入 2102H 单元。
2011LOP:	SJMP LOP	; 循环结束
	ORG 0100H	; 从 2100H 单元开始存放数据
2100LABEL0:	DB 55H, 38H	; 55 存放在 2100H 单元中。
2101	(或 DW 5538H)	; 38 存放在 2101H 单元中。
	END	; 汇编程序结束。

16

0000 0100 0000 0000 0011 1101



1.1 程序设计语言

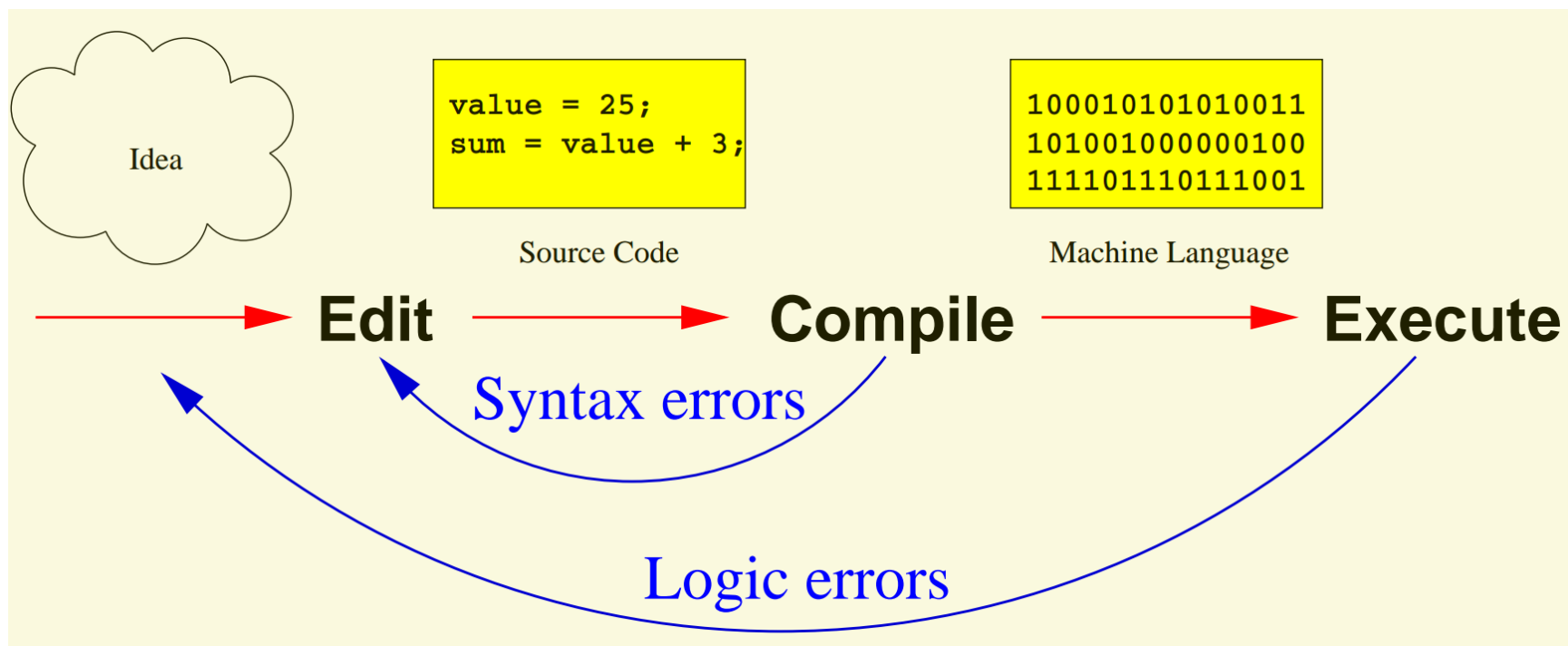
```
man(socrates).  
  
mortal(X) :- man(X).  
  
?- mortal(socrates).
```

第一行写为“苏格拉底是人”，它是一个基本逻辑子句，这是一个简单的事实。第二行是一条规则，并且转化为：“如果X是人，则X是凡人”，或者“所有的人都是凡人”。这个规则用于确定它的输入X何时是“凡人”。规则是语言的一个关键概念，并允许我们对对象及其关系作一般性陈述。它们由一个由一个类似十字旋转门的符号连接的头部和身体组成:-(发音为“如果”)。第三行的内容是：“苏格拉底是否是凡人？”?-“提出一个Prolog查询结果”。



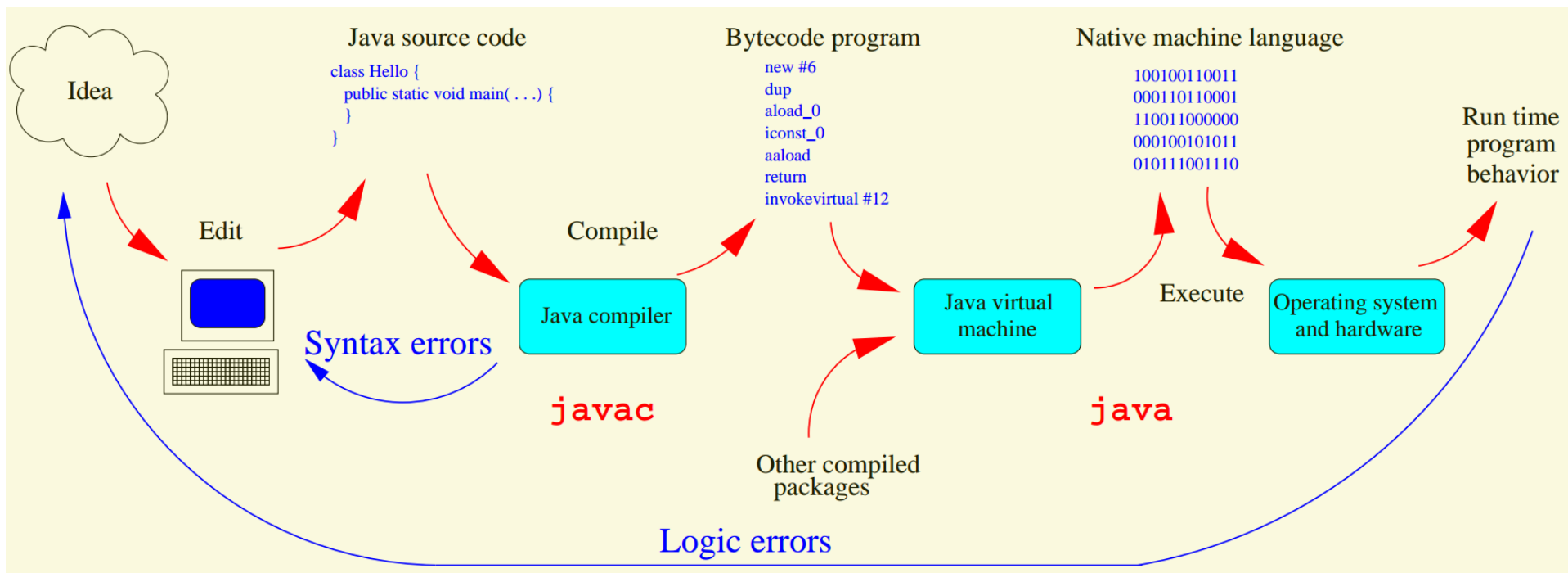
1.1 程序设计语言

高级语言的开发过程



1.1 程序设计语言

高级语言的开发过程

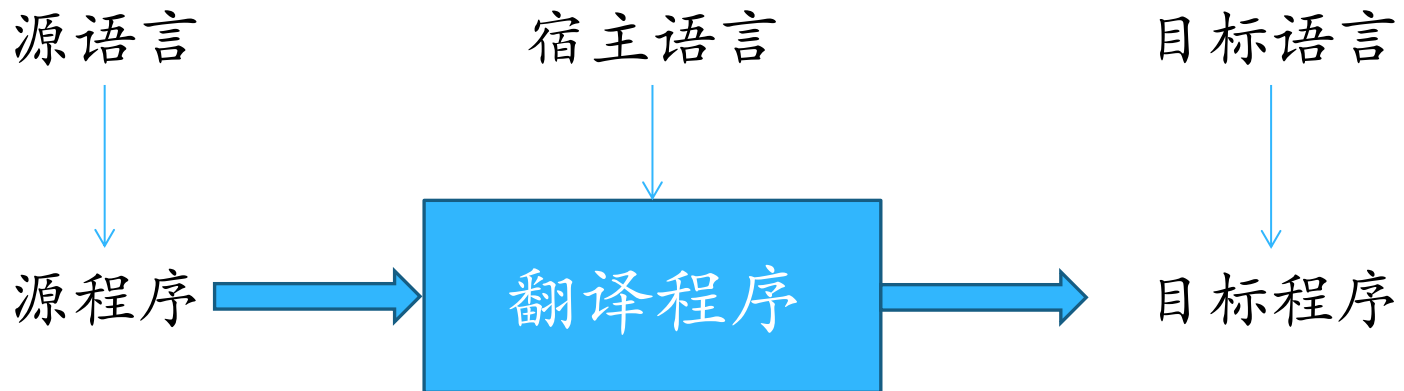


Java语言的开发过程



1.2 编译的作用与分类

- * 翻译程序：将源程序完整地转换成机器语言或者汇编语言程序，然后再处理、执行的程序。





1.2 编译的作用与分类

* 作用:

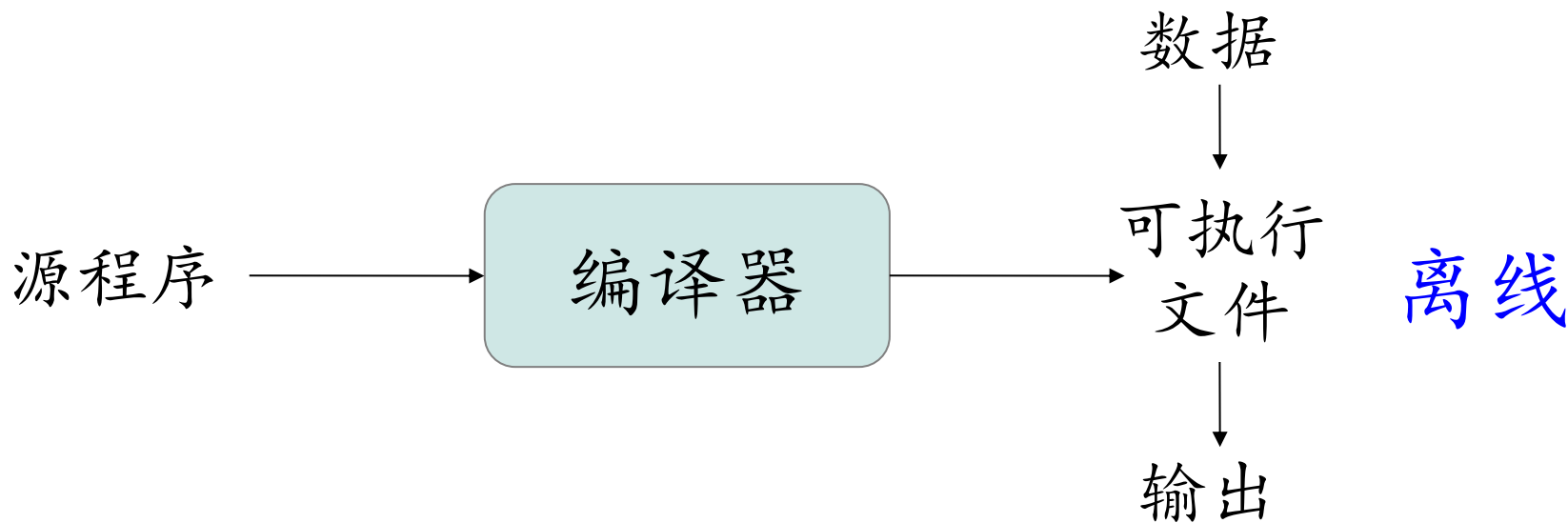
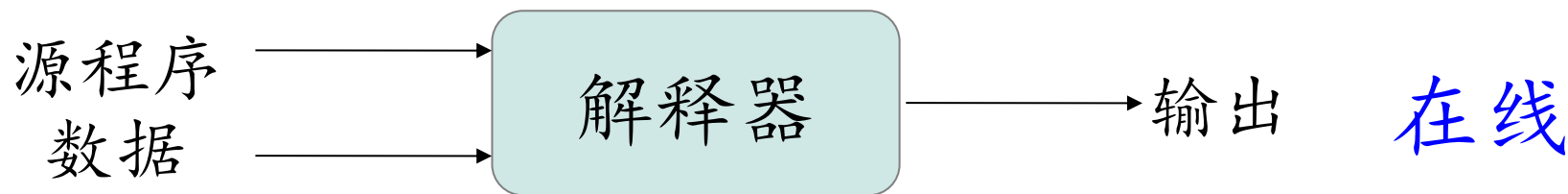
将高级语言转换成机器语言.

* 分类:

- 编译—将一种高级语言程序（源程序）转换成另一种语言的等价程序（目标程序），然后执行目标程序。
- 解释—将源程序的某一条语句翻译后立即执行这一条语句。



1.2 编译的作用与分类





1.2 编译的作用与分类

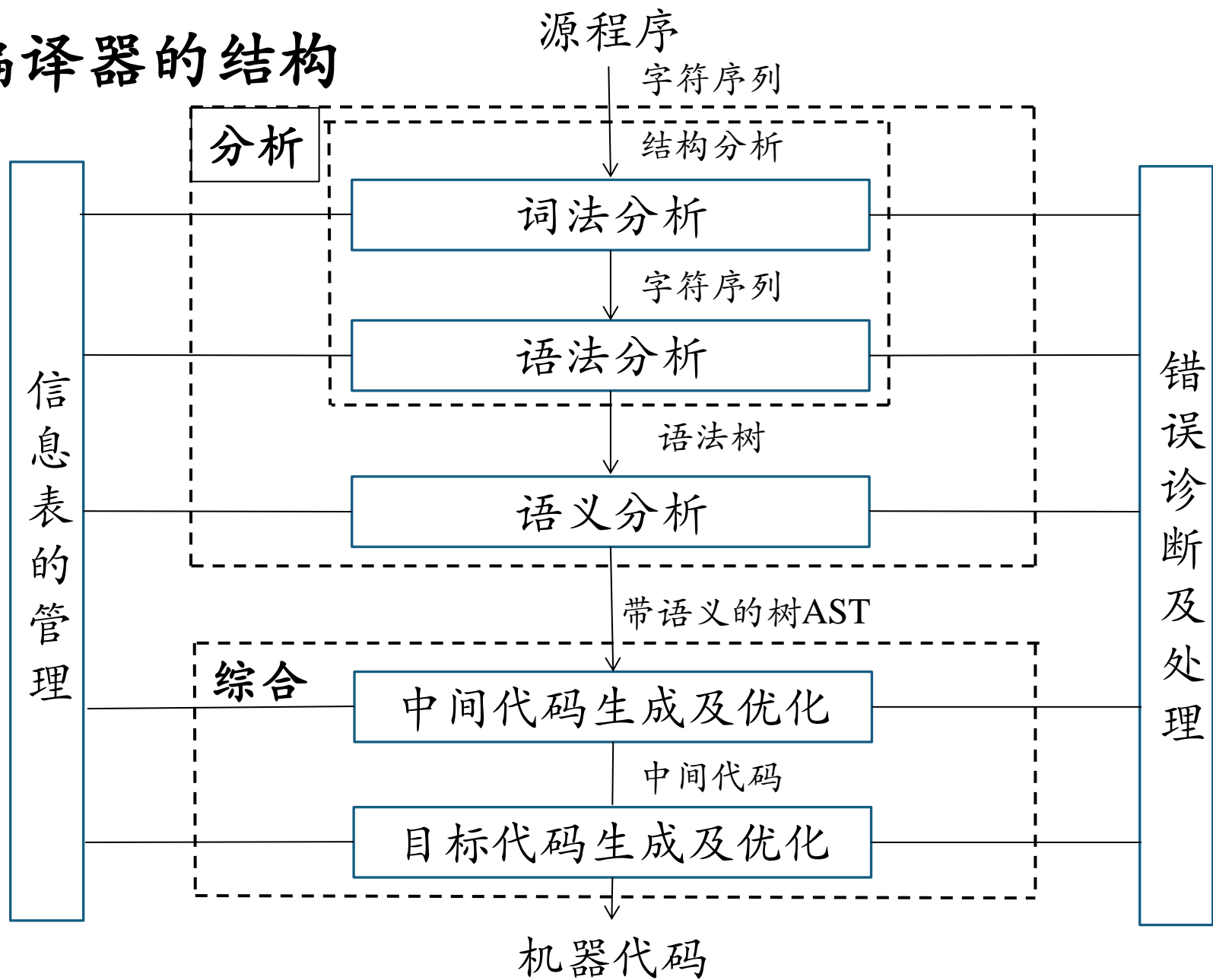
编译还是解释？

Pascal、C/C++、Object-C、Java、C#、Python、Go

编译的：**Fortran、Vala**

解释的：**JavaScript、PHP等**

1.3 编译器的结构





1.3 编译器的结构

词法分析

编译器将源程序当作一个字符串，对它进行从左到右的扫描，并分析，识别出符合词法规则的单词符号。如基本字、标识符、常量、运算符、界符等。

position	=	initial	+	rate	*	60
<id, 1>	<=>	<id, 2>	<+>	<id, 3>	<*>	<60>

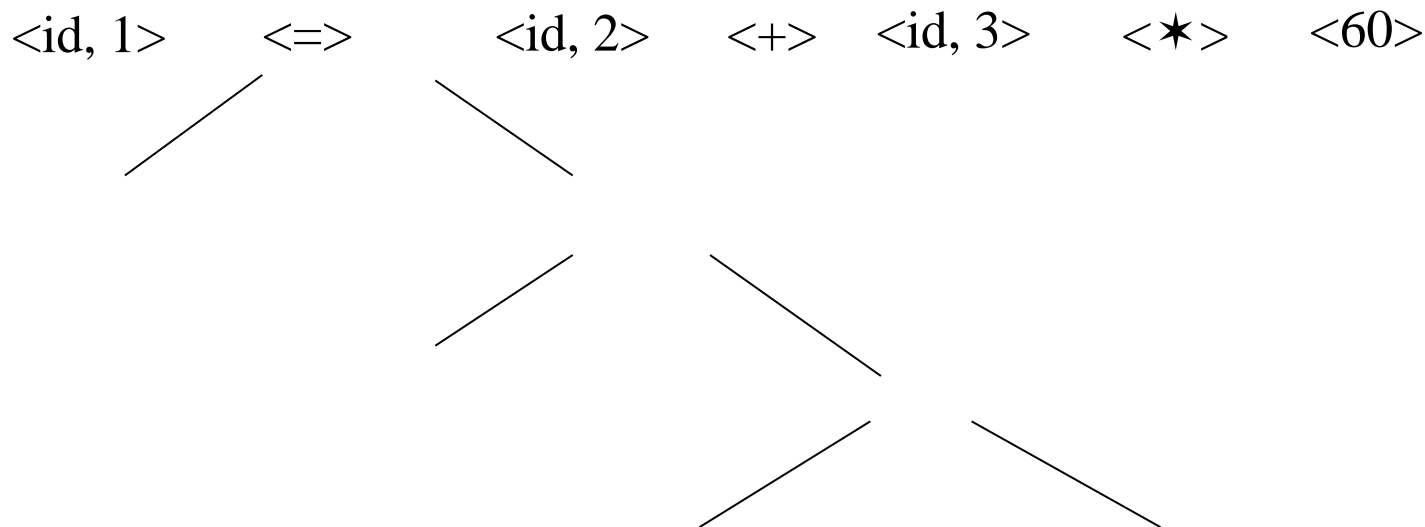
描述工具：有限自动机、正规表达式、状态转换图



1.3 编译器的结构

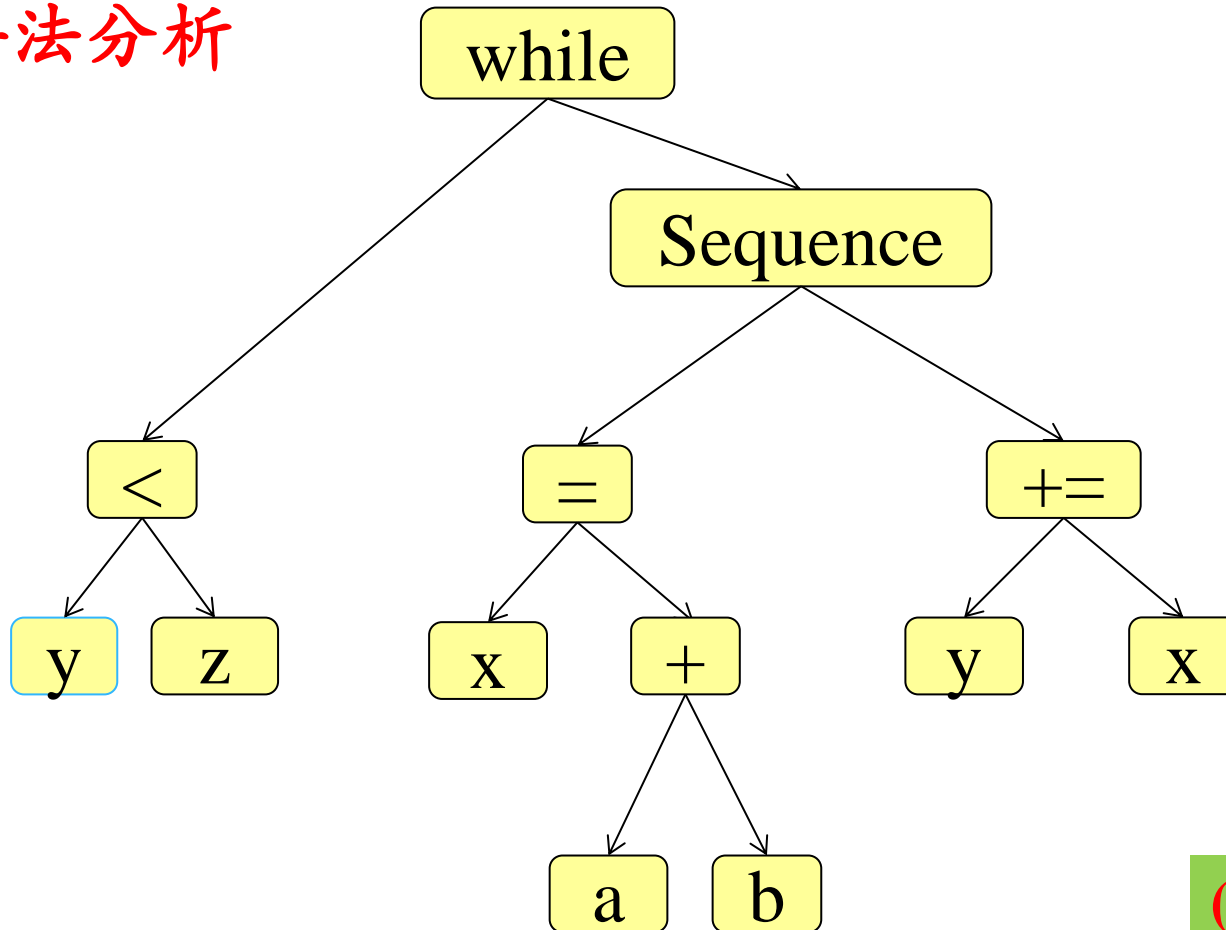
语法分析

根据语法规则，将单词符号构成各类语法单位，进行语法检查，生成语法树。



1.3 编译器的结构

语法分析



```
while (y < z) {  
    int x = a + b;  
    y += x;  
}
```

() {} ; 还需要吗?



1.3 编译器的结构

语义分析

Jack said Jerry left **his** assignment at home

Jack said **Jack** left his assignment at home.

程序设计语言中通常采用**严格**的规则来避免歧义。

编译器通常仅仅完成**有限**的语义分析。



1.3 编译器的结构

语义分析

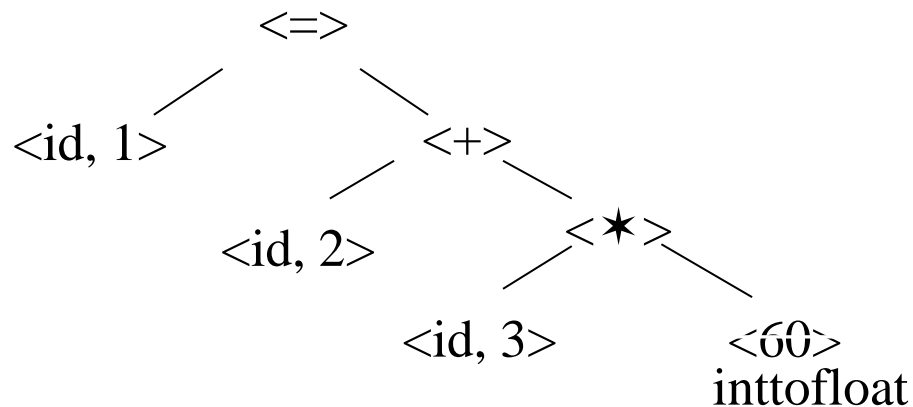
根据语义规则，对语法正确的语法单位进行初步翻译。

静态语义检查

类型检查、控制流检查、唯一性检查、关联名检查

动态语义检查

除零检查、数组溢出





1.3 编译器的结构

语法制导翻译

将静态检查和中间代码生成结合到语法分析中进行的技术。

主要思想：在进行语法分析的同时完成相应的语义处理。

分析过程

为每个产生式配上一个语义子程序

当对产生式进行推导或归约时

调用相应的语义程序

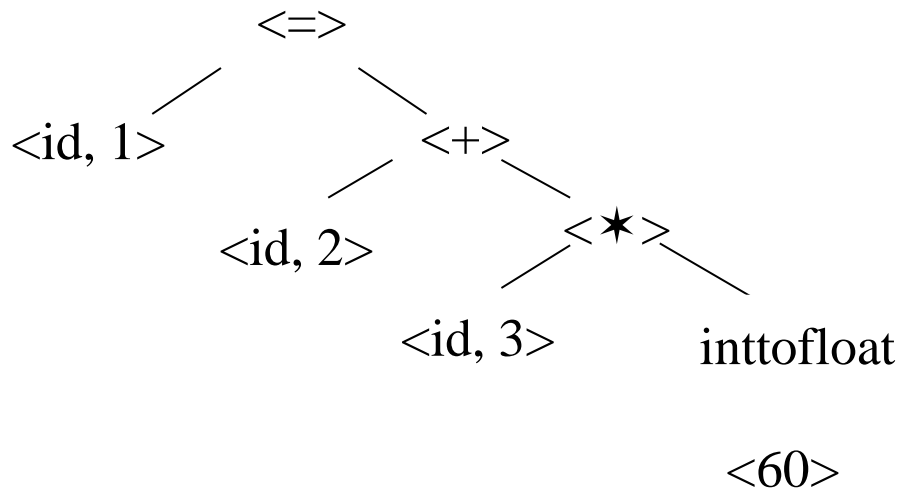
进行语义分析



1.3 编译器的结构

中间代码生成

在静态语义正确的前提下翻译成中间代码或者目标代码。



三地址代码

t1 = inttofloat (60)

t2 = id3 * t1

t3 = id2 + t2

id1 = t3



1.3 编译器的结构

$$a = b * c + b * d$$

逆波兰式:

$abc * bd * + =$

三元式

(1) $(*, b, c)$

(2) $(*, b, d)$

(3) $(+, (1), (2))$

(4) $(:=, a, (3))$

四元式

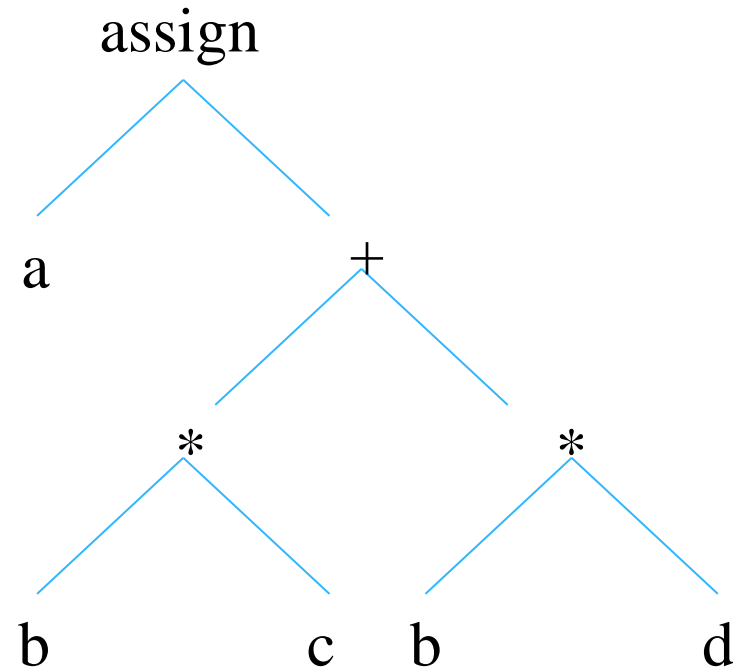
$(*, b, c, T_1)$

$(*, b, d, T_2)$

$(+, T_1, T_2, T_3)$

$(:=, a, T_3, a)$

图表示法:





1.3 编译器的结构

中间代码优化

为了提高目标程序的效率，对程序进行合理变换，在保持功能不变的前提下，使得目标代码具有更高的时间效率和/或空间效率。

代码优化贯穿整个编译过程。

主要方法：控制流分析、数据流分析、变换。

局部优化、循环优化、全局优化。



1.3 编译器的结构

中间代码优化

三地址代码

$t1 = \text{inttofloat}(60)$

$t2 = id3 * t1$

$t3 = id2 + t2$

$id1 = t3$

$t1 = id3 * 60.0$

$id1 = id2 + t1$



1.3 编译器的结构

目标代码生成

将中间代码转换成等价的、具有较高质量的目标代码，以充分利用目标机器的资源。

依赖于**硬件系统结构**和**机器指令**。

目标代码的形式：

绝对机器语言代码---可直接运行

可重定位的机器语言代码---需要连接装配

汇编语言代码---需要汇编



1.3 编译器的结构

目标代码生成

t1 = inttofloat (60)

t2 = id3 * t1

t3 = id2 + t2

id1 = t3

MOV R2, id3

MULF R2, R2, #60.0

MOV R1, id2

ADD R1, R1, R2

STF id1, R1



1.3 编译器的结构

目标代码生成

MIPS Instructions

Meaning

add \$s1,\$s2,\$s3

$\$s1 = \$s2 + \$s3$

sub \$s1,\$s2,\$s3

$\$s1 = \$s2 - \$s3$

bne \$s4,\$s5,Label

下一条指令在 Label if $\$s4 \neq \$s5$

beq \$s4,\$s5,Label

下一条指令在 Label if $\$s4 == \$s5$

j Label

下一条指令在 Label处

slt \$t1,\$s2,\$s3

if $\$s2 < \$s3$, $\$t1 = 1$ else $\$t1 = 0$



1.3 编译器的结构

符号表

用来保存已收集信息的一种数据结构

将信息从声明的地方传递到实际使用的地方

声明阶段：将信息存进符号表中

赋值阶段：取出符号表中的信息

符号表操作：

- 识别到的名字，查询是否已经在表中。

- 将未登记的名字添加到表中。

- 访问表中给定名字的信息。

- 添加或修改表中给定名字的信息。

- 删除一个或多个无用的项。



1.3 编译器的结构

符号表

每个名字对应一个表项

一个表项包括名字域和属性域

名字

属性信息

名字域

名字长度、存储方式

属性域

包含多个子域及标志位

类型、初值、存储大小、相对地址、

参数个数、形参标志、说明标志、赋值标志、



1.3 编译器的结构

出错处理

编译程序各个阶段都可能发现源程序中的错误。发现错误后，编译程序就要进行错误处理，包括：报告错误的性质，发生错误的位置等。

语法错误

不符合语法（词法）规则的错误；
非法字符、括号不匹配、缺少“;”...

语义错误

不符合语义规则的错误
说明错误、作用域错误、类型不一致、数组越界...



1.3 编译器的结构

运行时环境

程序存储空间，数据存储空间

静态存储分配

没有指针、动态分配、过程不可递归调用

栈式存储分配

无法预先知道存储空间大小

堆式存储分配

用户自由申请和释放数据空间

过程1的目标代码

过程2的目标代码

⋮

过程 n 的目标代码

全局和静态变量

栈

堆



1.4 编译技术的应用

* 提高编程效率、缩短调试时间，确保程序可靠性。

- 程序分析、验证、自动化测试工具
- 网络信息处理、网络协议的分析与实现
- 自然语言处理、文本分类、机器翻译系统
- 反病毒、恶意脚本查杀
- 逻辑门电路绘制

* 自动化测试工具

Best Automation Testing Tools for 2018



Best Automation Testing Tools for 2020



File Edit View Go Capture Analyze Statistics Telephony Tools Internals Help

Filter: tcp.s Expression... Clear Apply Save

No.	tcp.segment.multipletails	Protocol	Length	Info
4	tcp.segment.overlap	TCP	54	4643 > http [ACK] Seq=2 Ack=2 wi
5	tcp.segment.overlap.conflict	TCP	60	http > 4792 [ACK] Seq=2 Ack=2 wi
5	tcp.segment.toolongfragment	TCP	60	http > 4795 [SYN, ACK] Seq=0 Ack
5	tcp.segments	TCP	54	4795 > http [ACK] Seq=1 Ack=1 wi
5	tcp.seq	HTTP	287	GET /MFEwTzBNMESwSTAJBgUrDgMCGGU
5	tcp.srcport	TCP	60	http > 4795 [ACK] Seq=1 Ack=234
5	tcp.stream	HTTP	1185	[TCP Previous segment not captur
5		TCP	54	[TCP Dup ACK 53#1] 4795 > http [
5		TLSv1	55	Ignored Unknown Record
5		TCP	66	https > 4633 [ACK] Seq=1 Ack=2 w
5		TCP	66	4796 > http [SYN] Seq=0 win=8192
60	18.701172 199.7.57.72 192.168.0.82	TCP	60	http > 4796 [SYN, ACK] Seq=0 Ack
61	18.701294 192.168.0.82 199.7.57.72	TCP	54	4796 > http [ACK] Seq=1 Ack=1 wi

Frame 53: 287 bytes on wire (2296 bits), 287 bytes captured (2296 bits)

Ethernet II, Src: Hewlett-_2a:b0:76 (d8:d3:85:2a:b0:76), Dst: QnoTechn_01:60:0c (00:17:16:01:60:0c)

Internet Protocol Version 4, Src: 192.168.0.82 (192.168.0.82), Dst: 199.7.57.72 (199.7.57.72)

Transmission Control Protocol, Src Port: 4795 (4795), Dst Port: http (80), Seq: 1, Ack: 1, Len: 233

Source port: 4795 (4795)

Destination port: http (80)

[Stream index: 14]

Sequence number: 1 (relative sequence number)

[Next sequence number: 234 (relative sequence number)]

Acknowledgment number: 1 (relative ack number)

Header length: 20 bytes

Flags: 0x018 (PSH, ACK)

0020	39 48 12 bb 00 50 12 5a dc 13 dd ae 63 f7 50 18	9H...P.ZC.P.
0030	fa f0 d0 fc 00 00 47 45 54 20 2f 4d 46 45 77 54GE T /MFEwT
0040	7a 42 4e 4d 45 73 77 53 54 41 4a 42 67 55 72 44	ZBNMESwS TAJBgUrD
0050	67 4d 43 47 67 55 41 42 42 53 70 75 43 45 33 61	gMCGGUAB BSpuCE3a
0060	4b 33 47 69 76 5a 50 7a 47 51 4a 36 4c 35 42 52	K3GivZPz GQJ6L5BR
0070	79 5a 6f 66 77 51 55 6c 39 42 72 71 43 5a 77 79	yZofwQUl 9BrqCZwy
0080	4b 45 25 32 46 6c 42 38 49 4c 63 51 31 6d 36 53	KE%2F1B8 ILcQ1m6S
0090	68 48 76 49 43 45 44 32 46 45 4c 54 46 76 4a 78	hhVICED2 FELTFVjX
00a0	4a 69 53 4f 4d 48 41 56 5a 38 64 6b 25 33 44 20	JiSOMHAV Z8dk%3D
00b0	48 54 54 50 2f 31 2e 31 0d 0a 43 6f 6e 6e 65 63	HTTP/1.1 ..Connec
00c0	74 69 6f 6e 3a 20 4b 65 65 70 2d 41 6c 69 76 65	tion: Ke ep-Alive
00d0	0d 0a 41 63 63 65 70 74 3a 20 2a 2f 2a 0d 0a 55	..Accent : */*...ll

Source Port (tcp.srcport), 2 bytes

Packets: 3509 Displayed: 3509 Marked: 0 Load Profile: Default

iprotocalsec.com

* 自然

气象电文

【题目】新潟县天气预报
【日期】2010年5月17日14时47分
【发布】新潟县气象台
【原文】
受冷湿空气的影响，17日全天新潟县的气流将一直处于非常不稳定的状态。
截至17日18时的预计降雨量：
新潟县北部250毫米
新潟县南部200毫米
〈灾害情况〉
预警信息：泥石流灾害
注意事项：遇到冰雹、强风，或者雷云来临时，请不要外出，确保人身安全。



AI记者撰写的原稿

观众朋友们早上好，我们来介绍一下这一时段的天气情况。
根据气象台发布的数据显示，受冷湿空气的影响，17日全天新潟县的气流将一直处于不稳定的状态，并会伴随强降雨。
截至17日18时，新潟县北部的降雨量预计将超过250毫米。
其间还有可能伴随泥石流灾害，以及龙卷风等强风灾害，请大家出行时注意安全。

NTT DATA 的“AI记者”生成的天气预报稿件



电脑体检



木马查杀



电脑清理



系统修复



优化加速



功能大全



小金库



软件



扫描完成！共有4个需处理的危险项

已用时：00:07:04

扫描类型：快速扫描

扫描项目：21269 个

暂不处理



发现 4 个危险项



建议禁止

建议禁止开机自动运行的程序

无需启动的程序

D:\Program Files (x86)\GenuinePlatform 3\npgplib.dll



建议禁止

建议禁止开机自动运行的程序

无需启动的程序

D:\Program Files\Microvirt\MEmu\MemuService.exe



建议修复

系统图标被隐藏

系统设置篡改

修复此项可以将系统图标恢复显示，如果不是您将这些图标隐藏，建议修复。



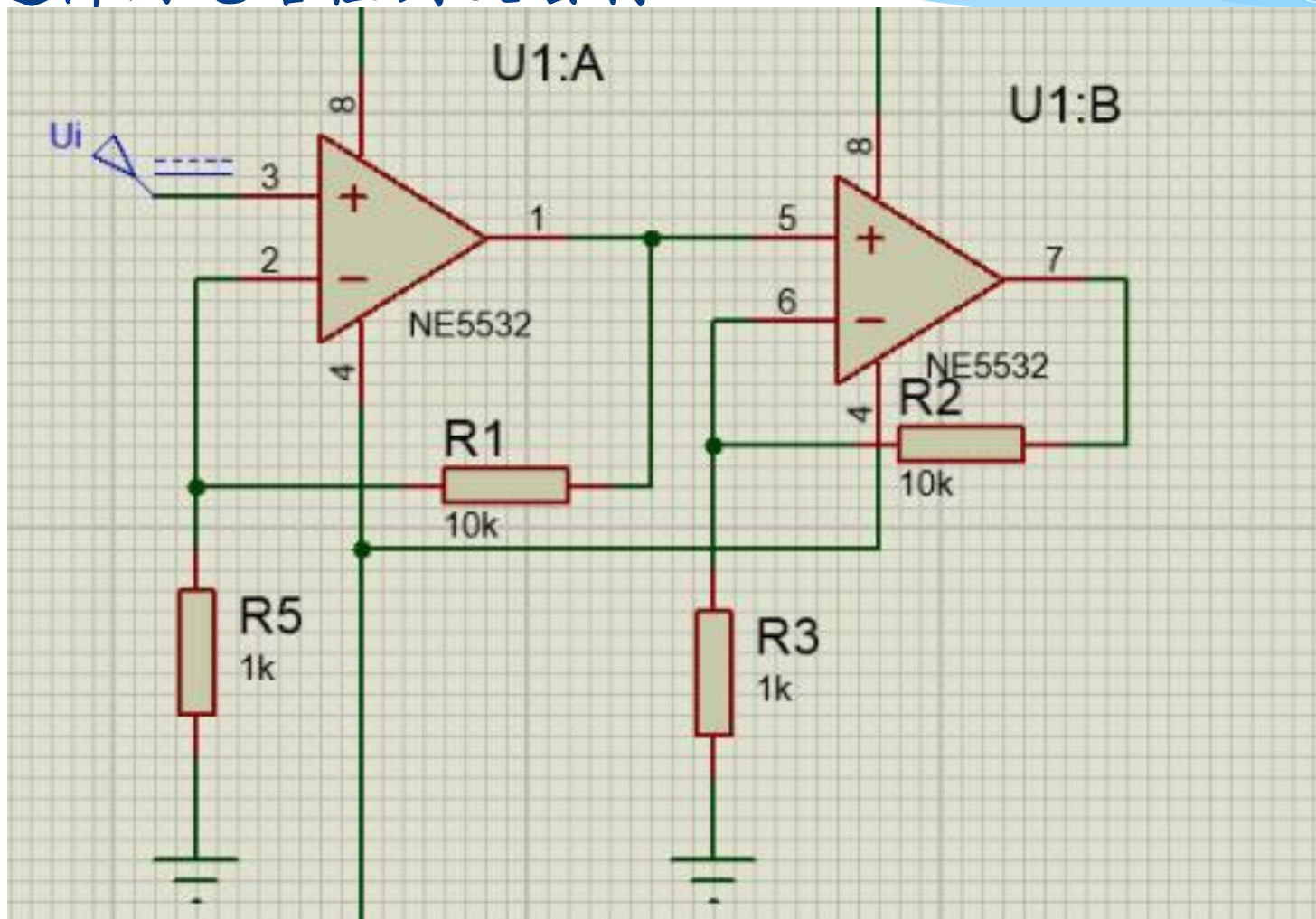
建议删除

无效的快捷方式：正版软件管理与服务平台.lnk

系统设置篡改

此快捷方式所指向的文件不存在，可能是由于程序卸载不彻底导致，或者是木马残留的快捷方式。

* 逻辑门电路检测及绘制





本章小结

- 课程概述
- 内容、意义、安排
 - 什么是编译程序
- 翻译、编译、解释
 - 编译的基本过程
 - 编译程序的结构
 - 编译技术的应用