

类 C 语言至 MIPS 编译器使用说明

顾城 2012 年 6 月 8 日

概述

类 C 语言至 MIPS 编译器（下文简称“编译器”）是将一种类 C 语言 LogoScript 编译为可被计算机组成实验所设计的类 MIPS CPU（仅包含 20 条基本指令）运行的汇编、二进制代码的实用工具。

LogoScript 语言起源于软件工程大作业所制作的 MY PC Logo 项目，在该项目中，项目组成员定义了类 C 的 LogoScript 语言，完成了基于栈的虚拟机指令集的设计，定义了 15 条指令，并开发了 LogoScript 语言解释器。

在此基础上，我又增加了两条新指令（IN 和 OUT），并使用 C# 语言编写完成了这 17 条栈 CPU 指令到 20 条 MIPS 指令的翻译器。由此实现了 LogoScript 语言到栈 CPU 指令再到 MIPS 指令的转化，实现了编译器。

使用说明

编译器是命令程序，使用方法为：

`Compiler.exe script codeMemorySize dataMemorySize`

script: 输入源文件名，包含要编译的程序代码。

codeMemorySize: 指令内存大小。如果不指定，默认为 32。(100%不够用)

dataMemorySize: 数据内存大小。如果不指定，默认为 64。

执行编译器后，如果代码正确编译，则在当前文件夹下产生三个文件：

`Asm.txt` 汇编源程序

`sc_datamem.mif` 数据内存文件

`sc_instmem.mif` 指令内存文件

如果发生编译错误则会提示错误所在行号和错误类型。另外，如果指令内存或者数据内存大小过小，会给出编译警告，提示用户扩大内存，否则编译出来的代码无法正常工作。

语言规约

详见《LogoScript》语言规约，该规约摘自软件工程项目。编译器实现的语言与该规约略有不同，主要有：

数字全部为 32 位整数，而非双精度浮点类型。

不作类型检查。

不支持除法。

支持 in 和 out 关键字，**identifier=***in*(port)表示从 port 端口读入 I/O，并存入变量 identifier 内；**out**(port,expression)指计算 expression 的值，并写入 port 端口。

不足和改进

由于是从基于栈的 CPU 翻译代码翻译而来，因此编译器产生的代码不够优化，冗余代码过多，寄存器使用率比较低。

一种比较好的方案是直接在代码生成的时候直接生成 MIPS 代码，但会遇到寄存器分配等问题；另一种是直接使用 MIPS 代码编写 LogoScript 虚拟机，并将基于栈的代码作为数据，CPU 执行时，通过虚拟机解释执行代码，但效率比较低。由于时间问题，因此上述两种方案都未实现。