

# 想为RISC添加定制指令，需要修改编译器后端，打算用LLVM或GCC，求大神介绍下？

同标题，打算在RISC扩展一些指令，需要改动编译器后端，打算用LLVM或者GCC也行，想知道后端的大致流程，比如DAG如何覆盖-匹配-调度这类顺序，或...显示全部

关注问题

写回答

邀请回答

1 条评论

分享

举报

你可以邀请下面用户，快速获得回答

委托刘看山邀请

开启后为你智能邀请潜在的回答者

开启邀请

依田聪

[object Object] 最近回答过该领域问题

邀请回答

cgoxopx

最近回答过该领域问题

邀请回答

子澄催眠治疗师

最近回答过该领域问题

邀请回答

Codetector

可能对该问题感兴趣

邀请回答

更多推荐结果

1 个回答

默认排序

「已注销」

弗吉尼亚理工大学 计算机专业博士在读

22 人赞同了该回答

在LLVM 3.8之前，LLVM的指令选择有两个，一个基于DAG覆盖的SelectionDAG，另外一个FastISel。前者在-O方式启用，后则在-O0的时候启用。从3.8开始，后端增加了一个GlobalSel的指令选择器(基于自动macro expanding，其实和FastISel类似，只不过FastISel是手写指令匹配模式)。由于GlobalSel尚未完全成熟(LLVM 6.0仅完整支持ARM后端，所以仅基于SelectionDAG讨论后端流程。

LLVM后端大致的工作流程如下所述

1. LLVM IR -> SelectionDAG[与机器无关的ISD操作符].

2. 中间经过若干个DAGCombiner和DAGTypeLegalizer，DAGLegalizer(这些类的作用是将目标机器不支持的类型和DAG操作转换为目标机器支持的操作，同时删除部分冗余计算).

3. 通过调用instructionSelect函数，变换过的SelectionDAG输入到指令选择器中，将机器无关的操作转换为机器支持的操作.

4. 将第3步生成的SelectionDAG[仅使用目标机器支持的操作]输入到指令调度器中，指令调度器通过满足data, output, anti-, control约束依照SelectionDAG构造一个ScheduleDAG.

5. 指令发射(instruction emit), 对ScheduleDAG进行拓扑排序，生成一个包含SUnit的Sequence列表，里面保存了最终的指令顺序流。然后遍历Sequence列表，依据每个SUnit中的SDNode的操作码得到TargetInstrInfo，然后调用buildMI生成一条MachineInstr，之后就是按照tii描述的操作数信息往mi中插入不同种类的操作数，比如：registerOperand将会创建一个虚拟寄存器，然后调用mi.addRegister()。然后把建立好的mi插入到对应的机器基本块MBB中。

示例

广告

相关问题

如何学好用好LLVM？ 5 个回答

请问LLVM与GCC之间的关系，网上说LLVM 是编译器的架构，在这个架构上可以搭建多个小编译器（类似C、C++/JAVA/），不理解的对不对，还请高手补充？ 8 个回答

现在国内还有哪些公司搞开源编译器（GCC、LLVM、OPEN64）？有哪些人？ 8 个回答

LLVM Pass编写完，应该如何编译成 .so 文件？ 7 个回答

有没有更高层次的 LLVM？ 4 个回答

相关推荐

Linux 从入门到精通

陈祥琳

41 人读过

阅读

深入分析 GCC

王亚刚

58 人读过

阅读

自己动手构造编译系统：编译、汇编与链接

126 人读过

阅读

广告

刘看山 · 知乎指南 · 知乎协议 · 知乎隐私保护指引

应用 · 工作 · 申请开通知乎机构号

侵权举报 · 网上有害信息举报专区

京 ICP 证 110745 号

京 ICP 备 13052560 号 - 1

京公网安备 11010802010035 号

互联网药品信息服务资格证书

(京) - 非经营性 - 2017 - 0067

违法和不良信息举报：010-82716601

儿童色情信息举报专区

前置条件

- 1. 在编译LLVM的时候，需要使用cmake -DLLVM\_ENABLE\_ASSERTIONS=ON才会启用view-dags等选项，这些选项下面会被用于查看输出的DAG图。
- 2. 安装xdot等dot查看工具，Ubuntu可以使用sudo apt-get install xdot命令安装。
- 3. 如果使用较新版本的LLVM，下面的clang-cc 命令名替换为clang -cc1。

```
// 文件名保存为add.c
int add(int a, int b)
{
    return a + b;
}
```

LLVM IR输入

使用命令clang-cc -emit-llvm -triple=i386-linux-gnu add.c，生成的LLVM IR如下：

```
; 舍去无关的target说明信息

define i32 @add(i32 %a, i32 %b) nounwind {
entry:
    %retval = alloca i32                ; <i32*> [#uses=2]
    %a.addr = alloca i32                ; <i32*> [#uses=2]
    %b.addr = alloca i32                ; <i32*> [#uses=2]
    store i32 %a, i32* %a.addr
    store i32 %b, i32* %b.addr
    %tmp = load i32* %a.addr             ; <i32> [#uses=1]
    %tmp1 = load i32* %b.addr            ; <i32> [#uses=1]
    %add = add nsw i32 %tmp, %tmp1       ; <i32> [#uses=1]
    store i32 %add, i32* %retval
    %0 = load i32* %retval               ; <i32> [#uses=1]
    ret i32 %0
}
```

然后调用SelectionDAGBuild类以基本块为单位构造一个SelectionDAG图，所有的alloca会被分配一个函数栈索引FrameIndex，然后构造一个对应的FrameIndexSDNode结点。store和load指令分别转换为ISD::store和ISD::load，add操作转换为ISD::add，ret会转换为X86ISD::RET\_FLAG。

SelectionDAG变换

使用clang-cc -S -triple=i386-linux-gnu add.c -view-dag-combine1-dags 查看输出的DAG如下图所示：



相关问题

[如何学好用好LLVM？](#) 5 个回答

[请问LLVM与GCC之间的关系，网上说LLVM 是编译器的架构，在这个架构上可以搭建多个小编译器（类似C、C++/JAVA/），不知理解的对不对，还请高手补充？](#) 8 个回答

[现在国内还有哪些公司搞开源编译器（GCC、LLVM、OPEN64）？有哪些人？](#) 8 个回答

[LLVM Pass编写完，应该如何编译成.so文件？](#) 7 个回答

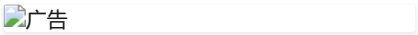
[有没有更高层次的 LLVM？](#) 4 个回答

相关推荐

**Linux 从入门到精通**  
陈祥琳  
41 人读过 阅读

**深入分析 GCC**  
王亚刚  
58 人读过 阅读

**自己动手构造编译系统：编译、汇编与链接**  
126 人读过 阅读



刘看山 · 知乎指南 · 知乎协议 · 知乎隐私保护指引

应用 · 工作 · 申请开通知乎机构号

侵权举报 · 网上有害信息举报专区

京 ICP 证 110745 号

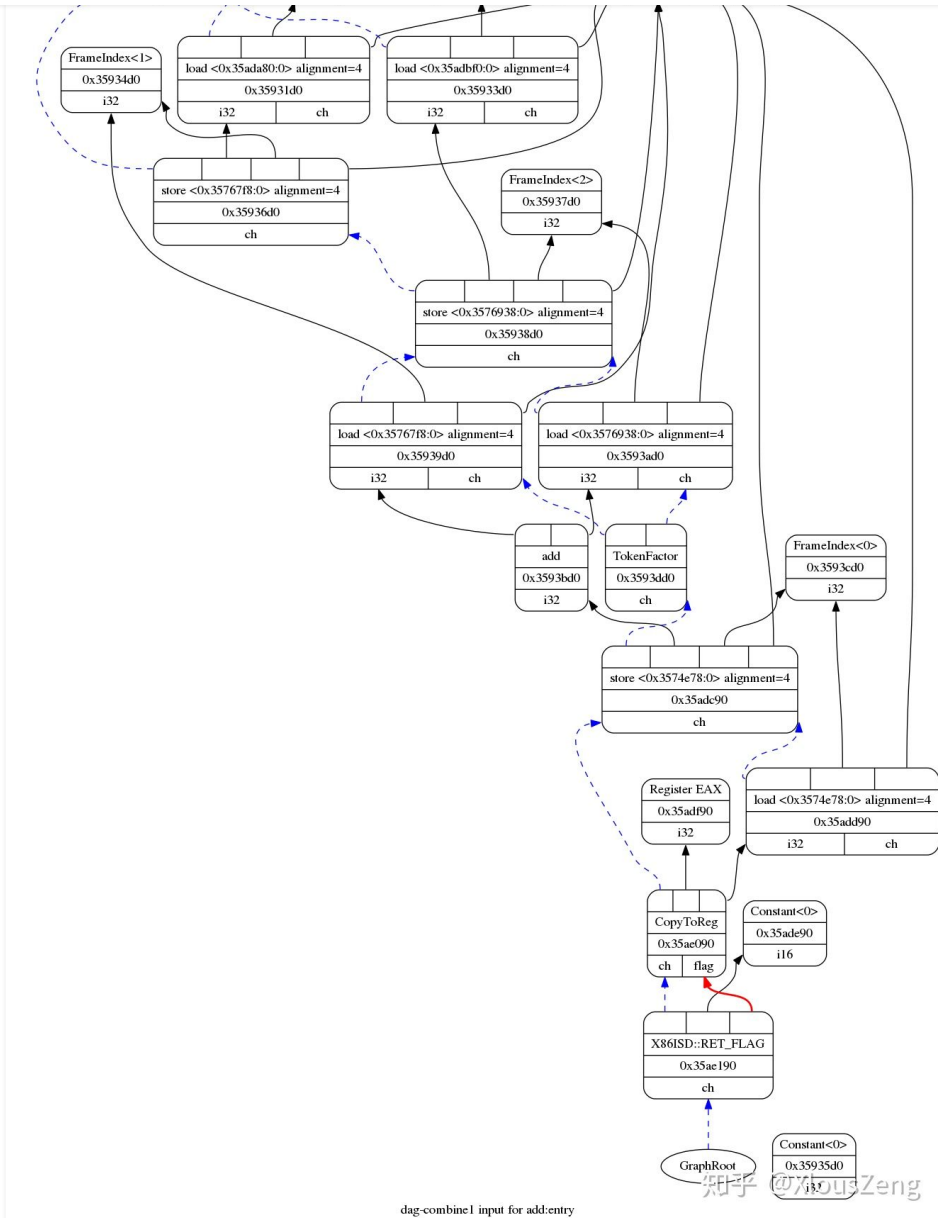
京 ICP 备 13052560 号 - 1

京公网安备 11010802010035 号

互联网药品信息服务资格证书  
(京) - 非经营性 - 2017 - 0067

违法和不良信息举报：010-82716601

儿童色情信息举报专区



经过若干次combine, legalization等变换后，查看命令clang-cc -S -triple=i386-linux-gnu add.c -view-isel-dags. 生成的DAG图如下：

广告

### 相关问题

如何学好用好LLVM？ 5 个回答

请问LLVM与GCC之间的关系，网上说LLVM 是编译器的架构，在这个架构上可以搭建多个小编译器（类似C、C++/JAVA/），不理解的对不对，还请高手补充？ 8 个回答

现在国内还有哪些公司搞开源编译器（GCC、LLVM、OPEN64）？有哪些人？ 8 个回答

LLVM Pass编写完，应该如何编译成 .so 文件？ 7 个回答

有没有更高层次的 LLVM？ 4 个回答

### 相关推荐

 **Linux 从入门到精通**  
陈祥琳  
41 人读过 阅读

 **深入分析 GCC**  
王亚刚  
58 人读过 阅读

 **自己动手构造编译系统：编译、汇编与链接**  
126 人读过 阅读

广告

刘看山 · 知乎指南 · 知乎协议 · 知乎隐私保护指引

应用 · 工作 · 申请开通知乎机构号

侵权举报 · 网上有害信息举报专区

京 ICP 证 110745 号

京 ICP 备 13052560 号 - 1

 京公网安备 11010802010035 号

互联网药品信息服务资格证书

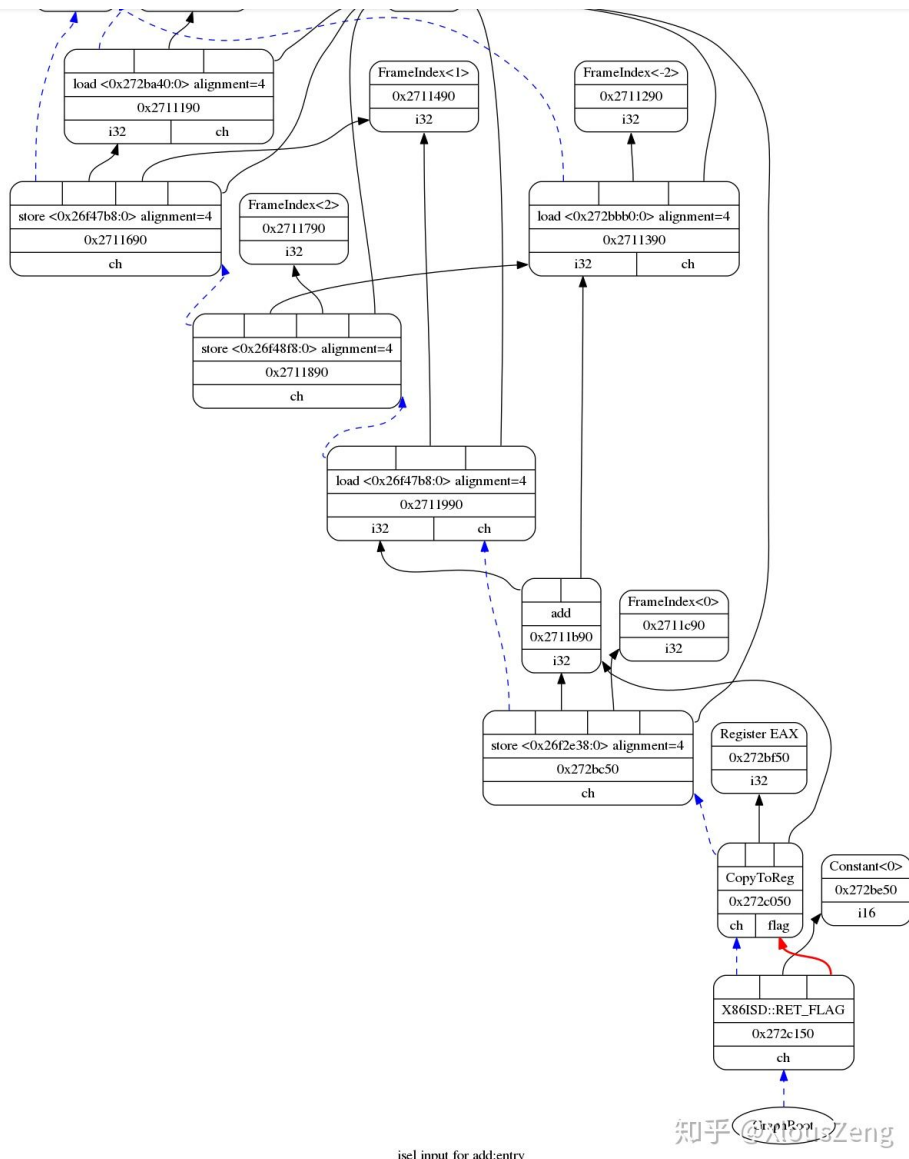
(京) - 非经营性 - 2017 - 0067

违法和不良信息举报：010-82716601

儿童色情信息举报专区







## 指令选择

指令选择会调用InstructionSelect()函数, 从root结点开始, 从后往前遍历SDNode, 进入X86GenDAGISel.inc:SelectCode函数, 该函数会为SelectionDAG上的每个SDNode根据操作码调用相关的函数, 然后将该子树替换成机器指令模式(此处会略过BasicBlock, Register, TargetConstant, TokenFactor, EntryToken, CopyFromReg, CopyToReg这些操作码)。如ISD::ADD会调用Select\_ISD\_ADD\_i32()函数将该子树替换X86ISD::ADD32rr指令模版, ISD::LOAD会被替换为X86ISD::MOV32rm, ISD::Store指令会被替换为X86ISD::MOV32mr。当然由于此处未开启优化, 所以add没有匹配X86ISD::ADD32rm模版。

clang-cc -S -triple=i386-linux-gnu add.c -view-sched-dags, 查看指令选择之后如图所示:

## 广告

### 相关问题

如何学好用好LLVM? 5 个回答

请问LLVM与GCC之间的关系, 网上说LLVM 是编译器的架构, 在这个架构上可以搭建多个小编译器 (类似C、C++/JAVA/), 不知理解的对不对, 还请高手补充? 8 个回答

现在国内还有哪些公司搞开源编译器 (GCC、LLVM、OPEN64)? 有哪些人? 8 个回答

LLVM Pass编写完, 应该如何编译成 .so 文件? 7 个回答

有没有更高层次的 LLVM? 4 个回答

### 相关推荐



Linux 从入门到精通

陈祥琳

41 人读过

阅读



深入分析 GCC

王亚刚

58 人读过

阅读



自己动手构造编译系统: 编译、汇编与链接

126 人读过

阅读

## 广告

刘看山 · 知乎指南 · 知乎协议 · 知乎隐私保护指引

应用 · 工作 · 申请开通知乎机构号

侵权举报 · 网上有害信息举报专区

京 ICP 证 110745 号

京 ICP 备 13052560 号 - 1

京公网安备 11010802010035 号

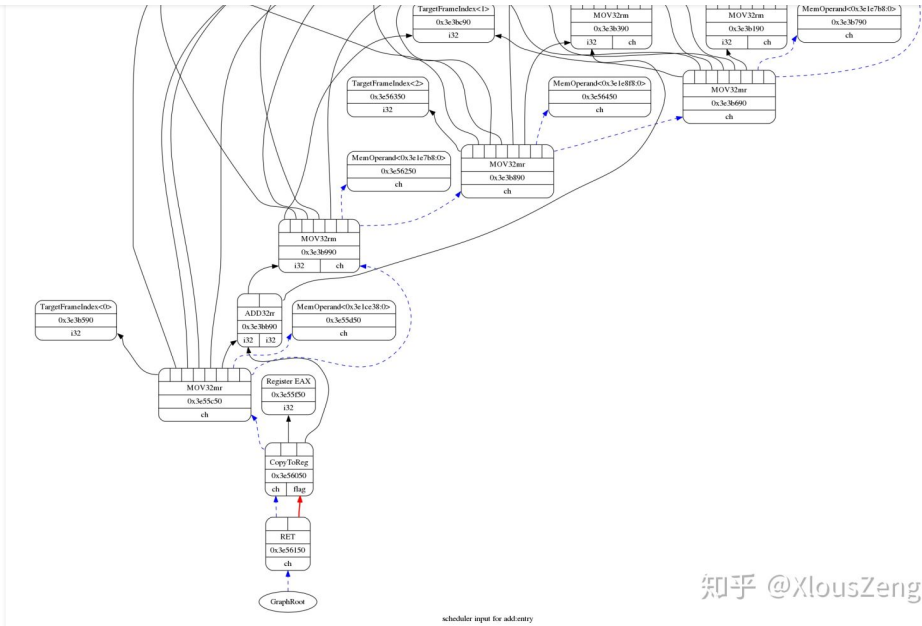
互联网药品信息服务资格证书

(京) - 非经营性 - 2017 - 0067

违法和不良信息举报: 010-82716601

儿童色情信息举报专区





知乎 @XlousZeng

指令调度

使用命令clang-cc -S -triple=i386-linux-gnu add.c view-sunit-dags,  
查看最终经过ScheduleDAGSDFast指令调度器生成的ScheduleDAG如图所示:

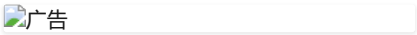


相关问题

- 如何学好用好LLVM? 5 个回答
- 请问LLVM与GCC之间的关系，网上说LLVM 是编译器的架构，在这个架构上可以搭建多个小编译器（类似C、C++/JAVA/），不知理解的对不对，还请高手补充? 8 个回答
- 现在国内还有哪些公司搞开源编译器（GCC、LLVM、OPEN64）?有哪些人? 8 个回答
- LLVM Pass编写完，应该如何编译成 .so 文件? 7 个回答
- 有没有更高层次的 LLVM? 4 个回答

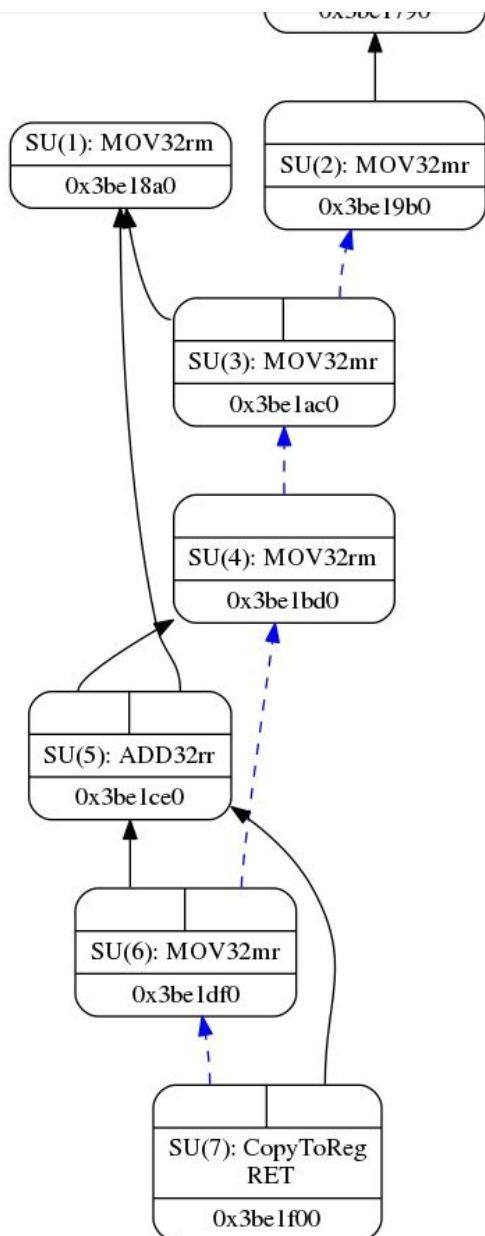
相关推荐

- Linux 从入门到精通**  
陈祥琳  
41 人读过 阅读
- 深入分析 GCC**  
王亚刚  
58 人读过 阅读
- 自己动手构造编译系统：编译、汇编与链接**  
126 人读过 阅读



刘看山 · 知乎指南 · 知乎协议 · 知乎隐私保护指引  
应用 · 工作 · 申请开通知乎机构号  
侵权举报 · 网上有害信息举报专区  
京 ICP 证 110745 号  
京 ICP 备 13052560 号 - 1  
京公网安备 11010802010035 号  
互联网药品信息服务资格证书  
(京) - 非经营性 - 2017 - 0067  
违法和不良信息举报: 010-82716601  
儿童色情信息举报专区





Scheduling-Units Graph for add:entry

## 代码发射

经过CodeEmit, 最终生成的机器相关的machine basic block如下:

```
# Machine code for add():
<fi#-2>: size is 4 bytes, alignment is 1 byte, fixed at location [SP+8]
<fi#-1>: size is 4 bytes, alignment is 1 byte, fixed at location [SP+4]
<fi#0>: size is 4 bytes, alignment is 4 bytes, at location [SP+4]
<fi#1>: size is 4 bytes, alignment is 4 bytes, at location [SP+4]
<fi#2>: size is 4 bytes, alignment is 4 bytes, at location [SP+4]
Live Outs: EAX

entry: 0x371b008, LLVM BB @0x36d5780, ID#0:
%reg1024<def> = MOV32rm <fi#-2>, 1, %reg0, 0, %reg0, Mem:LD(4,4) [FixedStack-2 + 0]
%reg1025<def> = MOV32rm <fi#-1>, 1, %reg0, 0, %reg0, Mem:LD(4,4) [FixedStack-1 + 0]
MOV32mr <fi#1>, 1, %reg0, 0, %reg0, %reg1025, Mem:ST(4,4) [a.addr + 0]
MOV32mr <fi#2>, 1, %reg0, 0, %reg0, %reg1024, Mem:ST(4,4) [b.addr + 0]
%reg1026<def> = MOV32rm <fi#1>, 1, %reg0, 0, %reg0, Mem:LD(4,4) [a.addr + 0]
%reg1027<def> = ADD32rr %reg1026, %reg1024, %EFLAGS<imp-def>
MOV32mr <fi#0>, 1, %reg0, 0, %reg0, %reg1027, Mem:ST(4,4) [retval + 0]
%EAX<def> = MOV32rr %reg1027
RET
```

## 广告

### 相关问题

如何学好用好LLVM? 5 个回答

请问LLVM与GCC之间的关系, 网上说LLVM 是编译器的架构, 在这个架构上可以搭建多个小编译器 (类似C、C++/JAVA/), 不知理解的对不对, 还请高手补充? 8 个回答

现在国内还有哪些公司搞开源编译器 (GCC、LLVM、OPEN64) ? 有哪些人? 8 个回答

LLVM Pass编写完, 应该如何编译成 .so 文件? 7 个回答

有没有更高层次的 LLVM? 4 个回答

### 相关推荐



Linux 从入门到精通

陈祥琳

41 人读过

阅读



深入分析 GCC

王亚刚

58 人读过

阅读



自己动手构造编译系统: 编译、汇编与链接

126 人读过

阅读

## 广告

刘看山 · 知乎指南 · 知乎协议 · 知乎隐私保护指引

应用 · 工作 · 申请开通知乎机构号

侵权举报 · 网上有害信息举报专区

京 ICP 证 110745 号

京 ICP 备 13052560 号 - 1

京公网安备 11010802010035 号

互联网药品信息服务资格证书

(京) - 非经营性 - 2017 - 0067

违法和不良信息举报: 010-82716601

儿童色情信息举报专区



知乎

首页发现等你来答

黑龙江女童遭继母虐待

提问

82

上述的指令未经任何优化，所以生成的代码不是最优的。存在冗余的指令，并且没有利用X86的指令集优势(add操作可以使用内存操作数)，但是用于叙述原理是完全足够的。

编辑于 2018-06-23

赞同 22

添加评论

分享

收藏

喜欢

收起

写回答