

Especificação do Hidrasm

PET Computação da UFRGS

June 11, 2013

Contents

1	Base	1
2	Arquivo de Descrição	2
2.1	Endereçamentos	2
2.2	Instruções	2
2.2.1	Exemplo	3
2.3	Máquina	3
2.3.1	Exemplo	3
2.4	Registradores	3
2.4.1	Exemplo	3
3	Expressão	4
3.1	Exemplo	4
4	Números	4
4.1	Exemplos	5
5	Erros e Avisos	5
5.1	Arquivo	5
5.1.1	Variáveis	5
5.1.2	Exemplo	6

1 Base

Para simplicidade, o programa do montador possui somente uma interface por linha de comando. Para integração com o simulador, a interface chama esse programa. Isso é feito para, caso deseje-se mudar o montador, não será necessário alterar o código fonte da interface ou do simulador.

Cada máquina é especificada em um arquivo, o qual contém as seguintes seções:

- Endereçamentos
- Instruções
- Máquina
- Registradores

2 Arquivo de Descrição

2.1 Endereçamentos

Descreve os modos de endereçamento de cada máquina, como o formato pelo qual é identificado, sobre quais tipos é válido e se é relativo ao PC ou não.

Identificador addressings

Formato <nome> [= | -] <código binário> <identificador>

onde:

nome nome do modo de endereçamento (usado internamente). Pode ser qualquer sequência de caracteres a-z e A-Z;

= | - Determina se o valor do endereço deve ser calculado em relação ao PC ou não. "=" indica que deve ser simplesmente copiado, "-" que é relativo ao PC. (Padrão: =);

código binário número;

identificador Expressão que identifica o modo de endereçamento. Somente um registrador ou endereço pode aparecer por expressão.

2.2 Instruções

Descreve qual o mnemônico, formato, modos de endereçamento e registradores possíveis para cada instrução.

Identificador instructions

Formato <tamanho> <mnemônico> <operandos> <endereçamento0, ...>
< <registrador0, ...> | - | *> <formato binário>

onde:

tamanho número de bits da instrução;

mnemônico qualquer sequência dos caracteres a-z e A-Z (case insensitive);

operandos Expressão que indica como os operandos da instrução devem aparecer. Uma expressão sem variáveis será ignorada e indicará que nenhum operando deve ser utilizado;

endereçamento0 ... quais modos de endereçamento podem ser usados com essa instrução. * indica qualquer um;

formato binário como que a instrução é montada. São usados os símbolos:

a[[n]][(m)] n-ésimo endereço (se o n não for informado, são colocados na ordem em que aparecem). m indica quantos bits terá o endereço, sendo desnecessário quando há somente um endereço no formato. Se houver mais de um endereço na expressão, todos usarão o mesmo número de bits caso este não seja informado;

m[n] n-ésimo modo de endereçamento (se o n não for informado, são colocados na ordem em que aparecem);
r[n] n-ésimo registrador (se o n não for informado, são colocados na ordem em que aparecem);
0 bit 0;
1 bit 1;
qualquer outro caractere ignorado

2.2.1 Exemplo

16 ADD a d - 0011000ma[0](8)

2.3 Máquina

Descrever diversos aspectos da máquina, como tamanho do PC e a endianness usada.

Identificador machine

Formato <little-endian | big-endian> <bits do PC>

onde:

little-endian | **big-endian** endianness usada;

bits do pc quantos bits o PC possui (determina o tamanho máximo da memória).

2.3.1 Exemplo

little-endian 8

2.4 Registradores

Descreve cada registrador usado pela máquina que pode aparecer no código assembly, definindo seu identificador.

Identificador registers

Formato <nome> <valor>

onde:

nome nome do registrador (para o código Assembly).

valor número que indica o valor do registrador (para o código de máquina)

2.4.1 Exemplo

R0 0b

3 Expressão

Uma expressão identifica um padrão de forma similar a uma expressão regular. São usados os símbolos:

r registrador

n endereço numérico

l label

a n ou l

o qualquer um dos acima (o de operando)

**** caractere de escape

qualquer outro caractere indica algum que deve aparecer na frase

As variáveis identificadas podem ter somente os caracteres em a-zA-Z0-9 e `_`. Elas devem estar separadas por um caractere em branco se a expressão exigir algum dos caracteres que as compõem (ver exemplo).

Ao ler uma expressão, espaços entre os caracteres são ignorados. Por isso, uma expressão não pode conter espaços (se contiver, estes serão ignorados).

3.1 Exemplo

As seguintes frases satisfazem "(as),a":

- "(125 s),red" (match de 125 e red)
- " (reds s) , 125" (match de reds e 125)

As seguintes frases não satisfazem "(as),a":

- "1(24 s),red"
- "(reds),32"
- "((red s),32"

4 Números

São suportados números nas seguintes bases:

binário seguido de um "b" ou "B"

decimal seguido de um "d" ou "D" ou nada

hexadecimal precedido por um algarismo e seguido de um "h" ou "H". Os caracteres dos número podem ser minúsculos ou maiúsculos.

4.1 Exemplos

010b 2

010 10

010d 10

010h 16

ah Não corresponde a um número válido

5 Erros e Avisos

As mensagens de erro e os avisos estarão em um arquivo separado, de forma a facilitar eventuais traduções do programa além de permitir diferentes níveis de rigidez para os erros.

5.1 Arquivo

Formato <código> <e | w> <mensagem>

onde:

código número da mensagem

e erro

w aviso

mensagem a mensagem que será usada.

Comentários são feitos com o símbolo '#'.

5.1.1 Variáveis

Para que o usuário possua mais informações, as mensagens podem conter variáveis que descrevem a linha em que ocorreu o problema. São elas:

ADDRESSING_MODE modo de endereçamento usado;

DISTANCE valor, em bytes, do último endereço relativo ao PC (sem truncamentos);

EXPECTED_OPERANDS número de operandos esperado pela diretiva ou instrução;

FOUND_OPERANDS número de operandos encontrados;

LABEL nome da label;

LAST_ORG_LINE número da linha do último ORG encontrado;

LINE a linha atual do código;

MNEMONIC mnemônico da instrução ou diretiva;

OPERAND_SIZE valor máximo que o operando pode assumir.

5.1.2 Exemplo

#Erros

- (1) 0 e Instrução ou diretiva desconhecida: \$MNEMONIC
- (2) 1 e Número incorreto de operandos: esperava-se \$EXPECTED_OPERANDS, encontrou-se \$FOUND_OPERANDS
- (3) 2 e Endereço relativo ao PC muito grande: \$LABEL está a uma distância de \$DISTANCE bytes do PC, mas o máximo possível são \$OPERAND_SIZE bytes
- (4) #Avisos
- (5)
- (6) 3 w Região possivelmente sobrescrita a partir da linha \$LAST_ORG_LINE
- (7) 4 w Modo de endereçamento desconhecido: \$ADDRESSING_MODE

onde os números entre () indicam as linhas no arquivo