Tema 8. Generación de Representaciones Intermedias

Francisco José Ribadas Pena

PROCESADORES DE LENGUAJES 4º Informática

ribadas@uvigo.es

7 de mayo de 2006

8.1 Introducción

Fich.

Fuente

→

A. LÉXICO

A. SINTÁCTICO

A. SEMÁNTICO

→ GENERACIÓN REPRESENT. INTERMEDIA

 \rightarrow repr. inter.

GENERACIÓN OPTIMIZ. DE CÓDIGO

Fich. Objeto →

Representación Intermedia (RI): Lenguaje de una máquina abstracta usado como interfaz entre Análisis y Gen. de Código.

- Etapa intermedia, previa a la generación de código objeto.
- Permite especificar operaciones sobre la máquina objetivo sin entrar en detalles de bajo nivel específicos de la arquitectura.
- Aisla el análisis (léxico, sintáctico, semántico) de la generación de código objeto.
- Muy relacionado con el desarrollo de intérpretes

Ventajas:

- 1. Mayor modularidad. (Ofrece abstracción para componentes de alto nivel)
 - Aisla elementos de más alto nivel de los dependientes de la máquina
- 2. Facilita optimización y generación de código
 - Elimina/simplifica características específicas máquina objetivo
 - nº limitado de registros, tipos de instrucciones
 - alineación de datos, modos de direccionamiento, etc,...
 - Permite optimizaciones independientes de la máquina
 - lacktriangle Ofrece representación simple y uniforme \Rightarrow fácil generar/optimizar código
- 3. Mayor transportabilidad
 - Independencia de la máquina objetivo y del lenguaje fuente
 - Análisis no depende de arquitectura destino
 - Generación código no depende del leng. original
 - Crear un compilador para otra máquina ⇒ basta crear una nueva etapa final de generación
 - Ejemplo: multicompiladores

Inconvenientes:

- 1. Necesidad de una fase extra para traducir a código máquina
 - Mayor coste computacional
- 2. Dificultad para definir un leng. intermedio adecuado
 - Compromiso entre la representación de elementos código fuente y del código máquina.

8.2 Tipos de Representaciones Intermedias

Características deseables de una R.I.:

- Sencillo de producir durante fase de A. Semántico
- Facilitar traducción al leng. máquina final para todas las posibles máquinas objetivo
- Construcciones claras, sencillas y uniformes, con significado unívoco
 - Facilita la especificación de la traducción a la cada máquina objetivo

Tipos de R.I.:

- Representaciones arbóreas: árboles de sintaxis abstracta, GDAs
- Representaciones lineales: polaca inversa, código 3 direcciones

Ejemplos de R.I:

- Diana: represent. arbórea usada en compiladores de ADA
- RTL (register transfer language): usada en familia de compiladores
 GCC
- Código-P: compiladores Pascal (también común en intérpretes)
- WAM (Warren abstract machinne): intérpretes Prolog
- Bytecodes JAVA

Ejemplo GCC.

8.2.1 R.I. Arbóreas

Basadas en árboles de sintaxis abstracta,

- Árboles de análisis sintáctico son info. superflua
- Nodos hoja → operandos
- Nodos internos → operadores del Lenguaje Intermedio

Ejemplos

Pueden usarse representaciones de más bajo nivel.

```
IF (a < 5) THEN c := b + 1;
```

También se usan GDAs (grafos dirigidos acíclicos)

- Representaciones condensadas de árboles abstractos
 - Subestructuras idénticas representadas una única vez
 - Uso de múltiples referencias a regiones comunes
- Mejora de tamaño y optimización de código implícita
- Ejemplo: a := b * (-c) + b * (-c)

8.2.2 R.I. Lineales

(a) Notación Polaca Inversa (RPN)

Notación <u>postfija</u> con los operadores situados a continuación de los operandos.

Ejemplo: $d := (a + b) * c \rightarrow d * a * b + c * :=$

Muy usado como código intermedio en intérpretes.

- Interpretación muy sencilla (basta una pila)
- Ejemplo: Postscript

Ventajas:

- Muy sencillo para expresiones aritméticas.
 - No necesita paréntesis
 - ullet Precedencia está implícita o valores temporales implícitos
- Interpretación/Generación de código muy simple
 - Sólo necesita una pila. Algoritmo interpretación:
 - 1. Recorrer lista, apilando operandos hasta llegar a un operador
 - 2. Tomar los operandos necesarios de la pila y aplicarles el operador
 - 3. Apilar resultado y continuar
 - Muy usado en primeras calculadoras comerciales

Inconvenientes:

- Difícil de entender (una sóla línea de código intermedio)
- Complicado representar control flujo (saltos)
- No es útil para optimización de código.

(b) Código de 3 Direcciones

Generalización del código ensamblador de una máquina virtual de 3 direcciones.

- Cada instrucción consiste en un operador y hasta 3 direcciones (2 operandos, 1 resultado)
- FORMATO GENERAL: x := y OP z
- x, y, z sin direcciones. Referencias a:
 - Nombres (dir. de variables)
 - Constantes
 - Variables temporales (creadas por el compilador durante la gen. de la R. I.)
 - Etiquetas (direcciones de instrucciones)
- Ejemplos:

```
d := x + 9 * z

temp1 = 9 * z

temp2 = x + temp1
d = temp2

etq1:
    x := 9
    etq2:
    ...
IF (a<b) THEN x:=9
```

Existe una referencia explícita a los resultados intermedios de las operaciones mediante las vars. temporales.

→ En RPN, referencias implícitas a resultados almacenados en pila

Posibilidades de representación: tercetos tercetos indirectos

(b.1) CUARTETOS

Tuplas de 4 elementos.

```
(<operador>, <operando1>, <operando2>, <resultado>)
```

Ejemplos:

(b.1) TERCETOS

Tuplas de 3 elementos.

```
(<operador>, <operando1>, <operando2>)
```

La dir. destino del resultado esta asociada de forma implícita a cada terceto.

- Existe una var. temporal asociada a cada terceto donde se guarda su resultado
- Valores intermedios se referencian indicando el nº del terceto que lo crea

Ejemplos:

Características:

- Más concisos que cuartetos (menos espacio)
- Evita manejo explícito de vars. temporales
- Complica optimización: mover/copiar/borrar tercetos más complejo
- Equivalen a arboles de sintaxis abstracta (con ramificación limitada a 2 descendientes)

```
s := a + b * c

1:(*,b,c)

2:(+,a,[1])

3:(:=,s,[2])
```

(b.1) TERCETOS INDIRECTOS

- Orden de ejecución de los tercetos determinada por un vector de apuntadores a triples (VECTOR de SECUENCIA)
- Referencias a vars. temporales se hacen directamente al terceto, no al vector de secuencia
- Representación más compacta → no repetición de tercetos
- Indirección facilita mover/copiar/borrar → simplifica optimiz.
- Ejemplo:

```
COD. 3 DIR.
               TERCETOS
                                       VECTOR SECUENCIA
               101:(/,c,d)
                                       1: 101
a := c / d
               102:(:=,a,[101])
c := c + 1
                                       2: 102
               103:(+,c,1)
b := c / d
                                       3: 103
               104:(:=,c,[103])
                                       4: 104
m := a - b
               105:(:=,b,[101])
                                       5: 101 (el valor de c ha cambiado)
               106: (-, [102], [105])
                                       6: 105
               107:(:=,m,[106])
                                       7: 106
                                       8: 107
```

8.3 Generación de Código Intermedio

Ejemplo de generación de C.I. para los constructores típicos de los lenguajes de programación.

(1) Lenguaje Intermedio:

- Instrucción de asignación: "a := b" \rightsquigarrow (:=,b, ,a)
- Operadores binarios: "x := a OP b" \rightarrow (OP,a,b,x)
 - Enteros ADDI SUBI MULI *EDIVI /E
- Operadores unarios: "x := OP a" \rightarrow (OP,a, ,x)

 - Cambio signo: MINUS
 Conversión tipos { CONVF (convierte int a float) CONVI (convierte float a int)

Saltos

- Incondicional: "goto etq" → (goto,etq,)
- Condicional: "if x goto etq" → (if,x,etq,) Si x es true salta a etq. Además, saltos con ops. relacionales (>, <, ==, etc..) "if x > y goto etq" \rightarrow (if_GT,x,y,etq)
- Declaración etiquetas: "etq: <...>" \longrightarrow (label,etq, ,) Etiqueta "etq" estará asociada a la siguiente instrucción.

Llamadas a procedimiento

Llamada al procedimiento proc(x1, x2, ..., xN, con n argumentos.

```
PARAM x1
PARAM x2
...
PARAM xN
call proc, N
```

Fin llamada a procedim.: "return valor" → (return, valor, ,)

- Acceso a memoria
 - Acceso indirecto:

```
"x := y[i]" \rightsquigarrow (:=[],y,i,x) (acceso)
"x[i] := y" \rightsquigarrow ([]:=,y,i,x) (asignación)
```

 ${\tt "a[j]"}={\sf posición}$ de memoria j unidades (bytes/palabras) después de la dirección de a.

```
x[10] := a+b
temp100 := a+b
x[10] := temp100
```

• Punteros:

```
"\mathbf{x} := \mathbf{\&y}" (asigna a x la dirección de y)
"\mathbf{x} := \mathbf{`y}" (asigna a x el contenido apuntado por la dir. guardada en y)
"\mathbf{`x} := \mathbf{y}" (guarda en la dir. guardada en la var. x el valor de y)
```

(2) Suposiciones:

- Mantendremos el cod. generado en un atributo asociado a los símbolos de la gramática
- 1 TDS por bloque + 1 TDS asociada a cada tipo registro
- Contenido entradas TDS
 - <u>Información de direccionamiento:</u> posición relativa (*offset*) respecto al inicio del área de datos del bloque de código actual
 - \longrightarrow variables locales se sitúan consecutivamente según orden de declaración
 - Tamaño de las variables:

char: 1 real: 8 record: \sum tamaño campos int: 4 puntero:4 array : tamaño elementos \times num. elementos

- Etiquetas: asociadas a una instrucción del lenguaje intermedio
 - → se refieren a direcciones de memoria de la zona de instrucciones
- Temporales: dir. de memoria destinadas al almacenamiento de valores intermedios
 - → al generar cod. objeto se decidirá si se refieren a registro o una pos. de memoria

(3) Atributos

- dir: referencia a la dirección en memoria/registro asociada a un identificador(variable) o a un temporal
 - ightarrow todo no terminal de la gramática usado en expresiones tendrá asociada siempre una var. temporal
- código: guarda el C.I.(lista de cuartetos o tercetos) generado para una construcción del lenguaje

Funciones Se supondrá que están definidas las dos funciones siguientes

- crearTemporal(): genera una nueva variable temporal
- generarCl(instruccion): genera la representación en Cl de una instrucción de 3 direcciones
- nuevaEtiqueta(): genera una nueva etiqueta

8.3.1 CI para expresiones aritméticas

```
S \to \mathsf{id} := E
                      { buscarTDS(id.texto)
                        /* obtenemos: id.tipo, id.dir, id.tamaño */
                        S.codigo = E.codigo +
                                     generarCI('id.dir = E.dir')
                      }
E \to E op E
                      { E0.dir = crearTemporal()
(op:+,-,*,/,mod,...)
                        E0.codigo = E1.codigo +
                                      E2.codigo +
                                      generarCI('E0.dir = E1.dir OP E2.dir')
                      }
  E \to \mathsf{op}\ E
                      { E0.dir = crearTemporal()
   (op: -, \wedge, \dots)
                        E0.codigo = E1.codigo +
                                      generarCI('E0.dir = OP E1.dir')
                      }
  E \rightarrow (E)
                      { E0.dir = E1.dir
                        E0.codigo = E1.codigo
                      }
  E \rightarrow {\rm const}
                      { E.dir = crearTemporal()
                        E.codigo = generarCI('E.dir = const')
                      }
    E \to \mathsf{id}
                      { buscarTDS(id.texto)
                        /* obtenemos: id.tipo, id.dir, id.tamaño */
                        E.dir = id.dir
                        E.codigo = <>
                      }
```

8.3.1 (cont.) Conversión de Tipos

- Arrastrar y consultar tipo asociado a construcciones (tema anterior)
- Incluir código para conversión implícita cuando sea necesario
- Aplicar la instrucción de C.I. que corresponda a cada tipo de dato

Ejemplo: Operador sobrecargado +

```
E \rightarrow E + E
     { E0.dir = crearTemporal()
       E0.codigo = E1.codigo +
                    E2.codigo
       if (E1.tipo = REAL) and (E2.tipo = INTEGER)
                EO.tipo = REAL
                E0.codigo = E0.codigo +
                             generarCI('E2.dir = CONV_F E2.dir')+
                             generarCI('E0.dir = E1.dir +_R E2.dir')
       else if (E1.tipo = INTEGER) and (E2.tipo = REAL)
                EO.tipo = REAL
                E0.codigo = E0.codigo +
                             generarCI('E1.dir = CONV_F E1.pos')+
                             generarCI('E0.dir = E1.dir +_R E2.dir')
       else if (E1.tipo = REAL) and (E2.tipo = REAL)
                EO.tipo = REAL
                E0.codigo = E0.codigo +
                             generarCI('E0.dir = E1.dir +_R E2.dir')
       else if (E1.tipo = INTEGER) and (E2.tipo = INTEGER)
                EO.tipo = INTEGER
                E0.codigo = E0.codigo +
                             \texttt{generarCI('E0.dir = E1.dir} +_I \texttt{E2.dir')}
       else
                E0.tipo = ERROR
                E0.codigo = <>
     }
```

8.3.1 (cont.) Casos Especiales

(1) REGISTROS

Suponemos campos almacenados de forma consecutiva.

Acceso a campos

```
E \rightarrow \text{id.id} \qquad \{ \text{ buscarTDS(id1.texto)} \\ /* \text{ obtenemos: tipo, dir. inicio, TDS registro (R) */} \\ \text{R.buscarTDS(id2.texto)} \\ /* \text{ obtenemos: tipo, dir. relativa en registro */} \\ \text{E.dir = id1.dir + id2.dir} \\ /* \text{ no es necesario generar código */} \\ \}
```

Escritura de campos

```
S \rightarrow \mathbf{id}.\mathbf{id} := E \qquad \{ \text{ buscarTDS(id1.texto)} \\ /* \text{ obtenemos info. de registro R */} \\ \text{R.buscarTDS(id2.texto)} \\ \text{S.codigo} = \text{E.codigo +} \\ \text{generarCI('id1.dir[id2.dir]} = \text{E.dir')} \\ \}
```

(2) ARRAYS

Suponemos arrays de 1 dimensión de tipo array(I,T)

- T: tipo de los elementos, guardamos su tamaño en TDS
- I: rango del índice, guardamos valores limite en TDS

Acceso

```
E \rightarrow \operatorname{id} \left[E\right] \quad \{ \text{ buscarTDS(id.texto)} \\ /* \text{ obtenemos: tipo 'array(I,T)', dir inicio,} \\ \text{ tamaño tipo base T, limite inferior de I */} \\ \text{E0.dir = crearTemporal()} \\ \text{despl = crearTemporal() /* despl. en array */} \\ \text{indice = crearTemporal() /* despl. en memoria */} \\ \text{E0.codigo = E1.codigo +} \\ \text{generarCI('despl = E1.dir - I.lim_inferior')+} \\ \text{generarCI('indice = despl * T.tamaño')+} \\ \text{generarCI('E0.dir = id.dir[indice]')} \\ \}
```

Escritura

```
S \rightarrow \operatorname{id}[E] := E \qquad \{ \text{ buscarTDS(id.texto)} \\ /* \text{ obtenemos info. del array */} \\ \text{S.dir = crearTemporal()} \\ \text{despl = crearTemporal() /* despl. en array */} \\ \text{indice = crearTemporal() /* despl. en memoria */} \\ \text{S0.codigo = E2.codigo +} \\ \text{E1.codigo +} \\ \text{generarCI('despl = E1.dir - I.lim_inferior')+} \\ \text{generarCI('indice = despl * T.tamaño')+} \\ \text{generarCI('id.dir[indice] = E2.dir')} \\ \}
```

(4) EJEMPLOS

8.3.2 CI para expresiones lógicas

Dos posibilidades

- 1. Codificando valores true/false numéricamente
 - $true \neq 0$, false = 0
 - Evaluar expr. lógicas del mismo modo que las aritméticas
- 2. Mediante control de flujo (saltos)
 - El valor de una expresión booleana se representa implícitamente mediante una posición alcanzada en el programa
 - \rightarrow si true salta a un punto; si false salta a otro
 - Ventajoso para evaluar exptr. boolenas en instrucciones de control de flujo
 - → facilita evaluar expresiones lógicas en cortocircuito

(Veremos sólo la primera posibilidad)

```
E \rightarrow E \text{ op } E \qquad \{ \text{ E0.dir = crearTemporal()} \\ \text{(OR, AND, XOR)} \qquad \text{E0.codigo = E1.codigo +} \\ \text{E2.codigo +} \\ \text{generarCI('E0.dir = E1.dir OP E2.dir')} \\ \}
```

8.3.2 (cont...)

```
E \to Eop_relE
                     { E0.dir = crearTemporal()
(=, <, >, !=, ...)
                      ETQ_TRUE = nuevaEtiqueta()
                      ETQ_FIN = nuevaEtiqueta()
                      E0.codigo = E1.codigo +
                          E2.codigo +
                          generarCI('if (E1.dir OPREL E2.dir) goto ETQ_TRUE')
                          generarCI('E0.dir = 0')+
                          generarCI('goto ETQ_FIN)+
                          generarCI('ETQ_TRUE:')+
                          generarCI('E0.dir = 1')+
                          generarCI('ETQ_FIN:')
                     }
   E \to {\rm true}
                     { E.dir = crearTemporal()
                       E.codigo = generarCI('E.dir = 1')
                     }
   E \rightarrow \mathsf{false}
                     { E.dir = crearTemporal()
                       E.codigo = generarCI('E.dir = 0')
                     }
```

8.3.3 Instrucciones de control de flujo

Versión sencilla, considerando expr. booleanas con valores numéricos Uso de etiquetas para dar soporte al control de flujo en el C.I.

SALTOS

```
if expresion then sentencias
if expresion then sentencias1 else sentencias2
```

BUCLES

```
while expresion do sentencias
```

for identificador := expresion1 **to** expresion2 **do** sentencias

EXTENSIONES

```
begin
    case valor1 : sentencias1
    case valor2 : sentencias2
    ...
    case valorN : sentenciasN
    default sentenciasD
    end
```

8.3.3 (cont...) SALTOS if - then E.codigo (evaluar E E.codigo en temp) if temp goto S1_inicio (evaluar E en temp) if temp goto S1_inicio S2.codigo goto S0_final S1_inicio: goto S0_final S1.codigo S1_inicio: S1.codigo S0_final: S0_final: $S \to \text{if } E \text{ then } S$ { S1_inicio = nuevaEtiqueta() SO_final = nuevaEtiqueta() S0.codigo = E.codigo+ generarCI('if E.dir goto S1.inicio')+ generarCI('goto S0_final)+ generarCI('S1_inicio:')+

S1.codigo+

}

generarCI('SO_final:')

$S \to \text{if } E \text{ then } S \text{ else } S$

```
{ S1_inicio = nuevaEtiqueta()
   S0_final = nuevaEtiqueta()
   S0.codigo = E.codigo +
        generarCI('if E.dir goto S1_inicio')+
        S2.codigo+
        generarCI('goto S0_final)+
        generarCI('S1_inicio:')+
        S1.codigo+
        generarCI('S0_final:')
}
```

8.3.3 (cont...) BUCLES

for ascendente (usa nueva etiqueta TEST) while E1.codigo (límite inferior) E.codigo S0_inicio: E2.codigo (evaluar E (límite superior) en temp) id = E1if temp goto S1_inicio if (id <= E2) goto S1_inicio S0_test: goto S0_final goto S0_final S1_inicio: S1_inicio: S1.codigo S1.codigo goto S0_inicio id = id + 1S0_final: goto S0_test S0_final: $\overline{S \to \text{while } E \text{ do } S}$ { SO_inicio = nuevaEtiqueta() S0_final = nuevaEtiqueta() S1_inicio = nuevaEtiqueta() S0.codigo = generarCI('S0_inicio:')+ E.codigo+ generarCI('if E.dir goto S1_inicio')+ generarCI('goto S0_final)+ generarCI('S1_inicio:')+ S1.codigo+ generarCI('goto S0_inicio)+ generarCI('SO_final:') }

```
S \rightarrow \text{for } id = E \text{ to } E \text{ do } S
                              { SO_inicio = nuevaEtiqueta()
                                S0_final = nuevaEtiqueta()
                                S0_test
                                           = nuevaEtiqueta()
                                S1_inicio = nuevaEtiqueta()
                                S0.codigo = generarCI('S0_inicio:')+
                                    E1.codigo +
                                   E2.codigo +
                                    generarCI('id.dir = E1.dir')
                                    generarCI('S0_test:')+
                                    generarCI('if (id.dir <= E2.dir)</pre>
                                                        goto S1_inicio')+
                                    generarCI('goto S0_final)+
                                    generarCI('S1_inicio:')+
                                    S1.codigo+
                                    generarCI('id.dir = id.dir + 1')+
                                    generarCI('goto S0_test)+
                                    generarCI('SO_final:')
                              }
```

8.3.3 (cont...) CASOS ESPECIALES Alteraciones de flujo dentro de bucles.

Salida: (exit, break) goto a la etiqueta S0.final

Cancelar iteración: (continue)

→ en **while**: **goto** al incio de la expresión de comparación (*S0.inicio*)

→ en **for**: **goto** a la instrucción de incremento

Instrucción CASE

S0.inicio:

E.codigo (evaluación expresión)

goto S0.test

S1.inicio:

S1.codigo

goto S0.final

switch expresion

begin

case valor1 : sentencias1

case valor2 : sentencias2

. . .

case valorN : sentenciasN

default sentenciasD

end

S2.inicio:

SN.inicio:

SN.codigo

goto S0.final

SD.inicio:

SD.codigo

goto S0.final

S0.test:

if E.dir = valor1 goto S1.inicio

if E.dir = valor2 goto \$2.inicio

...

if E.dir = valorN goto SN.inicio

goto SD.inicio

S0.final ...

NOTAS

- En C, si no se incluye break, continúa la ejecución en el siguente caso.
 - ightarrow se incluirá el goto S0.final en los casos donde exista ${f break}$
 - → consistente con comportamiento de **break** en **while** y **for** indicado anteriormente
- Si hubiera muchos casos, usar tabla de pares (caso, etiqueta)
 - ightarrow la selección de caso se haría dentro de un bucle que iteraría sobre esa tabla

8.3.4 Procedimientos y Funciones

	S0.inicio:	E1.codigo
Esquema de llamada		(evaluación
		parametro1)
<pre>procedimeinto(E, E,, E);</pre>		E2.codigo
		(evaluación
		parametro2)
Figure		
Ejemplos		EN.codigo
		(evaluación
<pre>imprimir(nombre, DNI, edad);</pre>		parametroN)
mostrar_resultados;		param E1.dir
b := media(A[5], velocidad*15):		param E2.dir
b . modia(n[o], volocidad 10).		
		param EN.dir
		call procedimiento
	S0.final	•••

El uso de C.I. oculta las operacioens adicionales necesarias en la llamada y el retorno de procedimientos.

LLAMADA

- Evaluar valores de los parámetros
- Crear el registro de actvación asociado al nuevo procedimiento
- Paso de parámetros
- Guardar estado del procedimeinto actual
- Guardar dirección de retorno
- Salto al inicio del código del procedmiento llamado

RETORNO

- Recuperar estado del procedimiento llamador
- Salto a siguiente instrucción
- Recuperar valor devuelto

Secuencias de llamada/retorno: alta dependencia de la máquina objetivo

convenciones de llamada + instrucciones específicas