

Bài 10

Phân tích ngữ nghĩa

Nội dung

- **Những vấn đề ngữ nghĩa**
- **Kiểm tra kiểu (Type checking)**
 - Hệ thống kiểu trong ngôn ngữ lập trình
 - Đặc tả một bộ kiểm tra kiểu
 - Chuyển đổi kiểu
- **Bảng ký hiệu**
 - Luật về phạm vi ảnh hưởng của biến
 - Các sơ đồ dịch để xây dựng bảng ký hiệu

Phân tích ngữ nghĩa

- Tìm ra các lỗi sau giai đoạn phân tích cú pháp
 - Kiểm tra sự tương ứng về kiểu
 - Kiểm tra sự tương ứng giữa việc sử dụng hàm, biến với khai báo của chúng
 - Xác định phạm vi ảnh hưởng của các biến trong chương trình
- Phân tích ngữ nghĩa thường sử dụng cây cú pháp

Khái niệm kiểm tra kiểu

- Kiểm tra xem chương trình có tuân theo các luật về kiểu của ngôn ngữ không
- Trình biên dịch quản lý thông tin về kiểu
- Việc kiểm tra kiểu được thực hiện bởi bộ kiểm tra kiểu (type checker), một bộ phận của trình biên dịch

Ví dụ về kiểm tra kiểu

- Toán tử % của C chỉ thực hiện khi các toán hạng là số nguyên
- Chỉ có mảng mới có chỉ số và kiểu của chỉ số phải nguyên
- Một hàm phải có một số lượng tham số nhất định và các tham số phải đúng kiểu

Kiểm tra kiểu

- Có hai phương pháp tĩnh và động
- Phương pháp áp dụng trong thời gian dịch là tĩnh
- Trong các ngôn ngữ như C hay Pascal, kiểm tra kiểu là tĩnh và được dùng để kiểm tra tính đúng đắn của chương trình trước khi nó được thực hiện
- Kiểm tra kiểu tĩnh cũng được sử dụng khi xác định dung lượng bộ nhớ cần thiết cho các biến
- Bộ kiểm tra kiểu được xây dựng dựa trên
 - Các biểu thức kiểu của ngôn ngữ
 - Bộ luật để định kiểu cho các cấu trúc

Biểu thức kiểu (Type Expression)

Biểu diễn kiểu của một cấu trúc ngôn ngữ
 Một biểu thức kiểu là một kiểu dữ liệu chuẩn hoặc được xây dựng từ các kiểu dữ liệu khác bởi cấu trúc kiểu (*Type Constructor*)

1. Kiểu dữ liệu chuẩn (int, real, boolean, char) là biểu thức kiểu
2. Biểu thức kiểu có thể liên hệ với một tên. Tên kiểu là biểu thức
3. Cấu trúc kiểu được ứng dụng vào các biểu thức kiểu tạo ra biểu thức kiểu

Cấu trúc kiểu

(a) Mảng (*Array*). Nếu T là biểu thức kiểu thì $array(l, T)$ là biểu thức kiểu biểu diễn một mảng với các phần tử kiểu T và chỉ số trong miền l

Ví dụ : `array [10] of integer` có kiểu `array(1..10, int);`

(b) Tích Descartes. Nếu T_1 và T_2 là các biểu thức kiểu thì tích Descartes $T_1 \times T_2$ là biểu thức kiểu

(c) Bản ghi (*Record*). Tương tự như tích Descartes nhưng chứa các tên khác nhau cho các kiểu khác nhau,

Ví dụ

```

struct
{
  double r;
  int i;
}
Có kiểu ((r x int) x (i x char))
  
```

Cấu trúc kiểu (tiếp)

(d) *Con trỏ*: Nếu T là biểu thức kiểu thì $\text{pointer}(T)$ là biểu thức kiểu

(e) *Hàm*: Nếu D là miền xác định và R là miền giá trị của hàm thì kiểu của nó được biểu diễn là $\text{biểu_thức} : D : R$.

Ví dụ hàm của C

`int f(char a, b)`

Có kiểu: `char × char : int`.

Hệ thống kiểu (Type System)

- Tập các luật để xây dựng các biểu thức kiểu trong những phần khác nhau của chương trình
- Được định nghĩa thông qua định nghĩa tựa cú pháp
- Bộ kiểm tra kiểu thực hiện một hệ thống kiểu
- Ngôn ngữ định kiểu mạnh: Chương trình dịch kiểm soát được hết các lỗi về kiểu

Đặc tả một bộ kiểm tra kiểu

- Ngôn ngữ đơn giản với mỗi tên được liên kết với một kiểu
- Văn phạm thuộc tính để khai báo biểu thức

$P \rightarrow D; E$

$D \rightarrow D; D \mid \text{id} : T$

$T \rightarrow \text{char} \mid \text{int} \mid \text{array}[\text{num}] \text{ of } T \mid \uparrow T$

$E \rightarrow \text{literal} \mid \text{num} \mid \text{id} \mid E \text{ mod } E \mid E[E] \mid E \uparrow$

Thuộc tính

- Thuộc tính là khái niệm trừu tượng biểu diễn một đại lượng bất kỳ, chẳng hạn một số, một xâu, một vị trí trong bộ nhớ
- Thuộc tính được gọi là **tổng hợp** nếu giá trị của nó tại một nút trong cây được xác định từ giá trị của các nút con của nó.
- Thuộc tính **kế thừa** là thuộc tính tại một nút mà giá trị của nó được định nghĩa dựa vào giá trị nút cha và/hoặc các nút anh em của nó.

Định nghĩa tựa cú pháp (syntax directed definition)

Định nghĩa tựa cú pháp là dạng tổng quát của văn phạm phi ngữ cảnh để đặc tả cú pháp của ngôn ngữ vào.

- Mỗi ký hiệu của văn phạm liên kết với một tập thuộc tính.
- Mỗi sản xuất $A \rightarrow \alpha$ liên hệ với một tập các quy tắc ngữ nghĩa để tính giá trị thuộc tính liên kết với những ký hiệu xuất hiện trong sản xuất. Tập các quy tắc ngữ nghĩa có dạng

$$b = f(c_1, c_2, \dots, c_n)$$

f là một hàm và b thoả một trong hai yêu cầu sau:

- b là một thuộc tính tổng hợp của A và c_1, \dots, c_n là các thuộc tính liên kết với các ký hiệu trong vế phải sản xuất $A \rightarrow \alpha$.
- b là một thuộc tính thừa kế một trong những ký hiệu xuất hiện trong α , và c_1, \dots, c_n là thuộc tính của các ký hiệu trong vế phải sản xuất $A \rightarrow \alpha$.

Ví dụ

Sản xuất	Quy tắc ngữ nghĩa
$L \rightarrow E \text{ return}$	Print (E.val)
$E \rightarrow E1 + T$	$E.val = E1.val + T.val$
$E \rightarrow T$	$E.val = T.val$
$T \rightarrow T1 * F$	$T.val = T1.val * F.val$
$T \rightarrow F$	$T.val = F.val$
$F \rightarrow (E)$	$F.val = E.val$
$F \rightarrow \text{digit}$	$F.val = \text{digit.lexval}$

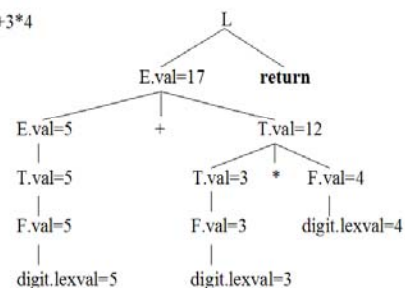
- Các ký hiệu E, T, F liên hệ với thuộc tính tổng hợp val
- Từ tổ *digit* có thuộc tính tổng hợp $lexval$ (Được bộ phân tích từ vựng đưa ra)

Cây phân tích cú pháp có chú giải

- Cây cú pháp có chỉ ra giá trị các thuộc tính tại mỗi nút được gọi là cây cú pháp có chú giải.

Ví dụ

Input: 5+3*4



Bộ kiểm tra kiểu của định danh

SẢN XUẤT	QUY TẮC NGŨ NGHĨA
$D \rightarrow \text{id} : T$	$\text{addtype}(\text{id.entry}, T.\text{type})$
$T \rightarrow \text{char}$	$T.\text{type} := \text{char}$
$T \rightarrow \text{int}$	$T.\text{type} := \text{int}$
$T \rightarrow \uparrow T_1$	$T.\text{type} := \text{pointer}(T_1.\text{type})$
$T \rightarrow \text{array}[\text{num}] \text{ of } T_1$	$T.\text{type} := \text{array}(1..\text{num.val}, T_1.\text{type})$

Bộ kiểm tra kiểu của biểu thức

SẢN XUẤT	QUY TẮC NGŨ NGHĨA
$E \rightarrow \text{literal}$	$E.\text{type} := \text{char}$
$E \rightarrow \text{num}$	$E.\text{type} := \text{int}$
$E \rightarrow \text{id}$	$E.\text{type} := \text{lookup}(\text{id.entry})$
$E \rightarrow E_1 \text{ mod } E_2$	$E.\text{type} := \text{if } E_1.\text{type} = \text{int and } E_2.\text{type} = \text{int} \text{ then } \text{int} \text{ else } \text{type_error}$
$E \rightarrow E_1[E_2]$	$E.\text{type} := \text{if } E_2.\text{type} = \text{int and } E_1.\text{type} = \text{array}(s, t) \text{ then } t \text{ else } \text{type_error}$
$E \rightarrow E_1 \uparrow$	$E.\text{type} := \text{if } E_1.\text{type} = \text{pointer}(t) \text{ then } t \text{ else } \text{type_error}$

Bộ kiểm tra kiểu của lệnh

SẢN XUẤT	QUY TẮC NGŨ NGHĨA
$S \rightarrow \text{id} := E$	$S.\text{type} := \text{if id.type} = E.\text{type} \text{ then } \text{void} \text{ else } \text{type_error}$
$S \rightarrow \text{if } E \text{ then } S_1$	$S.\text{type} := \text{if } E.\text{type} = \text{boolean} \text{ then } S_1.\text{type} \text{ else } \text{type_error}$
$S \rightarrow \text{while } E \text{ do } S_1$	$S.\text{type} := \text{if } E.\text{type} = \text{boolean} \text{ then } S_1.\text{type} \text{ else } \text{type_error}$
$S \rightarrow S_1; S_2$	$S.\text{type} := \text{if } S_1.\text{type} = \text{void and } S_2.\text{type} = \text{void} \text{ then } \text{void} \text{ else } \text{type_error}$

Bộ kiểm tra kiểu của hàm

SẢN XUẤT	QUY TẮC NGŨ NGHĨA
$D \rightarrow \text{id} : T$	$\text{addtype}(\text{id.entry}, T.\text{type}); D.\text{type} := T.\text{type}$
$D \rightarrow D_1; D_2$	$D.\text{type} := D_1.\text{type} \times D_2.\text{type}$
$\text{Fun} \rightarrow \text{fun id}(D) : T; B$	$\text{addtype}(\text{id.entry}, D.\text{type}: T.\text{type})$
$B \rightarrow \{S\}$	
$S \rightarrow \text{id}(E\text{List})$	$E.\text{type} := \text{if } \text{lookup}(\text{id.entry}) = t_1 : t_2 \text{ and } E\text{List.type} = t_1 \text{ then } t_2 \text{ else } \text{type_error}$
$E\text{List} \rightarrow E$	$E\text{List.type} := E.\text{type}$
$E\text{List} \rightarrow E\text{List}, E$	$E\text{List.type} := E\text{List.type} \times E.\text{type}$

Hàm kiểm tra biểu thức kiểu tương đương

```
function sequiv(s, t): boolean;
begin
  if s và t là cùng kiểu dữ liệu chuẩn then
    return true;
  else if s = array(s1, s2) and t = array(t1, t2) then
    return sequiv(s1, t1) and sequiv(s2, t2);
  else if s = s1 x s2 and t = t1 x t2 then
    return sequiv(s1, t1) and sequiv(s2, t2);
  else if s = pointer(s1) and t = pointer(t1) then
    return sequiv(s1, t1);
  else if s = s1 → s2 and t = t1 → t2 then
    return sequiv(s1, t1) and sequiv(s2, t2);
  else
    return false;
end;
```

Chuyển đổi kiểu

■ Kiểu của x+i với

x kiểu real

i kiểu int

Khi dịch sang lệnh máy, phép cộng với kiểu real và kiểu int có mã lệnh khác nhau

■ Tùy ngôn ngữ và bộ luật chuyển đổi sẽ quy đổi các toán hạng về một trong hai kiểu

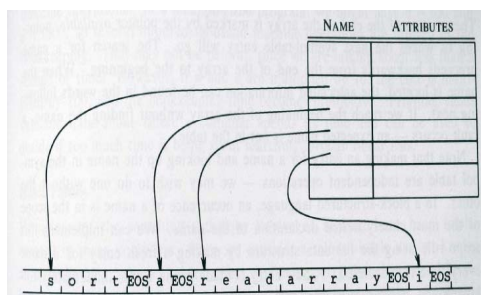
Áp đặt kiểu trong biểu thức

SẢN XUẤT	QUY TẮC NGŨNGHĨA
$E \rightarrow \text{num}$	$E.type := \text{int}$
$E \rightarrow \text{num.num}$	$E.type := \text{real}$
$E \rightarrow \text{id}$	$E.type := \text{lookup}(\text{id.entry})$
$E \rightarrow E_1 \text{ op } E_2$	$E.type := \text{if } E_1.type = \text{int and } E_2.type = \text{int}$ then <i>int</i> else if $E_1.type = \text{int and } E_2.type = \text{real}$ then <i>real</i> else if $E_1.type = \text{real and } E_2.type = \text{int}$ then <i>real</i> else if $E_1.type = \text{real and } E_2.type = \text{real}$ then <i>real</i> else <i>type_error</i>

Bảng ký hiệu

- Lưu trữ thông tin về tên, kiểu, phạm vi và kích cỡ bộ nhớ cần phân phối
- Bổ sung dữ liệu và tìm kiếm nhanh chóng
- Thích hợp với các cấu trúc dữ liệu động
 - Thường dùng danh sách tuyến tính và bảng băm
 - Mỗi lối vào có dạng bản ghi với một trường cho mỗi loại thông tin

Lưu trữ tên

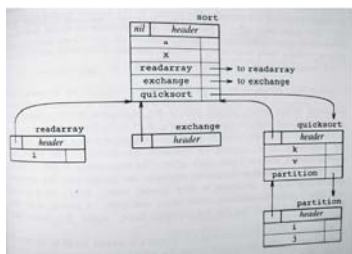


Bảng ký hiệu và các luật về phạm vi sử dụng

- Khối trong ngôn ngữ lập trình là tập các cấu trúc ngôn ngữ có chứa khai báo
- Một ngôn ngữ là có cấu trúc khối nếu
 - Các khối được lồng bên trong những khối khác
 - Phạm vi của khai báo trong mỗi khối là chính khối đó và các khối chứa trong nó
- Luật lồng nhau gần nhất
 - Cho nhiều khai báo của cùng một tên. Khai báo có hiệu lực là khai báo nằm trong khối gần nhất

Giải pháp nhiều bảng ký hiệu

- Các bảng cần được kết nối từ phạm vi trong ra phạm vi ngoài và ngược lại



Các luật về phạm vi lồng nhau

- Toán tử insert vào bảng ký hiệu không được ghi đè những khai báo trước
- Toán tử lookup vào bảng ký hiệu luôn luôn tham chiếu luật phạm vi gần nhất
- Toán tử delete chỉ được xóa khai báo gần nhất

Bảng ký hiệu hoạt động như một stack

Xây dựng trong giai đoạn phân tích cú pháp

- Chỉ có thể bắt đầu nhập thông tin vào bảng ký hiệu từ khi phân tích từ vựng nếu ngôn ngữ lập trình không cho khai báo tên trùng nhau.
- Nếu cho phép các phạm vi, bộ phân tích từ vựng chỉ trả ra tên của định danh cùng với token
 - Định danh được thêm vào bảng ký hiệu khi vai trò cú pháp của định danh được phát hiện

Tính địa chỉ biến toàn cục offset

```

P → {offset := 0} D
D → D; D
D → id : T      {enter(id.name, T.type, offset);
                  offset := offset + T.width}
T → int          {T.type := int; T.width := 4}
T → real         {T.type := real; T.width := 8}
T → array[num] of T1 {T.type := array(num.val, T1.type);
                     T.width := num.val * T1.width}
T → ↑T1         {T.type := pointer(T1.type); T.width := 4}

```

Địa chỉ liên quan

- Địa chỉ liên quan: Thông tin về phân phối bộ nhớ logic

- Biến toàn cục offset lưu vết của địa chỉ còn rồi tiếp theo
- Trước khai báo đầu tiên, offset được gán giá trị 0.
- Mỗi khi tìm thấy một định danh, offset được tăng lên

Lưu trữ thông tin về các phạm vi lồng nhau

- Xét các thủ tục lồng nhau: Khi một thủ tục nằm trong thủ tục khác được gọi, các khai báo của thủ tục bên ngoài tạm dừng hoạt động
- Dùng stack để lưu trữ dấu vết của các thủ tục lồng nhau
- Tạo bảng ký hiệu mới cho mỗi thủ tục
 - Khi thêm một định danh mới vào bảng ký hiệu, cần chỉ rõ bảng ký hiệu cần thêm

Xử lý các khai báo trong thủ tục lồng nhau

```

P → M D      {addwidth(top(tblptr),top(offset));
                pop(tblptr); pop(offset)}
M → ε         {t:=mktable(nil);
                push(t,tblptr); push(0,offset)}

D → D; D
D → proc id; N D1; S {t:=top(tblptr); addwidth(t,top(offset));
                    pop(tblptr); pop(offset);
                    enterproc(top(tblptr),id.name,t)}
D → id : T     {enter(top(tblptr),id.name, T.type, top(offset));
                top(offset) := top(offset) + T.width}
N → ε         {t:=mktable(top(tblptr));
                push(t,tblptr); push(0,offset)}

```

Lưu trữ dấu vết của các phạm vi

- *mktable(previous)* Tạo một bảng ký hiệu mới và trả lại con trỏ của bảng ký hiệu đó. Tham số *previous* là con trỏ tới thủ tục chứa nó.
- Stack *tblptr* chứa con trỏ tới các bảng ký hiệu và các thủ tục chứa nó.
- Stack *offset* chứa dấu vết các địa chỉ tương ứng ở mức lồng nhau nào đó
- *enter(table,name,type,offset)* tạo một lối vào mới cho định danh *name* trong bảng ký hiệu mà *table* trỏ tới, đồng thời chỉ ra các thuộc tính *type* và *offset*.
- *addwidth(table,width)* ghi giá trị *width* của mỗi lối vào trong *table* vào *header* của bảng ký hiệu
- *enterproc(table,name,newtable)* Tạo một lối vào mới cho thủ tục *name* trong bảng ký hiệu chỉ bởi *table*. Tham số *newtable* chỉ tới bảng ký hiệu cho thủ tục *name*.