

Especificación de un Lenguaje de Programación Mínimo y Construcción de su Procesador (3)

Procesadores de Lenguaje

Tabla de Símbolos

Contenidos

(Capítulos 1 y 2 de Aho et al.)

1. Especificación de un lenguaje de programación mínimo y construcción de su procesador.

- Introducción a los lenguajes de programación y a los procesadores de lenguajes: el modelo análisis/síntesis.
- Elementos de la teoría de lenguajes formales. Introducción a las gramáticas de atributos: definiciones dirigidas por la sintaxis y semántica de los lenguajes de programación.
- Definición del lenguaje fuente:
 - aspectos léxicos
 - aspectos sintácticos
 - **tabla de símbolos**
 - restricciones contextuales.

Definición del Lenguaje Fuente:

la Tabla de Símbolos

- **Justificación**
- **Estructura**
- **Especificación**

(en el libro de Aho et al., introducción en 2.7 y ampliación en 7.6 Ejemplos 5.3 y 5.6 en sección 5.1)

3

Definición del Lenguaje Fuente

Justificación de la Tabla de Símbolos

- Desde nuestro punto de vista podemos considerar que los lenguajes tienen asociados dos alfabetos o vocabularios distintos:
 - a) un conjunto de (pocos) símbolos (los códigos ASCII en los lenguajes de programación o las letras del alfabeto en los lenguajes naturales)
 - b) un segundo conjunto de símbolos en que se basa el significado del lenguaje (del orden de decenas, o miles, según se trate de un lenguaje de programación o natural) cuyos elementos se generan componiendo los símbolos del conjunto anterior.

Para los procesadores de lenguaje, este segundo conjunto es el vocabulario del lenguaje fuente y está formado por los *símbolos*, *tokens*, o palabras, a los que se refiere el diccionario/tabla de símbolos del lenguaje fuente.

ejercicio: identificar los dos alfabetos en varias frases tomadas de lenguajes de programación, ejemplo:

IF velocidad > LTE THEN puntos:=puntos-2 ELSE enviar_saludo(matricula);

4

Definición del Lenguaje Fuente

Justificación de la Tabla de Símbolos

- El significado (la semántica) de las frases de un lenguaje se puede derivar del significado individual de cada uno de los símbolos o *tokens* de su vocabulario.
- El diccionario es la estructura que recoge el significado de los símbolos individuales de un lenguaje.

discusión:

- *buscar ejemplos en los que la primera afirmación no sea cierta*
- *en el diccionario no están todos los símbolos del lenguaje ¿porqué? ¿algún ejemplo?*

5

Definición del Lenguaje Fuente

Justificación de la Tabla de Símbolos

- En los lenguajes de programación el diccionario tiene características especiales y recibe el nombre de tabla de símbolos

discusión: comparar diccionarios y tablas de símbolos

- Los lenguajes de programación pueden incluir mecanismos sintácticos para definir nuevos tipos de símbolos y nuevos símbolos para una frase concreta del lenguaje. Son los sub-lenguajes de declaración.

p.e.

TYPE producto: RECORD

inventario: Integer;

precio: Real;

END

VAR X,Y: Integer; P: producto

Define el tipo producto y declara que las variables X , Y son de tipo entero y la variable P de tipo producto.

6

Definición del Lenguaje Fuente

Justificación de la Tabla de Símbolos

- El significado/función de los símbolos o *tokens* puede ser de naturaleza muy distinta:
 - información sobre los objetos del dominio del lenguaje representados por los símbolos
 - operadores para componer el significado atribuido a estos objetos
 - marcas de puntuación (ayudan a reconocer las subestructuras de una frase), etc.

ejercicio: utilizar ejemplos de frases de un lenguaje de programación con identificadores, palabras reservadas, operadores etc. y describir el significado asociado a cada uno de los símbolos en las frases

7

Definición del Lenguaje Fuente

Justificación de la Tabla de Símbolos

- En los lenguajes de programación, el significado que se atribuye a los *tokens* del lenguaje se expresa mediante un conjunto de atributos asociado a cada símbolo o *token*:
 - el valor del símbolo en un dominio concreto (e.g. el valor 255)
 - la forma de codificarlo/utilizarlo (e.g. entero, real)
 - la clase de símbolo (e.g. constante, variable, acción/procedimiento, etc.)
 - la secuencia de caracteres básicos que lo distinguen de otros símbolos de su misma categoría (el lexema)
 - un localizador de la memoria de una máquina real o virtual (la dirección de memoria) donde se almacena el valor (*dir*) o el código (*etq*) de la acción asociada al símbolo
 - si el símbolo es compuesto, las referencias a los localizadores de sus componentes (e.g. *arrays*, *records*, *clases*, ...)

- *ejercicio: utilizar atributos para dar significado a los símbolos de los ejemplos anteriores*

8

Definición del Lenguaje Fuente

Justificación de la Tabla de Símbolos

- Algunos símbolos o *tokens* tienen características especiales, p.e. los números, ya que su significado, su valor, puede deducirse de la secuencia de elementos del alfabeto básico (dígitos) que lo componen.
- Normalmente no suelen formar parte de la tabla de símbolos los símbolos que tienen un significado intrínseco; p.e. un operador, o una marca de puntuación (*¿alguna excepción?*)
- En algunos casos el significado de estos símbolos puede depender del contexto en que se utilizan los mismos (el mismo símbolo puede tener significados diferentes en contextos diferentes).

9

Definición del Lenguaje Fuente

Justificación de la Tabla de Símbolos

- La tabla de símbolos se construye durante el proceso de análisis (*¿se sintetiza durante el análisis!*)
- La información en la tabla de símbolos se utiliza tanto en el análisis,
 - para especificar restricciones contextuales,
- como en la síntesis/traducción,
 - para interpretar/traducir el significado de los *tokens*.

10

Definición del Lenguaje Fuente

Estructura de la Tabla de Símbolos

- En el caso de los lenguajes de programación la estructura de este diccionario -la tabla de símbolos- tiene dos partes:
 - Una parte estática formada por las palabras clave del lenguaje (suelen ser de uso reservado y del orden de unas decenas de palabras en un lenguaje de programación típico),
 - Una parte dinámica definida por el programador: con el significado de los identificadores utilizados en una frase concreta (un programa).
- Inicialmente (en el lenguaje ejemplo) la estructura de la tabla de símbolos puede ser muy sencilla:
 - una tabla relacionando el código de cada *token*, o elemento del vocabulario del lenguaje fuente, con sus atributos: el lexema y el resto de la información semántica del símbolo.

Definición del Lenguaje Fuente

Estructura de la Tabla de Símbolos

- Cuando el lenguaje da facilidades lingüísticas al programador para definir nuevos símbolos cuyo significado se obtiene componiendo otros símbolos previamente definidos, la estructura de la tabla de símbolos es más compleja.
- Estas facilidades lingüísticas implican la utilización de categorías sintácticas que pueden mapearse, mediante una definición dirigida por sintaxis, a un conjunto de constructores básicos: producto cartesiano y funciones específicas (*array*, *record*, etc.)
- *discutir posibles estructuras para la tabla de símbolos de los ejemplos anteriores.*

Definición del Lenguaje Fuente

Estructura de la Tabla de Símbolos

- Otra fuente de complejidad de la estructura de la tabla de símbolos es cuando las facilidades lingüísticas del lenguaje incluyen poder organizar las frases del lenguaje - los programas- en módulos o bloques, anidados o no, y cada uno con un vocabulario específico del módulo y otro global compartido con todos, o con algunos, módulos del programa.
- En este caso lo normal es no tener una tabla de símbolos, si no un conjunto de tablas relacionadas de la misma forma que los módulos.

13

Definición del Lenguaje Fuente

Especificación de la Tabla de Símbolos

- La construcción de la tabla de símbolos debe seguirse de una correcta especificación de la misma
- Gramáticas de atributos y esquemas de traducción.
 - Como la definición de la parte dinámica de la tabla de símbolos está basada en las facilidades lingüísticas (la gramática) del lenguaje de programación, la especificación de la construcción de la tabla de símbolos está dirigida por la sintaxis del sub-lenguaje de declaración y utiliza las mismas técnicas ya utilizadas para especificar y construir otros elementos del procesador de lenguaje
 - La construcción de la tabla de símbolos es una tarea del módulo de análisis sintáctico en la que coopera inicialmente el módulo de análisis léxico
- Las ecuaciones semánticas de la gramática de atributos utilizan un conjunto de funciones (tipo abstracto de datos) que permiten crear, acceder y modificar la tabla de símbolos.

14

Definición del Lenguaje Fuente

Estructura y Especificación de la Tabla de Símbolos

- **Lenguaje ejemplo:**
 - Definición de la tabla de símbolos para el lenguaje ejemplo.
 - Definiciones léxicas y sintácticas.
 - Gramática de atributos: especificación de la construcción de la tabla de símbolos para el lenguaje ejemplo.

15

Definición del Lenguaje Fuente

Estructura y Especificación de la Tabla de Símbolos

Lenguaje Ejemplo

- Extensión del lenguaje ejemplo de las expresiones aritméticas.
- Los programas en el nuevo lenguaje son secuencias de sentencias de asignación.
- Cada sentencia de asignación es del tipo *variable := expresión*.
- Las expresiones pueden contener variables (sub-lenguaje de expresiones).
- Las variables deben haber sido previamente declaradas (sub-lenguaje de declaración).
- Ejemplo de programa en el nuevo lenguaje:

```
tiempo; espacio; velocidad
&
espacio := 25;
tiempo := 30;
velocidad := espacio / tiempo
```

16

Definición del Lenguaje Fuente

Estructura y Especificación de la Tabla de Símbolos

Lenguaje Ejemplo

- **Definición léxica**

```
parte-entera = 0|[1-9][0-9]*
parte-decimal = .[0-9]+
número = {parte-entera}{parte-decimal}?
SUMA = \+
RESTA = \-
MUL = \*
DIV = /
PAP = \(
PCIERRE = \)
letra = [a-z]|[A-Z]
dígito = [0-9]
iden = {letra}({letra}|{dígito})*
asig = :=
SEP = &
PYCOMA = ;
```

17

Definición del Lenguaje Fuente

Estructura y Especificación de la Tabla de Símbolos

Lenguaje Ejemplo

- **Definición sintáctica**

```
Prog --> Decs & Is
Decs --> Dec | Decs ; Dec
Dec --> id
Is --> I | Is ; I
I --> id := Exp
Exp --> Exp OpAd Term
Exp --> Term
Term --> Term OpMul Fact
Term --> Fact
Fact --> numero | id | ( Exp )
OpAd --> + | -
OpMul --> * | /
```

18

Definición del Lenguaje Fuente

Estructura y Especificación de la Tabla de Símbolos

Lenguaje de Ejemplo

- La tabla de símbolos para este lenguaje es sencilla: una tabla con el conjunto de identificadores utilizados en el programa.
- Funciones básicas utilizadas en la definición dirigida por sintaxis (en las ecuaciones semánticas):

creaTS: crea una nueva TS que incluye las palabras reservadas del lenguaje (en los lenguajes organizados en bloques incluiría una referencia al contexto)

añadeID: El resultado es la nueva TS que resulta de añadir a la anterior el identificador *id* y sus atributos. Para obtener los atributos de *id* puede ser necesario procesar varias estructuras sintácticas

obtenerID: El resultado es una entrada a la TS donde está toda la información sobre *id* (*null* si *id* no está en la TS).

19

Definición del Lenguaje Fuente

Estructura y Especificación de la Tabla de Símbolos

Gramática de atributos

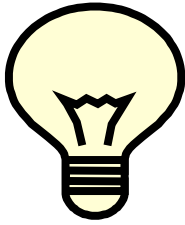
- Las categorías léxicas pueden tener atributos sintetizados.
- En concreto, se supondrá que:
 - todas las categorías léxicas que representan la clase de símbolos ID, los símbolos definidos por el programador para una frase concreta, tienen asociado un atributo sintetizado *lex*, cuyo valor es el lexema que permite identificar las diferentes ocurrencias del *token ID* en el programa.
 - los números tienen asociado un atributo *valor*
- En el caso de un lenguaje compilado a cada ID también se le asocia una dirección de memoria única.

20

Definición del Lenguaje Fuente

Estructura y Especificación de la Tabla de Símbolos

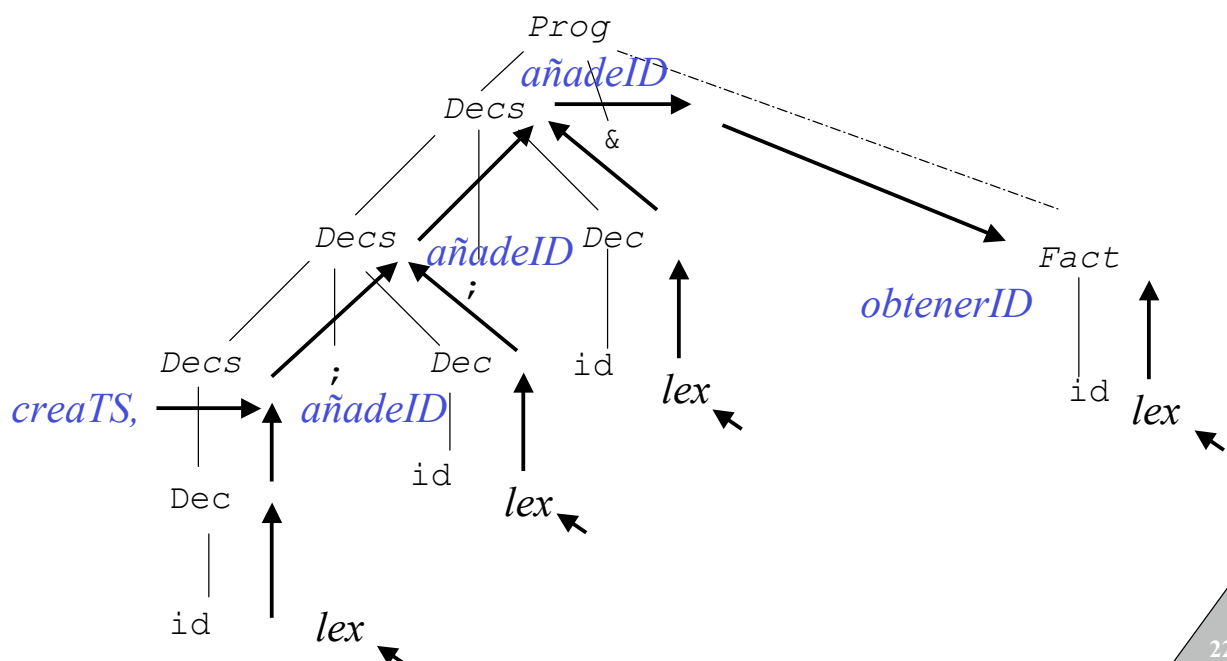
Construcción de la tabla de símbolos



Mediante una **gramática de atributos** se especifica como realizar la síntesis de la tabla de símbolos a partir del **sub-lenguaje** de las declaraciones.

21

Estructura y Especificación de la Tabla de Símbolos
Lenguaje ejemplo: grafo de dependencias utilizando atributos sintetizados para la creación de la TS



22

- 23



Definición del Lenguaje Fuente

Estructura y Especificación de la Tabla de Símbolos

Gramática de atributos: primera aproximación

Atributos de **Decs** (**secuencia de declaraciones**):

tablaSh: tabla de símbolos inicial
tablaS: tabla de símbolos sintetizada
dirh: dirección base de memoria libre
dir: dirección de memoria libre sintetizada

Atributos de **Is** (**secuencia de instrucciones**)

tablaSh: tabla de símbolos heredada

Atributos de **Dec** (**declara un identificador**)

lex: lexema del identificador

Atributos de **id** (**símbolo terminal**)

lex: lexema del identificador
(suministrado por el módulo de análisis léxico)

25

Definición del Lenguaje Fuente

Estructura y Especificación de la Tabla de Símbolos

Gramática de atributos: primera aproximación

Prog --> **Decs & Is**

Decs --> **Decs ; Dec**

Decs --> **Dec**

Decs.tablaSh = **creaTS()**

Decs.dirh = **DIR_BASE**

Decs.tablaS =

añadeID(Decs.tablaSh, Dec.lex, Decs.dirh)

Decs.dir = **Decs.dir + INCRMEM**

Dec --> **id**

Dec.lex = **id.lex**

Is --> **Is ; I**

...

26

Definición del Lenguaje Fuente
Estructura y Especificación de la Tabla de Símbolos
Gramática de atributos: primera aproximación

```
Prog--> Decs & Is
      Is.tablaSh = Decs.tabla

Decs --> Decs ; Dec
      Decs0.tablaS=
          añadeID(Decs1.tablaS,Dec.lex,Decs1.dir)
      Decs0.dir = Decs1.dir + INCRMEM

Decs --> Dec
      Decs.tablah = creaTS()
      ...
```

27

Definición del Lenguaje Fuente
Estructura y Especificación de la Tabla de Símbolos
Gramática de atributos

discusión:

criticar esta gramática y proponer una gramática más completa para el lenguaje de las instrucciones de asignación.

ejercicios:

- terminar de escribir la gramática de atributos dada***
- ampliar el lenguaje con palabras reservadas para mejorar la lectura de los programas e introducir los operadores enteros **div** , **mod** y especificar como se construiría la nueva tabla de símbolos***
- ampliar el lenguaje para poder expresar que en el dominio del lenguaje existen elementos de distinto tipo, p.e.: enteros, reales, caracteres, booleanos. Especificar como se construiría la nueva tabla de símbolos***

28