

Un ejercicio sobre análisis sintáctico LL(1)

IG29: Compiladores e intérpretes

Octava sesión de teoría

Vamos a definir un subconjunto reducido de las expresiones del lenguaje C y, para ello, comenzamos especificando los componentes léxicos que necesitaremos:

Categoría léxica	Expresión regular	Atributos	Acciones
blancos	$[\ \backslash t \backslash n]^+$	—	Omitir
comentario	$// [^ \backslash n]^*$	—	Omitir
opad	$[+-]$	<i>lex</i>	Copiar lexema en <i>lex</i> y emitir
asig	$=$	—	Emitir
int	$\backslash ?$	—	Emitir
dpt	$:$	—	Emitir
lit	$0 [1-9] [0-9]^*$	<i>val</i>	Calcular valor en <i>val</i> y emitir
ide	$[A-Za-z] [A-Za-z0-9_]^*$	<i>lex</i>	Copiar lexema en <i>lex</i> y emitir
ap	$\backslash ($	—	Emitir
cp	$\backslash)$	—	Emitir

La siguiente gramática incontextual G_0 , inutilizable en la práctica por su ambigüedad, es sin embargo perfectamente válida para especificar con precisión qué expresiones consideraremos sintácticamente válidas y cuáles no:

$$\begin{aligned} \langle E \rangle &\rightarrow \mathbf{lit} \mid \mathbf{ide} \\ \langle E \rangle &\rightarrow \mathbf{opad} \langle E \rangle \\ \langle E \rangle &\rightarrow \langle E \rangle \mathbf{opad} \langle E \rangle \\ \langle E \rangle &\rightarrow \langle E \rangle \mathbf{asig} \langle E \rangle \\ \langle E \rangle &\rightarrow \langle E \rangle \mathbf{int} \langle E \rangle \mathbf{dpt} \langle E \rangle \\ \langle E \rangle &\rightarrow \mathbf{ap} \langle E \rangle \mathbf{cp} \end{aligned}$$

Para poder interpretar correctamente la estructura lógica de las expresiones del subconjunto de C que estamos considerando, hace falta especificar, además, cuáles son los niveles de prioridad de algunos operadores¹. De mayor a menor prioridad, tenemos los siguientes: (1) signos más y menos; (2) suma y resta; (3) operador condicional; (4) asignación.

Con el objeto de integrar en el modelado sintáctico del lenguaje generado por G_0 las prioridades anteriores, se somete a nuestra consideración la siguiente GPDR, que llamaremos G_1 :

$$\begin{aligned} \langle E \rangle &\rightarrow \langle A \rangle (\mathbf{asig} \langle E \rangle)? \\ \langle A \rangle &\rightarrow \langle B \rangle (\mathbf{int} \langle B \rangle \mathbf{dpt} \langle E \rangle)^* \\ \langle B \rangle &\rightarrow \langle C \rangle (\mathbf{opad} \langle C \rangle)^* \\ \langle C \rangle &\rightarrow \mathbf{opad} \langle C \rangle \mid \mathbf{lit} \mid \mathbf{ide} \mid \mathbf{ap} \langle E \rangle \mathbf{cp} \end{aligned}$$

Las actividades que debes llevar a cabo son las siguientes:

A. Obtén a partir de G_1 una gramática incontextual equivalente G_2 utilizando las correspondientes transformaciones estándar:

$$\begin{aligned} (\alpha)^* &\equiv \langle X \rangle \rightarrow \alpha \langle X \rangle \mid \lambda \\ (\alpha)? &\equiv \langle X \rangle \rightarrow \alpha \mid \lambda \end{aligned}$$

B. Construye una tabla donde, para cada uno de los no terminales de tu nueva gramática G_2 se indique, en la correspondiente columna: (A) si el no terminal es o no anulable; (P) cuáles son sus primeros; (S) cuáles son sus siguientes.

C. A partir de la tabla anterior, construye la correspondiente tabla de análisis predictivo LL(1).

1. La asociatividad de los diferentes operadores (por la izquierda para suma, resta y condicional; por la derecha para la asignación) no es relevante para la resolución de este ejercicio.

D. Considera las tres expresiones C siguientes:

- `zzz2= -(x-7)`
- `x? (a=1): a=2 // error semántico, quizá`
- `1+ x? a=1: 2`

Para cada una de ellas:

- Obtén la correspondiente secuencia de categorías emitidas por el analizador léxico.
- Analiza con la gramática G_2 la secuencia anterior para determinar cuántos árboles de análisis tiene y cuáles son.

E. En vista de lo anterior, contesta a las tres siguientes preguntas sobre G_2 , justificando cada respuesta:

- ¿Genera el mismo lenguaje que G_0 ?
- ¿Es LL(1)?
- ¿Es ambigua?

F. Escribe una GPDR no ambigua G_3 que genere el mismo lenguaje que G_0 y refleje correctamente las prioridades de las diferentes operaciones.

Para contestar conjuntamente a los apartados B y C de este examen, puedes utilizar una tabla que siga esta estructura:

Tabla de análisis LL(1)												
	A	P	S	opad	asig	int	dpt	lit	ide	ap	cp	\$
(E)												
(A)												
(B)												
(C)												
()												
()												
()												