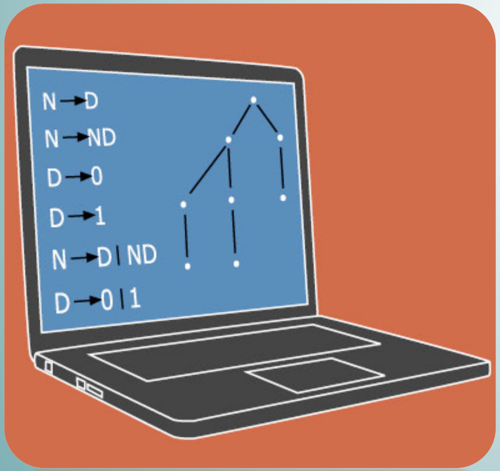
Procesadores de Lenguaje

Prefijos Viables



Cristina Tirnauca

Domingo Gómez Pérez

Dpto. de Matemáticas,

Estadística y Computación

Prefijos viables (I)

Posibles contenidos de la pila

Para cada regla de la gramática:

Hemos de decidir cuándo conviene usarla para reducir, es decir,

- sabiendo "lo que hay en la pila",
- y sabiendo cuál es el próximo símbolo de la entrada,

saber si hemos de reducir por esa regla o no.

Consideramos un paso intermedio de la derivación derecha: ¿qué es necesario conocer para saber cuál fue el paso anterior?

Dificultad: en caso de ambigüedad, es posible que haya varios candidatos a "paso anterior".

Definición de prefijos viables

Dado $\alpha\beta_1 \in (\Sigma \bigcup V)^*$ se dice que es prefijo viable de una gramática G si,

$$S \rightarrow_{rm} \ldots \rightarrow_{rm} \alpha A \omega \rightarrow_{rm} \alpha\beta_1\beta_2\omega.$$

Idea de LR

La idea del análisis ascendente expresada de otra forma es:

- Leer la palabra de izquierda a derecha.
- Encontrar el mayor prefijo viable (aquel, en el cual añadiendo un símbolo más deja de ser prefijo viable).
- Reducir por la regla adecuada.

No se esta diciendo que el prefijo viable sea único

Identifiquemos los prefijos viables

Derivaciones derechas con la gramática de las expresiones

Un ejemplo:

(Omitimos las aplicaciones de E:N y N:D para mayor claridad.)

- E
- E OP E
- E OP E OP E
- E OP E OP 3
- E OP E + 3
- E OP 2 + 3
- E + 2 + 3
- 1 + 2 + 3

- E
- E OP E
- E OP 3
- E + 3
- E OP E + 3
- E OP 2 + 3
- E + 2 + 3
- 1 + 2 + 3

Algunos prefijos viables son E OP 2, E +,E OP E OP E...

Más ejemplos de prefijos viables

Con una gramática de paréntesis

Otro ejemplo:

Secuencias bien parentizadas, con marca de fin de datos propia, no aceptamos λ, gramática no ambigua.

1. T
2. S ;
3. S A ;
4. S (A) ;
5. S (( )) ;
6. A (( )) ;
7. ( ) (( )) ;

Gramática:

- T : S ';'
- S : S A | A
- A : '(' A ')' | '(' ')'

Al no haber ambigüedad, no hay más que un "paso anterior" posible y, por tanto, la parte derecha "a reducir" es única.

Desplazar mientras nos encontremos en un prefijo viable.

Prefijos viables (IV)

La clave de bóveda

Teorema (Knuth, 1965):

Todo lenguaje libre de contexto determinista, en el que ninguna palabra sea prefijo de otra, admite una gramática en la cual el conjunto de prefijos viables para cada regla es un lenguaje regular.

Como siempre tendremos una marca de fin, en la práctica nunca una palabra aceptada es prefijo de otra. (Existe un teorema similar que no requiere esa condición.)

Consecuencia: Para cada regla, existe un autómata finito que puede recorrer la pila y decirnos si hay que reducir usándola.

Podemos combinarlos todos en uno solo, que recorre la pila y nos indica si hay que reducir o no, y por qué regla.

Ítems LR(0)

Qué estamos buscando y qué llevamos encontrado

Dada una gramática libre de contexto, un ítem LR(0) es un par formado por una regla de la gramática $A \to \alpha$ y un número natural entre 0 y $|\alpha|$.

Lo escribimos con frecuencia entre corchetes.

Cuando el número sea i, en vez de escribirlo explícitamente, pondremos un punto a continuación del i-ésimo símbolo de α:

$$[A \to \beta \bullet \gamma]$$

donde $\alpha = \beta\gamma$ y $|\beta| = i$.

Intuitivamente significa: Estamos buscando α y, tan pronto lo encontremos, reduciremos a A; y, de esa α que estamos buscando, llevamos visto β y nos falta por ver γ.

El Autómata LR(0)

Busca prefijos viables

Construimos un autómata indeterminista con λ-transiciones:

- Estados: todos los ítems LR(0).
- Estado inicial: $[T \to \bullet S]$, basado en la nueva regla inicial.
- Estados finales: los ítems de la forma $[A \to \alpha \bullet]$
- Transiciones:
  - De un ítem de la forma $[A \to \alpha \bullet a\gamma]$, donde a es un símbolo terminal, hay una transición a $[A \to \alpha a \bullet \gamma]$ etiquetada con a;
  - Del mismo modo, de un ítem de la forma $[A \to \alpha \bullet B\gamma]$, donde B es un símbolo gramatical no terminal, hay una transición a $[A \to \alpha B \bullet \gamma]$ etiquetada con B;
  - Además, de un ítem de la forma $[A \to \alpha \bullet B\gamma]$, donde B es un símbolo gramatical no terminal, hay una transición λ a cada uno de los ítems $[B \to \bullet\beta]$ para cada una de las reglas $B \to \beta$ de la gramática.

Y luego, determinizamos el autómata resultante.