

Ciencias de la Computación II

Propiedades de los LIC



Eduardo Contrera Schneider

Universidad de la Frontera

26 de octubre de 2016

Forma Normal de Chomsky

Cuando una gramática está en forma normal de Chomsky, podemos establecer una relación entre la longitud de sus cadenas y el número de pasos en su derivación. No es difícil probar por inducción que toda cadena de largo $|w|$ se puede derivar en exactamente $2|w|$ etapas.

Supongamos que G es una gramática independiente del contexto en forma normal de Chomsky y consideremos el árbol de derivación para una cadena cualquiera de $L(G)$. Hagamos las siguientes observaciones:

- Si un nodo tiene dos hijos, entonces los nodos hijos serán etiquetados con un no terminal y podremos tener al menos dos hijos más del mismo.
- En cada nivel se pueden duplicar el número de nodos con respecto al anterior, de manera que en el nivel k podríamos tener hasta 2^k nodos.
- Si un nodo tiene un único hijo, entonces dicho nodo hijo será etiquetado mediante un terminal.
- Si el camino más largo tiene $m + 2$ nodos y $m + 1$ aristas, entonces la cadena a lo mucho tiene longitud 2^m .

Lema del Bombeo

Sea L un lenguaje independiente del contexto que no contiene ϵ . Entonces existe un entero k para el cual, si $z \in L$ y $|z| > k$, entonces z se puede volver a escribir como $z = uvwxy$ con las siguientes propiedades:

- 1 $|vwx| \leq k$
- 2 Al menos o v o x no es ϵ .
- 3 $uv^iwx^{i^2}y \in L$ para todo $i \geq 0$.

Ejemplo

- Probemos que el lenguaje $L = \{a^i b^i \mid i \geq 0\}$ no es independiente del contexto.

Membresía

Dado un lenguaje independiente del contexto L sobre el alfabeto Σ y una cadena $w \in \Sigma^*$, ¿ $w \in L$ o no?

Lema

Sea $G = (N, \Sigma, N, P)$ una gramática independiente del contexto que no tiene producciones ϵ y que está en forma normal de Chomsky. Sea x una cadena de Σ^* . Se puede determinar, para cada $A \in N$ y para cadena subcadena w de x , si $A \Rightarrow^* w$.

Algoritmo CYK

Este algoritmo simplemente construye conjuntos N_{ij} de no terminales que generan las subcadenas w_{ij} de x . Una vez hecho, si $S \in N_{1n}$, entonces $x \in L(G)$ (donde $|x| = n$). Este algoritmo se enuncia como sigue:

CYK

- 1 Para cada $i = 1, 2, \dots, n$, sea

$$N_{i1} = \{A \mid A \rightarrow w_{i1}\}$$

- 2 Para $j = 2, 3, \dots, n$, hacer lo siguiente:

Para $i = 1, 2, \dots, n - j + 1$, hacer lo siguiente:

- 1 Inicializar $N_{ij} = \emptyset$.
- 2 Para $k = 1, 2, \dots, j - 1$, añadir a N_{ij} todos los no terminales A para los cuales $A \rightarrow BC$ con $B \in N_{ik}$ y $C \in N_{i+k, j-k}$

- 3 Si $S \in N_{1n}$ entonces $x \in L(G)$