TeoriaComputacion_2

October 30, 2019

Teoría de la Computación

Lenguajes y Gramáticas

Prof. Wladimir Rodriguez

wladimir.rodriguez@outlook.com

Departamento de Computación

1 Contenido

- Definiciones: palabras y lenguajes
- Operaciones con palabras y lenguajes
- Gramáticas
- Jerarquía de Chomsky

2 Alfabetos

- Un alfabeto es un conjunto finito A. Sus elementos se llamarán símbolos o letras. Notación:
- Alfabetos: $A, B, C, ...$
- Símbolos: $a, b, c, ...$ o números
- Ejemplos:
  - $A = \{0, 1\}$
  - $B = \{< 0, 0 >, < 0, 1 >, < 1, 0 >, < 1, 1 >\}$

3 Palabras

- Una palabra sobre el alfabeto A es una sucesión finita de elementos de A.

$$u = a_1 \cdots a_n$$

- donde $a_i \in A, \forall i = 1, \cdots, n$.
- Si $A = \{0, 1\}$ entonces 0111 es una palabra sobre este alfabeto
- El conjunto de todas las palabras sobre un alfabeto A se nota como A*.

- Notación:
  - Palabras: $u, v, x, y, z, \cdots$
- Si $u \in A^*$, entonces la longitud de la palabra u es el número de símbolos de A que contiene.
- Notación: $\mid u \mid$

- La palabra vacía es la palabra de longitud cero.
- Notación: ϵ
- Notación: El conjunto de cadenas sobre un alfabeto A excluyendo la cadena vacía se nota como A^+.

4 Operaciones: Concatenación

- Si $u, v\ A^*$, $u = a_1...a_n$, $v = b_1...b_m$, se llama concatenación de u y v a la cadena $u.v$ (o simplemente uv) dada por $a_1...a_nb_1...b_m$.
- Ejemplo:
  - Si $u = 011$, $v = 1010$, entonces $uv = 0111010$

- Propiedades:
  1. $\mid u.v \mid = \mid u \mid + \mid v \mid$, $\forall u, v \in A^*$
  2. Asociativa: $u.(v.w) = (u.v).w$, $\forall u, v, w \in A^*$
  3. Elemento Neutro: $u.\epsilon = \epsilon.u = u$, $\forall u \in A^*$
- Estructura de monoide

5 Iteración

- La Iteración n-esima de una cadena (u^n) es la concatenación con ella misma n veces.
- Si $u \in A^*$ entonces
  - $u^0 = \epsilon$
  - $u^{i+1} = u^i.u$, $\forall i \geq 0$
- Ejemplo:
  - Si $u = 010$, entonces $u^3 = 010010010$.

6 Cadena Inversa

- Si $u = a_1 \cdots a_n \in A^*$, entonces la cadena inversa de u es la cadena $u^{-1} = a_n \cdots a_1 \in A^*$.
- Ejemplo:
  - Si $u = 011$, entonces $u^{-1} = 110$.

7 Lenguajes

- Un lenguaje sobre el alfabeto A es un subconjunto del conjunto de las cadenas sobre $A : L \subseteq A^*$.
- Notación:

- Lenguajes: $L, M, N, \cdots$
- Ejemplos:
  - $L_1 = \{a, b, \epsilon\}$
  - $L_2 = \{a^i b^i \mid i = 0, 1, 2, \cdots\}$
  - $L_3 = \{uu^{-1} \mid u \in A^*\}$
  - $L_4 = \{a^{n^2} \mid n = 1, 2, 3, \cdots\}$

8 Operaciones: Concatenación

- Aparte de las operaciones de unión e intersección de lenguajes, dada su condición de conjuntos existe la operación de concatenación.
- Si L_1, L_2 son dos lenguajes sobre el alfabeto A, la concatenación de estos dos lenguajes se define como

$$L_1 L_2 = \{u_1 u_2 \mid u_1 \in L_1, u_2 \in L_2\}$$

- Propiedades:
  - $L\Phi = \Phi L = L$
  - Elemento Neutro: $\{\epsilon\}L = L\{\epsilon\} = L$
  - Asociativa: $L_1(L_2 L_3) = (L_1 L_2) L_3$

9 Operaciones: Iteración

- La iteración de lenguajes se define de forma recursiva:

$$L^0 = \{\epsilon\},$$

$$L^{i+1} = L^i L$$

- Si L es un lenguaje sobre el alfabeto A, la clausura de Kleene de L es:

$$L^* = \cup_{i \geq 0} L^i$$

$$L^+ = \cup_{i \geq 1} L^i$$

- Propiedades:
  - $L^+ = L^*$ si $\epsilon \in L$
  - $L^+ = L^* - \epsilon$ si $\epsilon \notin L$
- El lenguaje inverso de L es el lenguaje dado por:
  - $L^{-1} = \{u \mid u^{-1} \in L\}$

10 Ejemplos

- Si $L_1 = \{0^i 1^i : i \geq 0\}$, $L_2 = \{1^i 0^i : i \geq 0\}$ entonces,
- $L_1 L_2$?

- $L_1L_2 = \{0^i1^i1^j0^j : i, j\ 0\}$

- Si $L = \{0, 01\}$, entonces
- L^* = Conjunto de palabras sobre $\{0, 1\}$ en las que un uno va siempre precedido de un cero, incluyendo la palabra vacía
- L^+ = Conjunto de palabras sobre $\{0, 1\}$ en las que un uno va siempre precedido de un cero y sin incluir la palabra vacía

11 Operaciones: Cabecera

- La cabecera de L es el lenguaje dado por

$$CAB(L) = \{u \mid u \in A^* \ \ y \ \ \exists v \in A^* \ \ tal\ que \ \ uv \in L\}$$

12 Operaciones Homomorfismo

- Si A_1 y A_2 son dos alfabetos, una aplicación

$$h : A_1^* \to A_2^*$$

- se dice que es un homomorfismo si y solo si

$$h(uv) = h(u)h(v)$$

- Consecuencias:

$$h(\epsilon) = \epsilon$$

$$h(a_1 \cdots a_n) = h(a_1) \cdots h(a_n)$$

- Si $A_1 = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}, A_2 = \{0, 1\}$

$$h(0) = 0000, \ \ h(1) = 0001, \ \ h(2) = 0010, \ \ h(3) = 0011$$
$$h(4) = 0100, \ \ h(5) = 0101, \ \ h(6) = 0110, \ \ h(7) = 0111$$
$$h(8) = 1000, \ \ h(9) = 1001$$

$$h(034) = 000000110100, \ \ h() =$$

13 Gramáticas

- Una gramática describe la estructura de las frases de un lenguaje y se aplica por igual a:
  - Lenguajes naturales
  - Lenguajes formales
- Una gramática es un *ente formal* para especificar de manera finita el conjunto de cadenas de símbolos que constituyen un lenguaje.

14 Gramáticas Generativas

- Una gramática generativa es una cuádrupla (V, T, P, S) en la que
  - V es un alfabeto, llamado de variables o símbolos no terminales. Sus elementos se suelen representar con letras mayúsculas.
  - T es un alfabeto, llamado de símbolos terminales. Sus elementos se suelen representar con letras minúsculas.
  - P es un conjunto de pares (α, β), llamados reglas de producción, donde $\alpha, \beta \in (V \cup T)^*$ y α contiene, al menos un símbolo de V. El par (α, β) se suele representar como $\alpha \to \beta$
  - S es un elemento de V, llamado símbolo de inicio.

15 Gramática

- $G = (V, T, P, S)$
  - $V = \{E\}$
  - $T = \{+, *, (,), a, b, c\}$
  - P está compuesto por las siguientes reglas de producción

$$E \to E + E, \quad E \to E * E, \quad E \to (E)$$

$$E \to a, \quad E \to b, \quad E \to c$$

  - $S = E$

16 Lenguaje Generado: idea intuitiva

- Una gramática sirve para determinar un lenguaje

$$E \to E + E, \quad E \to E * E, \quad E \to (E)$$

$$E \to a, \quad E \to b, \quad E \to c$$

- Las palabras son las T^* que se obtienen a partir del símbolo inicial efectuando pasos de derivación. Cada paso consiste en elegir una parte de la palabra que coincide con la parte izquierda de una producción y sustituir esa parte por la derecha de la misma producción.

- E

- $E\ E * E$

- $E * E\ (E) * E$

- $(E) * E\ (E + E) * E$

- $(E + E) * E\ (a + E) * E$

- $(a + E) * E\ (a + b) * E$

- $(a + b) * E\ (a + b) * b$

- $(a + b) * b$ Palabra Generada

17 Paso de Derivación

- Gramática $G = (V, T, P, S)$ y dos palabras $\alpha, \beta \in (V \cup T)^*$, β es derivable a partir de α en un paso ($\alpha \Rightarrow \beta$) si y solo si
  - existen palabras $\delta_1, \delta_2 \in (V \cup T)^*$,
  - y una producción $\gamma \to \varphi$ tales que

$$\alpha = \delta_1 \gamma \delta_2, \quad \beta = \delta_1 \varphi \delta_2$$

18 Secuencia de Derivación

- β es derivable de α ($\alpha \Rightarrow^* \beta$) si y solo si existe una sucesión de palabras $\gamma_1, \cdots, \gamma_n (n \geq 1)$ tales que

$$\alpha = \gamma_1 \Rightarrow \gamma_2 \Rightarrow \cdots \Rightarrow \gamma_n = \beta$$

19 Lenguaje Generado

- Se define como lenguaje generado por una gramática $G = (V, T, P, S)$ al conjunto de cadenas formadas por símbolos terminales y que son derivables a partir del símbolo de partida.
- Es decir,
$$L(G) = \{u \in T^* \mid S \Rightarrow^* u\}$$

20 Ejemplo

- $G = (V, T, P, S)$, donde $V = \{S, A, B\}$, $T = \{a, b\}$. el símbolo de partida es S y las reglas son:
$$S \to aB, \quad S \to bA, \quad A \to a, \quad A \to aS$$
$$A \to bAA, \quad B \to b, \quad B \to bS, \quad B \to aBB$$
- Esta gramática genera el lenguaje
$$L(G) = \{u \mid u \in \{a, b\}^+ \ y \ N_a(u) = N_b(u)\}$$
- donde $N_a(u)$ y $N_b(u)$ son el número de apariciones de símbolos a y b, en u, respectivamente

21 Demostración

- Esto es fácil de ser interpretando
  - A palabra con una a de más
  - B palabra con una b de más
  - S palabra con igual número de a que de b.

- Hay que demostrar dos cosas:
  - Todas las palabras generadas por la gramática tienen el mismo número de a que de b.
  - Cualquier palabra con el mismo número de a que de b es generada
- Para lo primero, podemos considerar $N_{a,A}(u)$ (número de a + número de A) y $N_{b,B}(u)$ (número de b + número de B) y tener en cuenta lo siguiente para una generación $S\ {}^*u$:
  - Al principio de la generación tenemos:

$$N_{a,A}(S) = N_{b,B}(S) = 0$$

  - Al aplicar cualquier regla $u_1\ u_2$, si

$$N_{a,A}(u_1) = N_{b,B}(u_1),\ \ entonces\ \ N_{a,A}(u_2) = N_{b,B}(u_2)$$

- Luego la final $N_{a,A}(u) = N_{b,B}(u)$, y como u no contiene variables, $N_a(u) = N_b(u)$, como se quería demostrar

22 Algoritmo de Generación

- Generación por la izquierda, un símbolo a la vez.
- Para generar una a
  - Si a último símbolo de la palabra, aplicar $A \to a$
  - Si no es el último símbolo
    - Si la primera variable es S aplica $S \to aB$
    - Si la primera variable es B aplicar $B \to aBB$
    - Si la primera variable es A, aplicar $A \to a$ si hay mas variables, si no hay más variables aplicar $A \to aS$
- Generación por la izquierda, un símbolo a la vez.
- Para generar una b
  - Si b último símbolo de la palabra, aplicar $B \to b$
  - Si no es el último símbolo
    - Si la primera variable es S aplica $S \to bA$
    - Si la primera variable es A aplicar $A \to bAA$
    - Si la primera variable es B, aplicar $B \to b$ si hay mas variables, si no hay más variables aplicar $B \to bS$

23 Condiciones de Garantía

- Las palabras generadas tienen primero símbolos terminales y después variables.
- Se genera un símbolo de la palabra en cada paso de derivación
- Las variables que aparecen en la palabra pueden ser:
  - Una cadena de A (si hemos generado más b que a)
  - Una cadena de B (si hemos generado más a que b)
  - Una S si hemos generado las mismas a que b
- Antes de generar el último símbolo tendremos como variables:
  - Una A si tenemos que generar a

- Una B si tenemos que generar b

- Entonces aplicamos la primera opción para generar los símbolos y la palabra queda generada.

24 Ejemplo:

- Sea $G = (\{S, X, Y\}, \{a, b, c\}, P, S)$ donde P tiene las reglas,

$$S \rightarrow abc \quad S \rightarrow aXbc \quad Xb \rightarrow bX \quad Xc \rightarrow Ybcc$$

$$bY \rightarrow Yb \quad aY \rightarrow aaX \quad aY \rightarrow aa$$

25 Lenguaje Generado

- Esta gramática genera el lenguaje: $\{a^n b^n c^n \mid n = 1, 2, \cdots\}$.

$$aXbc \Rightarrow abXc \Rightarrow abYbcc \Rightarrow aYbbcc$$

- En este momento podemos aplicar dos reglas:
  - $aY \rightarrow aa$, en cuyo caso producimos

$$aabbcc = a^2b^2c^2 \in L(G)$$

  - $aY \rightarrow aaX$, en cuyo caso producimos $aaXbbcc$

26 Jerarquía de Chomsky

- Tipo 0: *Cualquier gramática.* Sin restricciones. Lenguajes recursivos
- Tipo 1: *Lenguajes dependientes del contexto.* Si todas las producciones tienen la forma

$$\alpha_1 A \alpha_2 \rightarrow \alpha_1 \beta \alpha_2$$

- donde α_1, α_2, $\beta \in (V \cup T)^*, A \in V$, y $\beta \neq \epsilon$, excepto posiblemente la regla $S \rightarrow \epsilon$, en cuyo caso S no aparece a la derecha de las reglas.

- Tipo 2: *Lenguajes independientes del contexto.* Si cualquier producción tiene la forma

$$A \rightarrow \alpha$$

- donde $A \in V$, $\alpha \in (V \cup T)^*$.
- Tipo 3: *Conjuntos regulares.* Si toda regla tiene la forma

$$A \rightarrow uB \qquad A \rightarrow u$$

- donde

$$u \in T^* \quad y \quad A, B \in V$$

.

27 Clases de Lenguajes

- Un lenguaje se dice que es de tipo $i (i = 0, 1, 2, 3)$ si y solo si es generado por una gramática de tipo i.
- La clase o familia de lenguaje de tipo i se denota por L_i.
- Propiedad

$$L_3 \subseteq L_2 \subseteq L_1 \subseteq L_0$$

28 Ejemplo 1

- Demostrar que la gramática

$$G = (\{S\}, \{a, b\}, \{S \to \epsilon, S \to aSb\}, S)$$

- genera el lenguaje $L = \{a^i b^i \mid i = 0, 1, 2, \cdots\}$
- Inicialmente tenemos dos opciones:

$$S \Rightarrow \epsilon, \quad S \Rightarrow aSb$$

- Con eso generamos la palabra vacía, o continuamos generando. Otra vez hay dos opciones:

$$S \Rightarrow S \Rightarrow aSb \Rightarrow ab, \quad S \Rightarrow aSb \Rightarrow aSb \Rightarrow aaSbb$$

- Si seguimos este procedimiento, nos encontramos que podemos ir generando todas las palabras de la forma $a^i b^i$, y siempre nos queda $a^i S b^i$ para seguir generando las palabras de mayor longitud
- Por otra parte, estas son las únicas palabras que se pueden generar.

29 Ejemplo 2

- Encontrar el lenguaje generado por la siguiente gramática $G = (\{S, A, B\}, \{a, b\}, P, S)$ donde P contiene las siguientes producciones:

$$S \to aAB \quad bB \to a \quad Ab \to SBb$$

$$Aa \to SaB \quad B \to SA \quad B \to ab$$

- El resultado es el Lenguaje vacío: nunca se puede llegar a generar una palabra con símbolos terminales. Siempre que se sustituye S aparece A, y siempre que se sustituye A aparece B.

30 Ejercicios

- Encontrar una gramática de libre contexto para generar cada uno de los siguientes lenguajes
  1. $L = \{a^i b^j \mid i, j \in N, i \leq j\}$
  2. $L = \{a^i b^j a^j b^i \mid i, j \in N\}$
  3. $L = \{a^i b^i a^j b^j \mid i, j \in N\}$

4. $L = \{a^i b^i \mid i \in N\} \cup \{b^i a^i \mid i \in N\}$
5. $L = \{uu^{-1} \mid u \in \{a,b\}^*\}$
6. $L = \{a^i b^j c^{i+j} \mid i,j \in N\}$

- donde N es el conjunto de los números naturales incluyendo el 0.

31 $L = \{a^i b^j \mid i, j \ N, i \ j\}$

- Este lenguaje es generado por la siguiente gramática:

$$S \to aSb$$

$$S \to \epsilon$$

$$S \to Sb$$

32 $L = \{a^i b^j a^j b^i \mid i, j \ N\}$

- Este lenguaje es generado por la siguiente gramática:

$$S \to aSb$$

$$S \to B$$

$$B \to bBa$$

$$B \to \epsilon$$

33 $L = \{a^i b^i a^j b^j \mid i, j \ N\}$

- Podemos generar $\{a^i b^i \mid i \ N\}$ con:

$$S_1 \to aS_1b, \quad S_1 \to \epsilon$$

- El lenguaje L se puede generar añadiendo:

$$S \to S_1 S_1$$

- Siendo S el símbolo inicial.

34 $L = \{a^i b^i \mid i\ N\}\ \{b^i a^i \mid i\ N\}$

- Podemos generar $\{a^i b^i \mid i\ N\}$ con:

$$S_1 \to aS_1b, \quad S_1 \to \epsilon$$

- y $\{b^i a^i \mid i\ N\}$ con:

$$S_2 \to bS_2a, \quad S_2 \to \epsilon$$

- El lenguaje L se puede generar añadiendo:

$$S \to S_1, \quad S \to S_2$$

- siendo S el símbolo inicial

35 $L = \{uu^{-1} \mid u\ \{a, b\}^*\}$

- Este lenguaje se genera con la gramática:

$$S \to aSa$$

$$S \to bSb$$

$$S \to \epsilon$$

36 $L = \{a^i b^j c^{i+j} \mid i, j\ N\}$

- Este lenguaje se genera con la gramática:

$$S \to aSc$$

$$S \to B$$

$$B \to bBc$$

$$B \to \epsilon$$

37 Ejercicio

- Determinar si la gramática $G = (\{S, A, B\}, \{a, b, c, d\}, P, S)$ donde P es el conjunto de reglas de producción:

$$S \rightarrow AB \quad A \rightarrow Ab \quad A \rightarrow a$$

$$B \rightarrow cB \quad B \rightarrow d$$

- genera un lenguaje de tipo 3.
- Esta gramática genera el lenguaje: $\{ab^i c^j d : i, j \; N\}$
- Y este lenguaje se puede generar mediante la gramática de tipo 3

$$S \rightarrow aB, \quad B \rightarrow bB, \quad B \rightarrow C, \quad C \rightarrow cC, \quad C \rightarrow d$$