

Teoría de Lenguajes

Clase Teórica 1 Gramáticas y Jerarquía de Chomsky

Primer cuatrimestre 2016

Material compilado por Julio Jacobo a lo largo de distintas ediciones de la materia Teoría de Lenguajes en el Departamento de Computación, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, y recientemente revisado por Verónica Becher.

Bibliografía:

Capítulo 1. *An Introduction to Formal Language Theory*, Michael Harrison, Addison Wesley, 1978.

Capítulo 1, *Introduction to Automata Theory, Languages and Computation*, J. Hopcroft, R. Motwani, J. Ullman, Second Edition, Addison Wesley, 2001.

Capítulo 1, *The Theory of Parsing, Translating and Compiling, Volume 1: Parsing*, Aho, J. Ullman, Prentice-Hall, 1971.

Alfabeto

Es un conjunto finito, no vacío, de elementos o caracteres.

Cadena

Es una secuencia finita de elementos (cero o más) de un alfabeto.

Ejemplo

Dado el alfabeto: $\Sigma = \{a, j, r\}$, tenemos como cadenas posibles: $aa, j, rja, raja, jarra$, etc.

Cadena nula λ

No tiene símbolos.

Lenguaje sobre un alfabeto Σ :

conjunto de cadenas sobre un alfabeto Σ .

Ejemplo:

- ϕ es un lenguaje
- $\{\lambda\}$ es un lenguaje (notar que es distinto de ϕ)
- Dado $\Sigma = \{0, 1\}$, $\{0, 01, 011, 0111, 01111, \dots\}$, es un lenguaje sobre Σ .

Concatenación \circ

Es una operación entre un símbolo del alfabeto Σ y una cadena sobre dicho alfabeto

$$\circ : \Sigma \times \{\text{cadenas sobre } \Sigma\} \rightarrow \{\text{cadenas sobre } \Sigma\}.$$

Ejemplo

Si el alfabeto es $\Sigma = \{a, b, c\}$, $\alpha = ab$ es una cadena, y entonces $a \circ ab = aab$ es también una cadena.

Cadena nula λ

Es el neutro de la concatenación:

$$\forall a \in \Sigma, a \circ \lambda = a$$

Concatenación de cadenas

Definimos la operación de concatenación de dos cadenas \circ . Supongamos cadenas x, y donde x tiene n símbolos $a_1 a_2 \dots a_n$. Entonces,.

$$x \circ y = x_1 \circ (x_2 \circ \dots (x_n \circ y) \dots)$$

Para facilitar la notación, en adelante no escribiremos el símbolo \circ .

Concatenación de lenguajes

Sea L_1 un lenguaje definido sobre el alfabeto Σ_1 , y sea L_2 un lenguaje definido sobre el alfabeto Σ_2 . Definimos la concatenación de L_1 y L_2 como el lenguaje

$$L_1 L_2 = \{xy : x \in L_1 \wedge y \in L_2\},$$

definido sobre el alfabeto $\Sigma_1 \cup \Sigma_2$.

Clausura de Kleene de un alfabeto Σ : el lenguaje Σ^*

- $\lambda \in \Sigma^*$
- Si $(a \in \Sigma \text{ y } \alpha \in \Sigma^*)$ entonces $a\alpha \in \Sigma^*$

Clausura positiva de un alfabeto Σ : el lenguaje Σ^+

Si $(a \in \Sigma \text{ y } \alpha \in \Sigma^*)$ entonces $a\alpha \in \Sigma^+$.

Ejemplo

Sea $\Sigma = \{a, j, r\}$, entonces $arj \in \Sigma^*$ porque $j\lambda = j \in \Sigma^*$, $rj = rj \in \Sigma^*$, y $arj = arj \in \Sigma^*$.

Clausura de Kleene de un lenguaje L : el lenguaje L^*

Se define por:

- $L^0 = \{\lambda\}$
- $L^n = LL^{n-1}$ para $n \geq 1$
- $L^* = \bigcup_{n \geq 0} L^n$.

Clausura positiva de un lenguaje L : el lenguaje L^+

Se define por:

- $L^+ = \bigcup_{n \geq 1} L^n$.

De lo anterior se ve que $L^+ = LL^* = L^*L$, y que $L^* = L^+ \cup \{\lambda\}$.

Y si L es un lenguaje definido sobre Σ , entonces, $L \subseteq \Sigma^*$.

Relaciones

Definición.

Dados los conjuntos A y B , se llama **relación** de A en B a cualquier subconjunto de $A \times B$.

Notación

- Si R es una relación de A en B , o sea, $R \subset A \times B$, podemos escribir $R : A \rightarrow B$.
- Si $B = A$ se dice que R es una relación sobre A .
- $a R b$ denota que el par (a, b) pertenece a la relación R , esto es, $(a, b) \in R$.

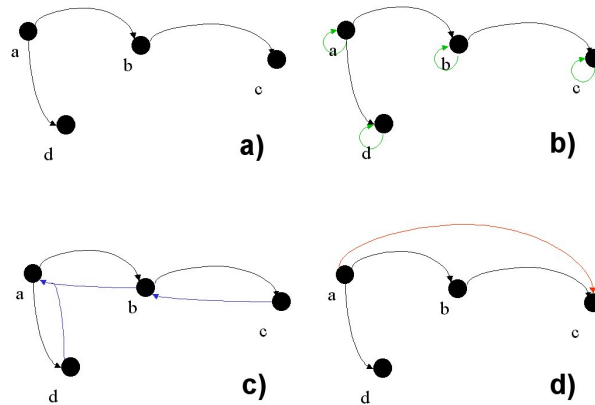


Figura 1: a) relación $R = \{(a, b), (a, d), (b, c)\}$, b) clausura reflexiva, c) clausura simétrica, d) clausura transitiva.

Propiedades de una relación R sobre A

Reflexividad

Una relación $R \subseteq A \times A$ es *reflexiva* cuando todo elemento de A está relacionado consigo mismo, o sea, cuando

$$\forall a \in A, (a, a) \in R.$$

Ejemplo

R: " \leq " sobre \mathbb{N} .

Simetría

Una relación $R \subseteq A \times A$ es *simétrica* cuando el hecho de que el par (a, b) pertenece a la relación R implica que el par (b, a) también pertenece a dicha relación, o sea, cuando

$$\forall a, b \in A, \left((a, b) \in R \Rightarrow (b, a) \in R \right).$$

Ejemplo

R: " \neq " sobre \mathbb{N} .

Transitividad

Una relación $R \subseteq A \times A$ es *transitiva* cuando el hecho de que los pares (a, b) y (b, c) pertenecen a la relación R implica que el par (a, c) también pertenece a dicha relación, o sea, cuando

$$\forall a, b, c \in A, \left((a, b) \in R \wedge (b, c) \in R \Rightarrow (a, c) \in R \right).$$

Ejemplo

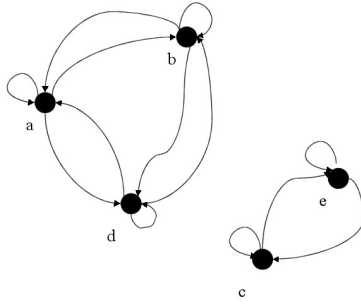
R: " a paralela a b ", en el conjunto de rectas del plano.

Relación de equivalencia

Una relación es de equivalencia, cuando es reflexiva, simétrica y transitiva.

Propiedad:

Una relación de equivalencia sobre un conjunto A particiona al mismo en subconjuntos disjuntos a los cuales se los llama **clases de equivalencia**.



Composición de relaciones:

Sean A, B y C tres conjuntos, y sean R y G dos relaciones tales que $R \subseteq A \times B$ y $G \subseteq B \times C$. La relación de composición $G \circ R \subseteq A \times C$ se define como

$$G \circ R = \{(a, c), a \in A, c \in C : \exists b \in B \text{ tal que } aRb \wedge bGc\}.$$

Relación de identidad:

Una relación R definida sobre A es de identidad (id_A) si se cumple que

$$\forall a, b \in A, a id_A b \text{ si y solo si } a = b.$$

Propiedad:

La relación de identidad es el elemento neutro de la composición. Dada una relación $R \subseteq A \times B$ es cierto que

$$id_B \circ R = R \circ id_A = R$$

Relación potencia:

Dada una relación $R \subseteq A \times A$, y dado n se define $R^n \subseteq A \times A$ como:

$$R^n = \begin{cases} id_A & \text{si } n = 0 \\ R \circ R^{n-1} & \text{si } n > 0 \end{cases}$$

con $R = R^1$.

Aclaración:

Notar que R^n es una relación (o sea, un *conjunto de pares*), cualquiera sea el valor de n .

Clausura transitiva

Dada una relación R sobre A , se define *clausura transitiva* R^+ como:

$$R^+ = \bigcup_{k=1}^{\infty} R^k,$$

Propiedades:

Una clausura transitiva cumple que

1. $R \subseteq R^+$
2. R^+ es transitiva

3. para toda relación G sobre A

$$\text{si } R \subseteq G \wedge G \text{ transitiva entonces } R^+ \subseteq G.$$

Demostración: R^+ es transitiva

Queremos probar que si aR^+b y bR^+c entonces aR^+c . Si aR^+b , entonces existe una secuencia de elementos d_1, \dots, d_n tal que $d_1 R d_2, \dots, d_{n-1} R d_n$, donde $d_1 = a$ y $d_n = b$. Por lo tanto, aR^+b . Análogamente, como bR^+c entonces existe una secuencia de elementos e_1, \dots, e_m tal que $e_1 R e_2, \dots, e_{m-1} R e_m$, donde $e_1 = b$ y $e_m = c$. Por lo tanto bR^+c . Concluimos que $aR^{n+m}c$, lo que a su vez implica que aR^+c .

Demostración: Si $R \subseteq G \wedge G$ transitiva entonces $R^+ \subseteq G$

si aR^+b entonces existe una secuencia de elementos c_1, \dots, c_n tal que $c_1 R c_2, \dots, c_{n-1} R c_n$, donde $c_1 = a$ y $c_n = b$. Como $R \subseteq G$ tenemos que $c_1 G c_2, \dots, c_{n-1} G c_n$, y como G es transitiva entonces, la aplicación repetida de la transitividad nos lleva a que $c_1 G c_n$, o sea aGb .

Clausura transitiva reflexiva: R^*

$$R^* = R^+ \cup id = \bigcup_{i=0}^{\infty} R^i.$$

Gramáticas

Definición

Una gramática es una 4-upla $G = \langle V_N, V_T, P, S \rangle$ donde

- V_N es un conjunto de símbolos llamados no-terminales (también, variables o categorías sintácticas)
- V_T es un conjunto de símbolos terminales (tal como lo era Σ en los ejemplos anteriores)
- P es el conjunto de "producciones", que es un conjunto finito de

$$(V_N \cup V_T)^* V_N (V_N \cup V_T)^* \times (V_N \cup V_T)^*,$$

estas producciones son entonces pares ordenados (α, β) , que usualmente son notados como $\alpha \rightarrow \beta$.

- $S \in V_N$ es el símbolo distinguido de V_N .

Forma sentencial de una gramática $G = \langle V_N, V_T, P, S \rangle$

- S es una forma sentencial de G .
- Si $\alpha\beta\gamma$ es una forma sentencial de G , y $(\beta \rightarrow \delta) \in P$, entonces $\alpha\delta\gamma$ es también una forma sentencial de G .

Derivación directa en G

Si $\alpha\beta\gamma \in (V_N \cup V_T)^*$ y $(\beta \rightarrow \delta) \in P$, se dice que $\alpha\delta\gamma$ se deriva directamente en G de $\alpha\beta\gamma$ y se denota como

$$\alpha\beta\gamma \xrightarrow{G} \alpha\delta\gamma.$$

Entonces, \xrightarrow{G} es una relación sobre $(V_N \cup V_T)^*$, es decir,

$$\xrightarrow{G} \subseteq (V_N \cup V_T)^* \times (V_N \cup V_T)^*.$$

Podemos componer la relación \xrightarrow{G} consigo misma, 0 o más veces...

Clausura de Kleene de la relación de derivación \rightarrow_G

$$\begin{aligned}\left(\rightarrow_G\right)^0 &= id_{(V_N \cup V_T)^*} \\ \text{Si } n > 0, \left(\rightarrow_G\right)^n &= \left(\rightarrow_G\right)^{n-1} \circ \rightarrow_G \\ \left(\rightarrow_G\right)^+ &= \bigcup_{k=1}^{\infty} \left(\rightarrow_G\right)^k \\ \left(\rightarrow_G\right)^* &= \left(\rightarrow_G\right)^+ \cup id_{(V_N \cup V_T)^*}\end{aligned}$$

Definición

Denotaremos con $\xrightarrow[k]{G}$ a la potencia k de la relación \rightarrow_G .

Definición

Denotaremos con $\xrightarrow{+}_G$ y con $\xrightarrow{*}_G$ a las clausura transitiva y a la clausura transitiva y reflexiva de \rightarrow_G respectivamente.

Definición

Lenguaje generado por una gramática $G = \langle V_N, V_T, P, S \rangle$, el cual se denotará como $\mathcal{L}(G)$,

$$\mathcal{L}(G) = \left\{ \alpha \in V_T^* : S \xrightarrow{+}_G \alpha \right\}$$

Clasificación de gramáticas (Chomsky)

3 (las más simples),
2,
1,
0 (las más sofisticadas)

Clasificación de gramáticas (Chomsky)

Gramáticas regulares (tipo 3)

- Las producciones son de la forma $A \rightarrow xB$ o $A \rightarrow x$, donde $A, B \in V_N$ y $x \in V_T^*$. La gramática es llamada regular "lineal a derecha".
- Las producciones son de la forma $A \rightarrow Bx$ o $A \rightarrow x$, donde $A, B \in V_N$ y $x \in V_T^*$. La gramática es llamada regular "lineal a izquierda".

Ambos tipos de gramática son llamados regulares.

Clasificación de gramáticas (Chomsky)

Gramáticas independientes del contexto (tipo 2) (también llamadas libres de contexto)

Cada producción es de la forma $A \rightarrow \alpha$, donde $A \in V_N$ y $\alpha \in (V_N \cup V_T)^*$.

Clasificación de gramáticas (Chomsky)

Gramáticas Dependientes del contexto (tipo 1) (también llamada sensitiva al contexto)

Cada producción es de la forma

$$S \rightarrow \lambda$$

o de la forma

$$\alpha A \gamma \rightarrow \alpha \beta \gamma$$

donde

$$A \in V_N; \alpha, \gamma \in (V_N \cup V_T)^*; \beta \in (V_N \cup V_T)^+.$$

(Notar que esta segunda forma impide la generación de la cadena nula λ)

Clasificación de gramáticas (Chomsky)

Sin restricciones (tipo 0)

No poseen ninguna restricción.

Lenguaje generado por una gramática

Un lenguaje generado por una gramática tipo t es llamado "lenguaje t ".

Ejemplo de gramática tipo 3 (regular)

$G = \langle \{S, A, B, C\}, \{a, b, c\}, S, P \rangle$, donde P está dado por

$$\begin{aligned} S &\rightarrow aA \\ A &\rightarrow aB \\ A &\rightarrow aA \\ B &\rightarrow bB \\ B &\rightarrow bC \\ C &\rightarrow cC \\ C &\rightarrow c \end{aligned}$$

Derivación de $aabbbbcccc$

$$\begin{array}{llll} S & \rightarrow aA & \rightarrow aaB & \rightarrow aabB \\ & \rightarrow aabbB & \rightarrow aabbbB & \rightarrow aabbbbC \\ & \rightarrow aabbbbcC & \rightarrow aabbbbccC & \rightarrow aabbbbcccc \end{array}$$

G genera el lenguaje $\{a^n b^m c^k : m \geq 2; n, k \geq 1\}$.

Ejemplo de gramática tipo 2 (independiente del contexto)

$G = \langle \{E, T, F\}, \{a, +, *, (,)\}, E, P \rangle$ donde P está dado por

$$\begin{aligned} E &\rightarrow E + T \\ E &\rightarrow T \\ T &\rightarrow T * F \\ T &\rightarrow F \\ F &\rightarrow (E) \\ F &\rightarrow a \end{aligned}$$

Derivación de $a * (a + a)$ donde

$$\begin{aligned} E &\rightarrow T \rightarrow T * F \\ &\rightarrow T * (E) \rightarrow F * (E) \\ &\rightarrow a * (E) \rightarrow a * (E + T) \\ &\rightarrow a * (T + T) \rightarrow a * (F + T) \\ &\rightarrow a * (a + T) \rightarrow a * (a + F) \\ &\rightarrow a * (a + a) \end{aligned}$$

Ejemplo de gramática tipo 1 (dependiente del contexto):

Gramática $G = \langle \{S, B, C\}, \{a, b, c\}, S, P \rangle$, donde P está dado por

$$\begin{array}{llll} S & \rightarrow aSBC & CB & \rightarrow BC \\ S & \rightarrow abC & bB & \rightarrow bb \\ & & cC & \rightarrow cc \\ & & bC & \rightarrow bc \end{array}$$

Derivación de $aaabbbccc$:

$$\begin{array}{lll}
 S \rightarrow & & \\
 \rightarrow aSBC & \rightarrow aaSBCBC & \rightarrow aaabCBCBC \\
 \rightarrow aaabBCCBC & \rightarrow aaabBCBCC & \rightarrow aaabBBCCC \\
 \rightarrow aaabbBCCC & \rightarrow aaabbbCCC & \rightarrow aaabbbccC \\
 \rightarrow aaabbbccC & \rightarrow aaabbbccc &
 \end{array}$$

Ejemplo de gramática tipo 0 (sin restricciones):

Gramática para generar $\{ww : w \in \{a, b\}^*\}$ $G = (\{S, A, B, C, D\}, \{a, b\}, S, P)$, donde P está dado por

$$\begin{array}{llllll}
 S \rightarrow CD & AD \rightarrow aD & Aa \rightarrow aA & Ba \rightarrow aB & C \rightarrow \lambda \\
 C \rightarrow aCA & BD \rightarrow bD & Ab \rightarrow bA & Bb \rightarrow bB & D \rightarrow \lambda \\
 C \rightarrow bCB & & & &
 \end{array}$$

derivación de $abaaabaa$

$$\begin{array}{lll}
 S \rightarrow CD & \rightarrow aCAD & \rightarrow abCBAD \\
 \rightarrow abaCABAD & \rightarrow abaaCAABAD & \rightarrow abaaAABAD \\
 \rightarrow abaaAABaD & \rightarrow abaaAAaBD & \rightarrow abaaAaABD \\
 \rightarrow abaaaAABD & \rightarrow abaaaAAbD & \rightarrow abaaaAbAD \\
 \rightarrow abaaabAAD & \rightarrow abaaabAaD & \rightarrow abaaabaAD \\
 \rightarrow abaaabaaD & \rightarrow abaaabaa &
 \end{array}$$