Capítulo 6

Gramáticas

6.1. Definiciones

6.1.1. Alfabeto

Un alfabeto es un conjunto finito y no vacío, cuyos elementos reciben el nombre de símbolos, letras o caracteres. Notaremos a los alfabetos con la letra sigma, por ejemplo $\Sigma = \{l_1, l_2, \ldots, l_k\}$.

6.1.2. Cadena

Una cadena o palabra sobre un alfabeto Σ es una secuencia finita de símbolos de Σ. Mas precisamente se la define como la función $p : \{1, \ldots, n\} \to \Sigma$ donde el entero positivo n es la longitud de la palabra, que se denota también como $|p|$. Ademas definimos la palabra especial $\lambda : \{\} \to \Sigma$ llamada nula, o vaciá, considerada de longitud 0. Dada la palabra $p : \{1, \ldots, n\} \to \Sigma$ por simplicidad de notación escribiremos la cadena p como $p\,(1)\,p\,(2) \ldots p\,(n) = l_1 l_2 \ldots l_n$.

6.1.3. Clausuras de Kleene

Dado un alfabeto Σ definimos $\Sigma^0 = \{\lambda\}$ y $\Sigma^n = \{p/p : \{1, \ldots, n\} \to \Sigma\}$ para todo $n \geq 1$, es decir, el conjunto de todas las palabra de una determinada longitud. Definimos ademas dos operadores: $\Sigma^* = \bigcup\limits_{n \geq 0} \Sigma^n$ y $\Sigma^+ = \bigcup\limits_{n \geq 1} \Sigma^n$, es decir que Σ^* es el conjunto de todas las palabras sobre un alfabeto Σ, incluyendo la palabra vacía.

Observaciones

- Dado un conjunto $A = \emptyset$ resulta $A^* = \{\lambda\}$.
- Si $A \neq \emptyset$ resulta A^* es infinito numerable pues es u.n.c.n.

6.1.4. Concatenación

Dado un alfabeto Σ y dos palabras $p : \{1, \ldots, m\} \to \Sigma$ y $q : \{1, \ldots, n\} \to \Sigma$ definimos la concatenación de p y q (denotada pq) como la palabra $pq : \{1, \ldots, m+n\} \to \Sigma$ tal que:

$$pq\left(i\right) = \begin{cases} p\left(i\right) & 1 \leq i \leq m \\ q\left(i-m\right) & m+1 \leq i \leq m+n \end{cases}$$

Observaciones

- En particular $p\lambda = \lambda p = p$.
- Es fácil ver que $p\left(qr\right) = \left(qp\right)r$ pero $pq \neq qr$, es decir, la concatenación es asociativa pero no conmutativa.

6.1.5. Cadena potencia

Dada una palabra $p \in \Sigma^*$ definimos su potencia p^n como:

$$p^n = \begin{cases} \lambda & n = 0 \\ p^{n-1}p & n \geq 1 \end{cases}$$

6.1.6. Cadena reversa

Dada una palabra $p \in \Sigma^*$ y un caracter $c \in \Sigma$ definimos inductivamente la cadena reversa p^R como:

$$\begin{aligned} \lambda^R = &\quad \lambda \\ \left(pc\right)^R = &\quad cp^R \end{aligned}$$

6.1.7. Subcadenas

Diremos que:

- s es *subcadena* de p si existen $q, r \in \Sigma^*$ tales que: $p = qsr$.
- s es *prefijo* de p si es una subcadena tal que: $p = \lambda sr$.
- s es sufijo de p si es una subcadena tal que: $p = qs\lambda$.
- s es *subcadena propia* de p si es subcadena de p y ademas $s \neq p$.

6.2. Lenguajes

6.2.1. Definición

Un lenguaje sobre un alfabeto Σ es un subconjunto de Σ^*. Para cada alfabeto Σ llamaremos $\mathcal{L}$ al conjunto de todos los lenguajes sobre Σ, es decir, $\mathcal{L} = \mathcal{P}(\Sigma^*)$.

Observación Para cualquier alfabeto, sabemos que $\#\Sigma^* = \aleph_0$ y en consecuencia $\#\mathcal{L} = \#\mathcal{P}(\Sigma^*) = \aleph_1$.

6.2.2. Unión

Dados dos lenguajes L_1 y L_2 sobre Σ definimos la unión de los lenguajes como: $L_1 \cup L_2 = \{p \in \Sigma^* / p \in L_1 \vee p \in L_2\}$.

6.2.3. Intersección

Dados dos lenguajes L_1 y L_2 sobre Σ definimos la intersección de los lenguajes como: $L_1 \cap L_2 = \{p \in \Sigma^* / p \in L_1 \wedge p \in L_2\}$.

6.2.4. Diferencia

Dados dos lenguajes L_1 y L_2 sobre Σ definimos la diferencia de los lenguajes como: $L_1 - L_2 = \{p \in \Sigma^* / p \in L_1 \wedge p \notin L_2\}$.

6.2.5. Complemento

Para cada lenguaje L, definimos el lenguaje complemento como: $\bar{L} = \Sigma^* - L$.

6.2.6. Concatenación

Dados dos lenguajes L_1 y L_2 sobre Σ definimos la concatenación de los lenguajes como: $L_1L_2 = \{p \in \Sigma^* / \exists q \in L_1, \exists r \in L_2, p = qr\}$.

Observaciones

- $L\emptyset = \emptyset L = \emptyset$.
- $L\{\lambda\} = \{\lambda\} L = L$.
- En general $L_1L_2 \neq L_2L_1$.

6.2.7. Potencia

Dado un lenguaje L definimos inductivamente su potencia L^n como:

$$\begin{aligned} L^0 = &\quad \{\lambda\} \\ L^{n+1} = &\quad L^n L \end{aligned}$$

6.2.8. Clausuras de Kleene

Dados un alfabeto Σ y un lenguaje L sobre Σ definimos: $L^* = \bigcup_{n\geq 0} L^n$ y $L^+ = \bigcup_{n\geq 1} L^n$, es decir que L^* es la unión infinita:

$$L^* = L^0 \cup L^1 \cup L^2 \cup \cdots \cup L^n \cup \cdots = \{\lambda\} \cup L \cup L^2 \cup \cdots \cup L^n \cup \cdots$$

mientras que L^+ es la unión infinita:

$$L^+ = L^1 \cup L^2 \cup \cdots \cup L^n \cup \cdots = L \cup L^2 \cup \cdots \cup L^n \cup \cdots$$

Observaciones

- Tanto L^* como L^+ contienen a L.
- Dado que $L^0 = \{\lambda\}$ resulta que $L^* = L^+ \cup \{\lambda\}$.
- $\emptyset^* = \{\lambda\}$.
- $L^+ = L^*L$.
- $(L^*)^* = L^*$.

- $L^* L^* = L^*$.

Al contrario que con los operadores anteriores, las clausuras de Kleene construyen siempre lenguajes infinitos.

6.3. Gramáticas

6.3.1. Definición

Una gramática es una tupla (N, T, P, σ) donde:

- N es un conjunto finito de símbolos llamados *no terminales*.
- T es un conjunto finito de símbolos llamados *terminales*, o alfabeto, tal que $N \cap T = \emptyset$.
- P es un conjunto finito de *reglas de producción* donde $P \subseteq [(N \cup T)^* - T^*] \times [N \cup T]^*$.
- $\sigma \in N$ es el llamado *símbolo inicial*.

Observaciones

- Nótese que $[(N \cup T)^* - T^*]$ es el conjunto de cadenas de no terminales y terminales que contienen al menos un no terminal. Dado que $\lambda \notin [(N \cup T)^* - T^*]$ no puede haber reglas del tipo (λ, β).
- A una regla de producción (α, β) la notaremos: $\alpha \to \beta$.
- Una regla del tipo $\alpha \to \lambda$ recibe el nombre de regla λ.

6.3.2. Derivación

Si $\alpha \to \beta$ es una regla de producción y $\gamma\alpha\delta \in [(N \cup T)^* - T^*]$, entonces diremos que $\gamma\beta\delta$ deriva directamente de $\gamma\alpha\delta$ o que $\gamma\alpha\delta$ produce directamente $\gamma\beta\delta$ y lo notaremos: $\gamma\alpha\delta \Rightarrow \gamma\beta\delta$. Si vale que $\alpha_1 \Rightarrow \alpha_2 \Rightarrow \ldots \Rightarrow a_{n-1} \Rightarrow \alpha_n$ donde $\alpha_i \in [(N \cup T)^* - T^*]$ y $\alpha_n \in [N \cup T]^*$ diremos que α_n deriva de α_1 o que α_1 produce α_n y lo notaremos: $\alpha_1 \Rightarrow^* \alpha_n$.

6.3.3. Lenguajes generados

Definimos el lenguaje generado por una gramática $G = (N, T, P, \sigma)$ como $L(G) = \{p \in T^* / \sigma \Rightarrow^* p\}$.

6.3.4. Gramáticas regulares

Las gramáticas regulares (también llamadas del tipo 3 o lineales) pueden ser clasificadas como derechas o izquierdas. Las reglas de producción de una gramática regular derecha tienen las siguientes restricciones:

1. El lado izquierdo debe consistir en un solo no terminal.
2. El lado derecho esta formado por un símbolo terminal que puede estar seguido o no, por un símbolo no terminal o la cadena vacía.

Es decir que las producciones de una gramática regular derecha tienen la forma $A \to a$, $A \to aB$ o $A \to \lambda$, donde $A, B \in N$ y $a \in T$. Alternativamente, en una gramática regular izquierda las reglas de producción son de la forma: $A \to a$, $A \to Ba$ o $A \to \lambda$.

Observaciones Toda gramática regular derecha puede ser convertida en una gramática regular izquierda (y viceversa). Algunas definiciones permiten reemplazar a por una cadena de uno o mas terminales, siendo ambas definiciones equivalentes.

6.3.5. Gramáticas libres de contexto

Las gramáticas libres o independientes del contexto son iguales a las regulares, pero sin restricciones sobre el lado derecho. Es decir que las reglas de producción tienen la forma $A \to \alpha$ donde $A \in N$ y $\alpha \in (N \cup T)^*$.

Observaciones

- Estas gramáticas reciben ese nombre debido a que en el lado izquierdo el no terminal aparece solo, por lo que la regla se puede aplicar sin importar el contexto en el que aparece dicho no terminal, al contrario de reglas de la forma $\gamma N \delta \to \gamma \alpha \delta$ donde el no terminal N solo se puede reemplazar por α cuando se encuentre en el contexto de γ y δ.
- Dado que no existen restricciones sobre el lado derecho, podría ocurrir que en el aparezcan mas de uno como en $A \to XY$. En esta situación el enfoque mas común consiste en reemplazar en el paso siguiente el no terminal situado mas a la izquierda.

- Alternativamente se podría reemplazar no terminales desde la derecha o algún otro patrón pues resultara que el orden en el que se apliquen las reglas, no afecta la determinación de si una cadena puede ser generada por la gramática.

6.3.6. Gramáticas sensibles al contexto

Las reglas de producción de una gramática dependiente del contexto son del tipo $\alpha A\beta \to \alpha\delta\beta$ donde $A \in N$, $\alpha, \beta \in [N \cup T]^*$ y $\delta \in [N \cup T]^+$. Adicionalmente se permite la regla $\sigma \to \lambda$ donde σ es el símbolo inicial siempre que σ no aparezca en el lado derecho de otra producción. Quizás a simple vista podría parecer que las gramáticas sensibles al contexto son mas restrictivas que las libre de contexto, sin embargo basta tomar $\alpha = \lambda = \beta$ para ver que las sensibles al contexto abarcan a las libres de contexto.

6.3.7. Gramáticas estructuradas por frases

Finalmente las gramáticas estructuradas por frases o irrestrictas, son aquellas que no tienen restricciones sobre la forma de las reglas de producción. Es decir que sus reglas son de la forma $\alpha \to \beta$ donde $\alpha \in [(N \cup T)^* - T^*]$ y $\beta \in [N \cup T]^*$.

Observación Dado que $\lambda \notin [(N \cup T)^* - T^*]$ no puede haber reglas del tipo $\lambda \to \beta$.

6.3.8. Relación entre gramáticas

Llamando $\mathcal{G}_i = \{G/G \text{ es del tipo } i\}$ con $i = 0, 1, 2, 3$, se tiene que: $\mathcal{G}_3 \subset \mathcal{G}_2 \subset \mathcal{G}_1 \subset \mathcal{G}_0$.

6.3.9. Tipos de lenguajes

Diremos que un lenguaje L es de tipo i (con $i = 0, 1, 2, 3$) si y solo si existe una gramática G del tipo i tal que $L(G) = L$.