Alfabetos — Palabras — Lenguajes

Horst H. von Brand
vonbrand@inf.utfsm.cl

Departamento de Informática
Universidad Técnica Federico Santa María

Contenido

Definiciones

Operaciones

Resumen

Introducción

Hablaremos de *lenguajes*, compuestos de *palabras*, a su vez compuestas de *símbolos*. Buscamos analizar las anteriores matemáticamente. Debemos partir definiendo nuestros términos.

Alfabeto

Definición

Un *alfabeto* es un conjunto finito de símbolos atómicos.

Usaremos letras griegas mayúsculas para designar alfabetos: $\Sigma, \Gamma, \Delta, \ldots$

Que los símbolos sean atómicos significa que los consideraremos como un todo, sin partes. Para nosotros, una letra como «á» es solo eso, no una «a» con acento, es un símbolo distinto.

Palabras

Definición

Una *palabra* es una secuencia finita de símbolos de un alfabeto Σ.

Usaremos letras griegas minúsculas para designar palabras: $\alpha, \beta, \gamma, \ldots, \varepsilon, \ldots$
La palabra de largo cero (vacía) la designaremos ε. Cuidado, algunos autores aún usan λ para designarla.
Note que siendo secuencias, pueden contener símbolos repetidos, y el orden de los mismos es relevante.
En inglés suelen llamarse *string*.

Lenguajes

Definición

Un *lenguaje* sobre el alfabeto Σ es un conjunto de palabras sobre Σ.

Lenguajes

Siendo conjuntos, el conjunto vacío $\varnothing$ es un lenguaje. Tenga cuidado de distinguir $\varnothing$ del lenguaje $\{\varepsilon\}$. El primero no tiene elementos, el segundo tiene un elemento.

Según esta definición son lenguajes C (el conjunto de todos los archivos válidos en C) y el castellano (todos los textos que se pueden escribir en castellano). Acá nos interesa solo si es o no legal, no nos enfrascamos con ortografía, gramática o sentido de las palabras.

Operaciones con palabras

Largo: El *largo* de la palabra α es su número de símbolos (el largo de la secuencia). Lo anotamos $|\alpha|$. Algunos anotan $\lg(\alpha)$.

Concatenación: *Concatenar* las palabras $\alpha = a_1 a_2 \dots a_m$ y $\beta = b_1 b_2 \dots b_n$ resulta en la palabra $a_1 a_2 \dots a_m b_1 b_2 \dots b_n$. Lo anotamos $\alpha \cdot \beta$ o simplemente $\alpha\beta$.

Operaciones con palabras

Potencias:

$$\alpha^n = \begin{cases} \varepsilon & n = 0 \\ \alpha^{n-1} \cdot \alpha & n \geq 1 \end{cases}$$

Reverso:

$$\begin{aligned} \varepsilon^{\mathrm{R}} &= \varepsilon \\ (\alpha a)^{\mathrm{R}} &= a \alpha^{\mathrm{R}} \qquad\qquad \alpha \in \Sigma^*, a \in \Sigma \end{aligned}$$

Operaciones con palabras

Algunas propiedades:

- $|\varepsilon| = 0$
- $|\alpha\beta| = |\alpha| + |\beta|$
- $|\alpha^{\mathrm{R}}| = |\alpha|$
- $\alpha \cdot \varepsilon = \varepsilon \cdot \alpha = \alpha$
- En general, $\alpha \cdot \beta \neq \beta \cdot \alpha$
- $(\alpha\beta)^{\mathrm{R}} = \beta^{\mathrm{R}}\alpha^{\mathrm{R}}$

Operaciones entre lenguajes

Son conjuntos, son aplicables las operaciones entre conjuntos. Suele anotarse $L_1 \mid L_2$ para la unión entre L_1 y L_2 (idea es «una palabra de L_1 o de L_2»).

Concatenar: $L_1 \cdot L_2 = \{\alpha \cdot \beta \colon \alpha \in L_1 \wedge \beta \in L_2\}$

Potencias:

$$L^n = \begin{cases} \{\varepsilon\} & n = 0 \\ L^{n-1} \cdot L & n \geq 1 \end{cases}$$

Estrella de Kleene:

$$L^* = \{\alpha \colon \exists k \geq 0, \alpha \in L^k\}$$

Más:

$$L^+ = \{\alpha \colon \exists k \geq 1, \alpha \in L^k\}$$

Operaciones entre lenguajes

Extendemos operaciones sobre palabras a lenguajes elemento a elemento. Por ejemplo:

$$L^{\mathrm{R}} = \{\alpha^{\mathrm{R}} \colon \alpha \in L\}$$

Operaciones entre lenguajes

Algunos ejemplos:

$$L_1 = \{a, ab\} \qquad L_2 = \{a, ba, bab\}$$

$$\begin{aligned}
L_1 \mid L_2 = L_2 \mid L_1 &= \{a, ab, ba, bab\} \\
L_1 \cdot L_2 &= \{aa, aba, abab, abba, abbab\} \\
L_2 \cdot L_1 &= \{aa, aab, baa, baab, baba, babab\} \\
L_2^{\mathrm{R}} &= \{a, ab, bab\}
\end{aligned}$$

Operaciones entre lenguajes

Potencias:

$$L = \{a, ab\}$$

$$\begin{aligned}
L^0 &= \{\varepsilon\} \\
L^1 &= L^0 \cdot L \\
&= \{\varepsilon\} \cdot L \\
&= L \\
&= \{a, ab\} \\
L^2 &= \{aa, aab, aba, abab\} \\
L^3 &= \{aaa, aaab, aaba, aabab, abaa, abaab, ababa, ababab\}
\end{aligned}$$

Operaciones entre lenguajes

Estrella de Kleene:

$$
\begin{aligned}
L^* &= L^0 \cup L^1 \cup L^2 \cup \cdots \\
&= \{\varepsilon\} \cup \{a, ab\} \cup \{aa, aab, aba, abab\} \cup \cdots \\
&= \{\varepsilon, a, ab, aa, aab, aba, abab, \ldots\} \\
L^+ &= L^1 \cup L^2 \cup \cdots \\
&= \{a, ab, aa, aab, aba, abab, \ldots\}
\end{aligned}
$$

Operaciones entre lenguajes

Note que:

$$
\begin{aligned}
\varnothing^0 &= \{\varepsilon\} & \{\varepsilon\}^0 &= \{\varepsilon\} \\
\varnothing^* &= \varnothing^0 \cup \varnothing^1 \cup \varnothing^2 \cup \cdots & \{\varepsilon\}^* &= \{\varepsilon\}^0 \cup \{\varepsilon\}^1 \cup \{\varepsilon\}^2 \cup \cdots \\
&= \{\varepsilon\} & &= \{\varepsilon\} \\
\varnothing^+ &= \varnothing^1 \cup \varnothing^2 \cup \cdots & \{\varepsilon\}^+ &= \{\varepsilon\}^1 \cup \{\varepsilon\}^2 \cup \cdots \\
&= \varnothing & &= \{\varepsilon\}
\end{aligned}
$$

Operaciones entre lenguajes

Algunas propiedades:

- $L \mid L = L$
- $L_1 \mid L_2 = L_2 \mid L_1$
- $L_1 \mid (L_2 \mid L_3) = (L_1 \mid L_2) \mid L_3$
- $\varnothing \mid L = L \mid \varnothing = L$
- $\{\varepsilon\} \cdot L = L \cdot \{\varepsilon\} = L$
- $\varnothing \cdot L = L \cdot \varnothing = \varnothing$
- $L_1 \cdot (L_2 \cdot L_3) = (L_1 \cdot L_2) \cdot L_3$
- $(L_1 \mid L_2) \cdot L_3 = (L_1 \cdot L_3) \mid (L_2 \cdot L_3)$ y
  $L_1 \cdot (L_2 \mid L_3) = (L_1 \cdot L_2) \mid (L_1 \cdot L_3)$
- $\{\varepsilon\} \mid L^+ = L^*$

En general, es $L_1 \cdot L_2 \neq L_2 \cdot L_1$.

Interpretación de las operaciones

- $L_1 \mid L_2$ es tomar una palabra de L_1 o una palabra de L_2
- $L_1 \cdot L_2$ es tomar una palabra de L_1 y concatenarla con una palabra de L_2
- L^n es tomar n palabras (posiblemente distintas) de L y concatenarlas en orden. Note que en general:

$$L^n \neq \{\alpha^n \colon \alpha \in L\}$$

- L^* es tomar cero o más palabras de L (posiblemente distintas) de L y concatenarlas en orden.
- L^+ es tomar una o más palabras de L (posiblemente distintas) de L y concatenarlas en orden.

Notación

Vimos que unión y concatenación de lenguajes son asociativas, por lo que el orden en que se efectúen estas operaciones no importa. Para operaciones mixtas (unión, concatenación, potencias, estrella) adoptamos la convención del álgebra: primero potencias y estrella, luego concatenaciones, luego uniones. Esto ahorra interminables paréntesis.

Resumen

- Definimos alfabetos, palabras, lenguajes
- Operaciones con palabras y con lenguajes, y propiedades simples

Es importante familiarizarse con el significado de las operaciones entre palabras y entre lenguajes. Las usaremos frecuentemente durante el curso.