UNIVERSIDAD DEL BIO-BIO

Alfabetos, Cadenas y lenguajes

Profesor: Nelson Contreras Oliva

Clase 2: Alfabetos, Cadenas y lenguajes

Alfabetos y cadenas

Un alfabeto es un **conjunto finito** no vacío cuyos elementos se llaman símbolos.

Denotamos un alfabeto arbitrario con la letra Σ.

- **Una cadena o palabra** sobre un alfabeto Σ es **cualquier sucesión finita** de elementos de Σ. Admitimos la existencia de una única cadena que no tiene símbolos, la cual se denomina cadena vacía y se denota con λ.
- La **cadena vacía** se ve representada por vacío Ø en la teoría de conjuntos.

Ejemplo Alfabeto

- Ejemplo **Sea $\Sigma = \{a, b\}$** el alfabeto que consta de los **dos símbolos a y b.**

- **Las siguientes son cadenas sobre Σ:**
    - **aba**
    - **ababaaa**
    - **aaaab.**

- Obsérvese que **aba $\neq$ aab**. **El orden de los símbolos en una cadena es significativo** ya que las cadenas se definen como sucesiones, es decir, conjuntos secuencialmente ordenados.

Ejemplo Alfabeto

- El **alfabeto Σ = {0, 1}** se conoce como alfabeto binario. Las cadenas sobre este alfabeto son secuencias finitas de ceros y unos, llamadas secuencias binarias, tales como:
  - 001
  - 1011
  - 001000001.
- **Sea Σ = {a, b, c, . . . , x, y, z},** el **alfabeto del idioma castellano**.
- Las palabras oficiales del castellano (las que aparecen en el diccionario DRA) son cadenas sobre Σ.

Ejemplo Alfabeto

- El **conjunto de todas las cadenas** sobre un alfabeto Σ, incluyendo la cadena vacía, se denota por $\Sigma*$

- Sea $\Sigma = \{a, b, c\}$, entonces

$$\Sigma* = \{\lambda, a, b, c, aa, ab, ac, ba, bb, bc, ca, cb, cc, aaa, aab, abc, baa, . . .\}.$$

UNIVERSIDAD DEL BIO-BIO

Notación a utilizar

Notación usada en la teoría de lenguajes	
Σ, Γ	denotan alfabetos.
Σ^*	denota el conjunto de todas las cadenas que se pueden formar con los símbolos del alfabeto Σ.
$a, b, c, d, e, \ldots$	denotan símbolos de un alfabeto.
$u, v, w, x, y, z, \ldots$ $\alpha, \beta, \gamma, \ldots$	denotan cadenas, es decir, sucesiones finitas de símbolos de un alfabeto.
λ	denota la cadena vacía, es decir, la única cadena que no tiene símbolos.
$A, B, C, \ldots, L, M, N, \ldots$	denotan lenguajes (definidos más adelante).

Concatenación de Cadenas

- **Dado un alfabeto Σ** y **dos cadenas u, v ∈ Σ∗**, la concatenación de u y v se denota como ***u · v o simplemente uv*** y se define descriptivamente así:

- **1. Si v = λ, entonces u · λ = λ · u = u.**

Es decir, la concatenación de cualquier cadena u con la cadena vacía, a izquierda o a derecha, es igual a u.

- **2. Si u = {$a_1, a_{2,\ldots,}\ a_n$, v =$b_1, b_{2,\ldots,}b_n$}, entonces u · v = {$a_1, a_{2,\ldots,}\ a_n b_1, b_{2,\ldots,}b_n$}.**

Es decir, u · v es la cadena formada escribiendo los símbolos de u y a continuación los símbolos de v.

Propiedades Concatenación

- **La concatenación de cadenas se puede definir** inductiva o recursivamente de la siguiente manera.
- **Si $u, v \in \Sigma*$, $a \in \Sigma$, entonces:**

1. $u \cdot \lambda = \lambda \cdot u = u$.
2. $u \cdot (va) = (u \cdot v)a$.

Propiedad: La concatenación de cadenas es una operación asociativa.

Es decir, si $u, v, w \in \Sigma*$, entonces $(uv)w = u(vw)$.

Demostración: Se puede hacer escribiendo explícitamente las cadenas u, v, w y usando la definición descriptiva de concatenación. También se puede dar una demostración inductiva usando la definición recursiva de concatenación.

UNIVERSIDAD DEL BIO-BIO
5 Años Desde Agosto 2014 Hasta Agosto 2019 ACREDITADA
ubiobio.cl

Potencias de una cadena

- **Dada u ∈ Σ∗ y n ∈ N**, se define (descriptivamente) u^n en la siguiente forma:
  - $u^0 = \lambda$,
  - $u^n = \underbrace{uu \ldots u}_{\text{n veces}}$

Longitud de una cadena

- La longitud de una cadena **u ∈ Σ∗ se denota |u|** y se define como el numero de símbolos de u (contando los símbolos repetidos). Es decir,

$$|u| = \begin{cases} 0, & \text{si } u = \lambda, \\ n, & \text{si } u = a_1, a_2 \ldots a_n \end{cases}$$

- $|aba| = 3$, $|baaa| = 4$

UNIVERSIDAD DEL BIO-BIO

Longitud de una cadena

Ejemplo Si $w \in \Sigma^*$, $n, m \in \mathbb{N}$, demostrar que

$$|w^{n+m}| = |w^n| + |w^m|$$

Solución:

Caso $n, m \geq 1$. $|w^{n+m}| = |\underbrace{ww\cdots w}_{n+m \text{ veces}}| = (n+m)|w|$. Por otro lado,

$$|w^n| + |w^m| = |\underbrace{ww\cdots w}_{n \text{ veces}}| + |\underbrace{ww\cdots w}_{m \text{ veces}}| = n|w| + m|w|.$$

Caso $n = 0, m \geq 1$. $|w^{n+m}| = |w^{0+m}| = |w^m|$. Por otro lado,

$$|w^n| + |w^m| = |w^0| + |w^m| = |\lambda| + |w^m| = 0 + |w^m| = |w^m|.$$

Caso $m = 0, n \geq 1$. Similar al caso anterior.

Caso $n = 0, m = 0$. $|w^{n+m}| = |w^{0+0}| = |\lambda| = 0$. Por otro lado,

$$|w^n| + |w^m| = |w^0| + |w^0| = |\lambda| + |\lambda| = 0 + 0 = 0.$$

UNIVERSIDAD DEL BIO-BIO

Reflexión o inversa de una cadena

La reflexión o inversa de una cadena u ∈ Σ∗ se denota u^R y se define descriptivamente así:

$$u^R = \begin{cases} \lambda, & \text{si } u = \lambda, \\ a_n \ldots a_2 a_1, & \text{si } u = a_1 a_2 \ldots\ a_n. \end{cases}$$

De la definición se observa claramente que la reflexión de la reflexión de una cadena es la misma cadena, es decir,

$$\left(u^R\right)^R = u, \text{ para } u \in \Sigma^*$$

Subcadenas, prefijos

Una cadena v es una **subcadena o una subpalabra** de ***u*** si existen cadenas ***x, y*** tales que ***u = xvy***.

Nótese que x o y pueden ser **λ** y, por lo tanto, la cadena vacía es una **subcadena** de cualquier cadena.

Un **prefijo** de ***u*** es una cadena **v tal que u = vw** para alguna cadena **w ∈ Σ∗.**

Se dice que **v es un prefijo propio si v ≠ u.**

Subcadenas, sufijos

Un **sufijo** de **u** es una cadena **v tal que u = wv** para alguna cadena **w ∈ Σ∗.**

Se dice que **v es un sufijo propio si v ≠ u.**

Obsérvese que **λ** es un prefijo y un sufijo de toda cadena **u** ya que **uλ = λu = u.**

- Por la misma razón, toda cadena u es prefijo y sufijo de sı misma.

Ejemplo Prefijo y Sufijos

Sea $\Sigma = \{a, b, c, d\}$, $u = bcbaadb$.

Prefijos de u :	Sufijos de u :
λ	λ
b	b
bc	db
bcb	adb
bcba	aadb
bcbaa	baadb
bcbaad	cbaadb
bcbaadb	bcbaadb

Lenguajes

Un lenguaje L sobre un alfabeto Σ es un subconjunto de **Σ∗, es decir L ⊆ Σ∗.**

Casos extremos:

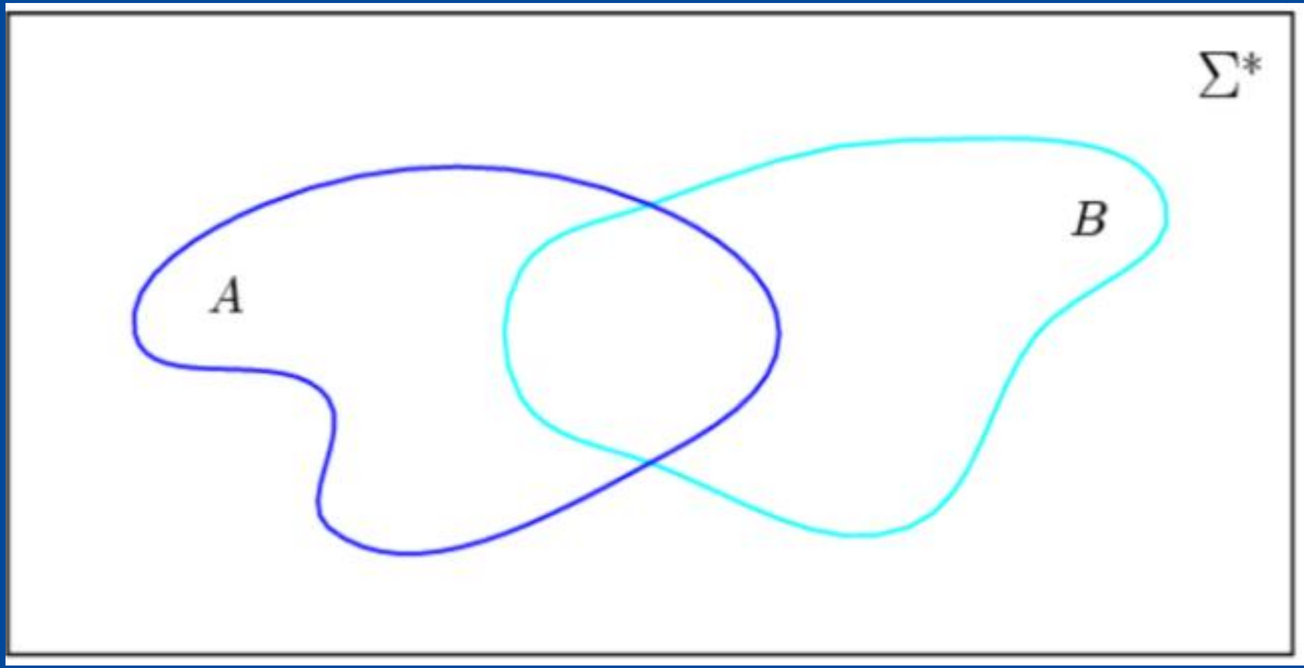
L = ∅, lenguaje vacío.
L = Σ∗, lenguaje de todas las cadenas sobre Σ.

Todo lenguaje **L satisface ∅ ⊆ L ⊆ Σ∗**, y puede ser **finito o infinito.**

Lenguajes

Los lenguajes se denotan con letras mayúsculas A, B, C, . . . , L, M, N, ...

En la siguiente grafica se visualizan dos lenguajes A y B sobre Σ.



Ejemplo Lenguajes

Σ = {a, b, c}. L = {a, aba, aca}.

Σ = {a, b, c}. L = {a, aa, aaa, . . .} = $\{a^n: n \geq 1\}$.

Σ = {a, b, c}. L = $\{\lambda, aa, aba, ab^2a, ab^3a, \ldots\} = \{ab^na : n \geq 0\} \cup \{\lambda\}$.

Σ = {a, b, c, . . . , x, y, z}. L = $\{u \in \Sigma{*}$: u aparece en el diccionario español DRA}.
L es un lenguaje finito.

Σ = {a, b, c}. L = $\{u \in \Sigma{*}$: u no contiene el símbolo c}.
Por ejemplo, abbaab ∈ L pero abbcaa ∉ L.

UNIVERSIDAD DEL BIO-BIO
5 Años Desde Agosto 2014 Hasta Agosto 2019
ACREDITADA
ubiobio.cl

Ejemplo Lenguajes

$\Sigma = \{0, 1\}$.
L = conjunto de todas las secuencias binarias que contienen un numero impar de unos.

$\Sigma = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$.
El conjunto $\mathbb{N}$ de los números naturales se puede definir como un lenguaje sobre Σ, en la siguiente forma:

$\mathbb{N} = \{u \in \Sigma*: u = 0$ ó 0 no es un prefijo de $u\}$.