ANÁLISIS COMPARATIVO DE LENGUAJES

La educación genera confianza. La confianza genera esperanza. La esperanza genera paz.-Confucio.

M. Cs. Patricia Roggero

Universidad Nacional de San Luis

Año 2021

Aspectos Formales de los Lenguajes de Programación

Notas de Clase

Capítulos III – Programming Languages – Design and Implementation – Terrence Pratt

Capítulos III Concepts of Programming Languages – Robert Sebesta

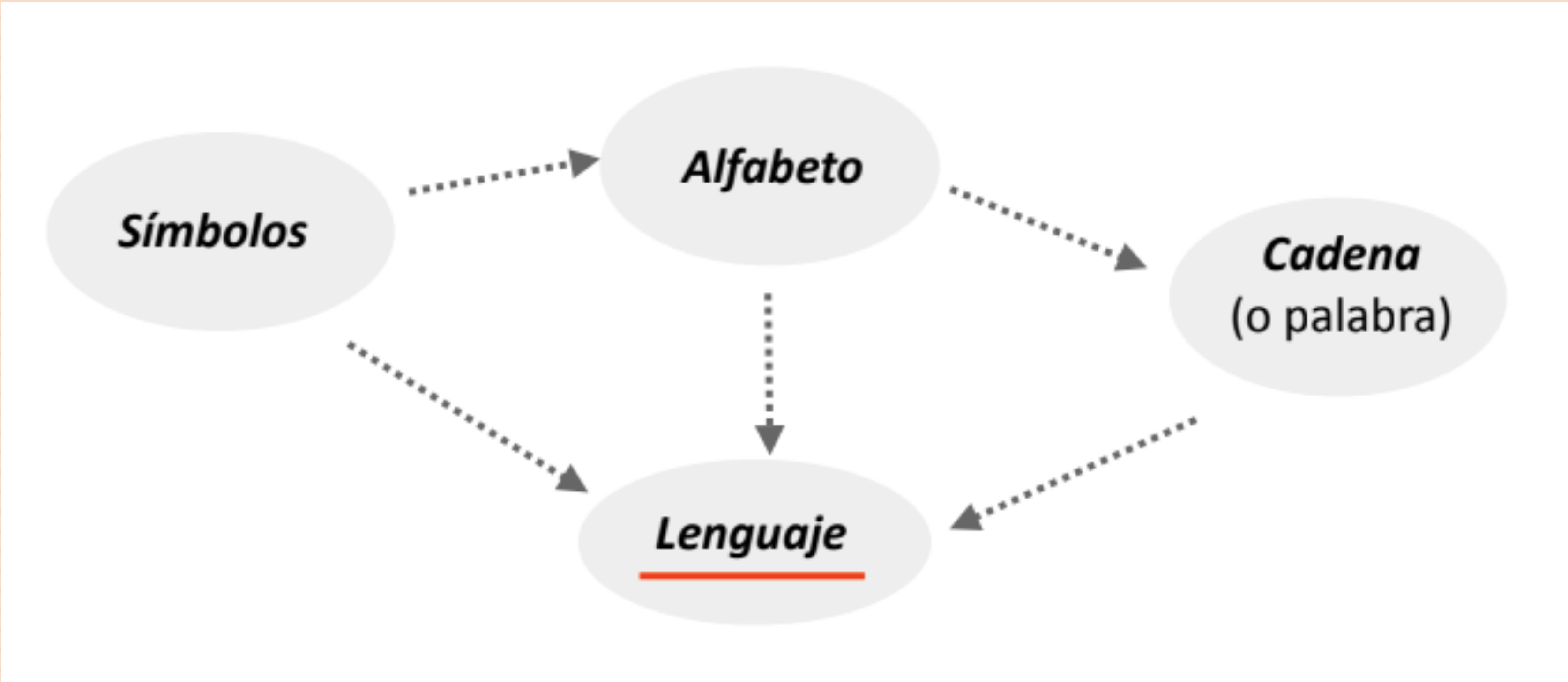
Apunte de la cátedra.

Modelos para la descripción de Lenguajes Formales

- A partir de ahora es nuestro objetivo estudiar métodos que permitan la descripción formal de lenguajes, es decir la sintaxis.
- Chomsky (1950, 1959). Lingüista, estudió la naturaleza teórica de los lenguajes naturales.
- En particular nuestro interés es estudiar cómo se definen los lenguajes de programación.
- Para comenzar debemos dar algunas definiciones previas necesarias para poder definir modelos formales:

Definiciones y conceptos introductorios

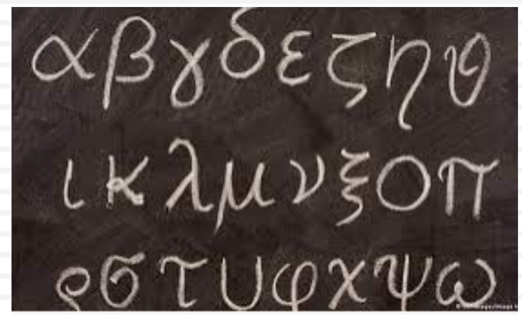
Cadenas, alfabetos y lenguajes



Definiciones y conceptos introductorios

Cadenas, alfabetos y lenguajes

Un **símbolo** es una entidad abstracta a la que no definiremos formalmente, como el punto en geometría. Letras y dígitos son ejemplos de símbolos frecuentemente usados.



Definiciones y conceptos introductorios

Cadenas, alfabetos y lenguajes

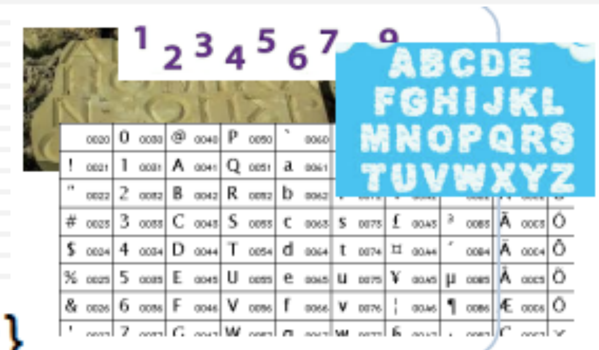
Un **alfabeto** es un conjunto finito, no vacío de símbolos. Se utiliza comúnmente el símbolo Σ, para denotarlo.

$\Sigma_1 = \{a,b\}$

$\Sigma_2 = \{0,1\}$

$\Sigma_3 = \{a,b,c,d,\ldots,z\}$

$\Sigma_4 = \{0,1,2,\ldots,9\}$



UNICODE

ASCII

Definiciones y conceptos introductorios

Cadenas, alfabetos y lenguajes

Una **cadena** o **palabra** es una secuencia finita de símbolos. Por ejemplo, *a*, *b* y *c* son símbolos pertenecientes a un alfabeto y *cccab* es una cadena.

Ejemplos

$\Sigma_1 = \{a,b\}$

x_1 = aabbbaa; x_2 = bbbaa ; ….

$\Sigma_2 = \{0,1\}$

x_1 = 1011011; x_2 = 1100 ; ….

Definiciones y conceptos introductorios

Cadenas, alfabetos y lenguajes

Definiciones y conceptos introductorios

Cadenas, alfabetos y lenguajes

- ***Cadena vacía***

  Es la cadena con 0 ocurrencias de símbolos y denotada por la letra griega λ

  (sin importar el alfabeto)

Ejemplos

$\Sigma_1 = \{a,b\}$

$x = \lambda$

$\Sigma_2 = \{0,1\}$

$x = \lambda$

Definiciones y conceptos introductorios

Cadenas, alfabetos y lenguajes

Cadenas: Operaciones sobre cadenas

Longitud de una cadena denotada por $|x|$

Cantidad de símbolos de la cadena x

$x = a_1 a_2 a_3 \dots a_n$ → $|x| = |a_1 a_2 a_3 \dots a_n| = n$

Ejemplo

$|x| = |1011011| = 7$

$|x| = |123+32*339+57/65| = 16$

Definiciones y conceptos introductorios

Cadenas, alfabetos y lenguajes

Cadenas: Operaciones sobre cadenas

Longitud de una cadena: cadena nula

$$x = \lambda \quad \Longrightarrow \quad |x| = |\lambda| = 0$$

Definiciones y conceptos introductorios

Cadenas, alfabetos y lenguajes

Cadenas: Operaciones sobre cadenas

Concatenación de cadenas:

Es una operación binaria que permite concatenar cadenas (se colocan los símbolos de una cadena y a continuación de los símbolos de la cadena x)

$x = a_1\, a_2\, a_3 \ldots a_n$

$y = b_1\, b_2\, b_3 \ldots b_m$

$\Rightarrow$

$x\, y = a_1\, a_2\, a_3 \ldots a_n\ b_1\, b_2\, b_3 \ldots b_m$

$|x\, y| = n + m$

La cadena vacía (λ) es el neutro de la concatenación

$$\lambda x = x\lambda = x$$

Definiciones y conceptos introductorios

Cadenas, alfabetos y lenguajes

Cadenas: Operaciones sobre cadenas

Potencia sobre cadenas:

Potencia sobre Cadenas $\Rightarrow$ x^n

$$x^0 = \lambda$$

$$x^n = x^{n-1}\, x$$

Ejemplos:

Si x= abaa entonces:

x^2= abaaabaa

x^3= abaaabaaabaa

¿Cómo se obtiene x^3 según la definición?

Definiciones y conceptos introductorios

Cadenas, alfabetos y lenguajes

Cadenas: Operaciones sobre cadenas

Potencia sobre Alfabetos Σ^n

Es el conjunto de todas las cadenas de longitud n cuyos símbolos están en Σ.

$\Sigma^0 = \{ \lambda \}$

$\Sigma^n = \{ x_1, x_2, ..., x_i, ... \}$ donde $\forall x_i \ |x_i| = n$

Ejemplos:

1. Notar que $\Sigma^0 = \{\lambda\}$, independientemente de lo que sea Σ. Dado que λ es la única cadena cuya longitud es 0.
2. Si $\Sigma = \{0, 1\}$, entonces $\Sigma^1 = \{0, 1\}$,
3. $\Sigma^2 = \{00, 01, 10, 11\}$,
4. $\Sigma^3 = \{000, 001, 010, 011, 100, 101, 110, 111\}$ y así siguiendo.

Definiciones y conceptos introductorios

Cadenas, alfabetos y lenguajes

Lenguajes:

Un lenguaje L es un conjunto (finito o infinito) de cadenas sobre algún alfabeto Σ

Ejemplos

$\Sigma_1 = \{a,b\}$

$L_1 = \varnothing$; $L_2 = \{aabbbaa, bbaa, bb, b, bbbaaaabbb, ...\}$; ....

$\Sigma_2 = \{0,1\}$

$L_1 = \{0,1\}$; $L_2 = \{1001, 100, 1, 000, 11111, ...\}$; ...

Observación: el lenguaje L no necesariamente debe incluir todos los símbolos del alfabeto

Definiciones y conceptos introductorios

Cadenas, alfabetos y lenguajes

Otros ejemplos de Lenguajes

Definiciones y conceptos introductorios

Cadenas, alfabetos y lenguajes

Lenguajes: Lenguaje Referencial

Es el conjunto de todas las cadenas sobre algún alfabeto Σ y es denotado por Σ^*

$$\Sigma^* = \Sigma^0 \cup \Sigma^1 \cup \Sigma^2 \cup \ldots = \{\lambda\} \cup \Sigma^1 \cup \Sigma^2 \cup \ldots = \bigcup_{n \in (\mathbb{N} \cup \{0\})} \Sigma^n$$

Ejemplo:

$\Sigma = \{0,1\}$ $\quad \Sigma^* = \{\lambda, 0, 1, 00, 11, 010, 000, \ldots\}$

Definiciones y conceptos introductorios

Cadenas, alfabetos y lenguajes

- Σ^* es un conjunto infinito de cadenas de longitud finita, ya que Σ es un conjunto finto no vacío de símbolos.

- Para cualquier lenguaje L se cumple que $L \subseteq \Sigma^*$

- Y además:

$$\Sigma^+ = \Sigma^1 \cup \Sigma^2 \cup \Sigma^3 \cup \ldots = \bigcup_{n \in \mathbb{N}} \Sigma^n$$

$$\Sigma^+ = \Sigma^* - \Sigma^0$$

Definiciones y conceptos introductorios

Cadenas, alfabetos y lenguajes

Lenguajes: formas de describirlos

Primera opción: descripción del conjunto por extensión

↓

$$L = \{x_1, x_2, x_3, x_4, \ldots\}$$

Ejemplo:

$\Sigma = \{b, c\}$ $L = \{c, cc, ccc, cccc\}$

Definiciones y conceptos introductorios

Cadenas, alfabetos y lenguajes

Lenguajes: *formas de describirlos*

Segunda opción: descripción del conjunto por comprensión

$$L = \{x \in \Sigma^* / \mathcal{P}(x)\}$$

Ejemplo:

$\Sigma = \{a, b\}$

$$L = \{x \in \Sigma^* / \overbrace{x \text{ contiene una cantidad par del símbolo a y no contiene el símbolo b}}^{\mathcal{P}(x)}\}$$

Definiciones y conceptos introductorios

Cadenas, alfabetos y lenguajes

Lenguajes: formas de describirlos

Otra opción: remplazar x por una expresión con parámetros y describir el lenguaje estableciendo las condiciones sobre los parámetros.

Ejemplo

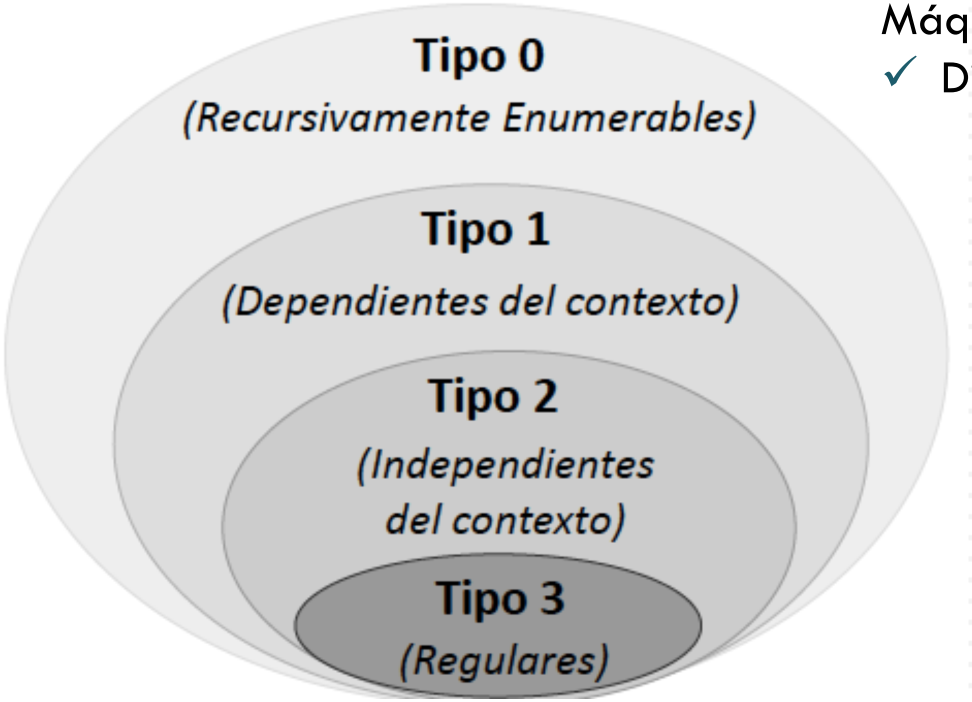
$\Sigma = \{a, b\}$

$$L = \{\overbrace{a^n}^{\text{(expresión, donde } n \text{ es el parámetro)}} / \underbrace{n \in \mathbb{N} \text{ y } n \text{ es par}}_{\text{(condición sobre el parámetro } n\text{)}}\}$$

Observación: también es posible describir un lenguaje, utilizando lenguaje natural.

Lenguajes Formales

Jerarquía de Chomsky



Modelos descriptores de Lenguajes:

- ✓ Dispositivos reconocedores: Autómatas o Máquinas abstractas o Máquinas de Estados
- ✓ Dispositivos generadores: Gramáticas

Noam Chomsky

Dispositivos descriptores de Lenguajes

Gramáticas: es un modelo matemático que permite generar a través de reglas sintácticas o gramaticales, cadenas miembros de un leguaje específico.

Autómatas: es un modelo matemático que representa la idea de computación o manipulación de cadenas vía la aplicación de acciones preestablecidas. Tiene como objetivo (en general), determinar la pertenencia de una cadena a un lenguaje específico.

Lenguajes Regulares

Lenguajes Regulares

$\Sigma = \{0, 1\}$

Finitos

$L = \{\lambda, 00, 11\}$

$L = \{0^i1 / 1 \leq i <= 5\}$

Infinitos

$L = \{x \in \Sigma^* \ / \ x \text{ comienza con } 0\}$

$L = \{1^k \ / \ k > 0\}$

$L = \{0^i1^k \ / \ i \geq 1 \text{ y } k > 0\}$

Lenguajes Regulares

- Otros ejemplos de lenguajes regulares son:

Los nombres de los identificadores, palabras claves, números, etc., en los lenguajes de programación

Expresiones Regulares (ER)

- Las **ER** son un tipo de notación particular que permite describir **lenguajes regulares**.
- Ofrecen una forma declarativa de expresar las cadenas.
- Se construyen utilizando los siguientes elementos:

- **operandos:**

  los símbolos σ *en* Σ

  la palabra vacía (λ)

  el conjunto vacío ($\emptyset$)

  expresiones regulares

- **operadores:**

  unión (+)

  concatenación (.)

  clausura de Kleene (*)

Expresiones Regulares (ER)

- La ER se definen recursivamente de la siguiente manera:

$\emptyset$ es una ER y denota el conjunto vacío, $L(\emptyset) = \emptyset$.

λ es una ER y denota el conjunto $\{\lambda\}$. $L(\lambda) = \{\lambda\}$.

$\forall a \in \Sigma$, a es una ER y denota el conjunto $\{a\}$. $L(a) = \{a\}$.

Si α y β son ER, entonces:

$\alpha + \beta$ es una ER y denota el conjunto $L(\alpha + \beta) = L(\alpha) \cup L(\beta)$

$\alpha\beta$ es una ER y denota el conjunto $L(\alpha\beta) = L(\alpha)L(\beta)$

α^* es una ER y denota el conjunto $L(\alpha^*) = (L(\alpha))^*$

(α) es una ER y denota el conjunto $L((\alpha)) = L(\alpha)$

Ninguna otra cosa es una ER.

Expresiones Regulares (ER)

La **Precedencia** de los operadores (de mayor a menor): clausura, concatenación y unión.
La **Asociatividad**: de izquierda a derecha.

Ejemplo (1):

Considerando $\Sigma = \{a, b\}$

- a es una ER que denota el lenguaje $L = \{a\}$
- b es una ER que denota el lenguaje $L = \{b\}$
- $a + b$ es una ER que denota el lenguaje $L = \{a\} \cup \{b\} = \{a, b\}$
- ab es una ER que denota el lenguaje $L = \{a\}.\{b\} = \{ab\}$
- $(ab)^*$ es una ER que denota el lenguaje $L = \{ab\}^*$
- $(a + b)(ab)^*$ es una ER que denota el lenguaje $L = \{a, b\}\{ab\}^*$

Expresiones Regulares (ER)

Construir una ER para cada uno de los siguientes lenguajes, considerando $\Sigma = \{a, b\}$:

- $\{\lambda, aab, baa\}$
- $L = \{x \in \Sigma^* / x \text{ comienza con b y termina con a}\}$
- $L = \{x \in \Sigma^* / x \text{ contiene la subcadena ab}\}$
- El lenguaje de todas las cadenas que tienen número par de símbolos.

Expresiones Regulares (ER)

- $(\lambda + aab + baa)$
- $b\ (a + b)^*\ a$
- $(a + b)^*\ ab\ (a + b)^*$
- $(aa + ab + ba + bb)^*$

¿Dudas?