UNIVERSIDAD DEL BIO-BIO

Lenguajes Formales

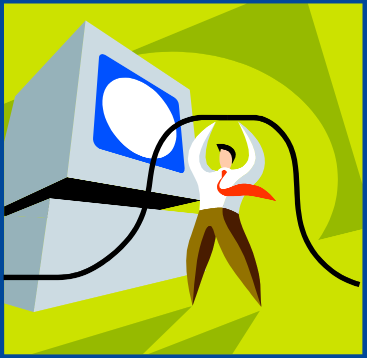
Profesor: Nelson Contreras Oliva

Clase 3: Lenguajes Formales

UNIVERSIDAD DEL BIO-BIO

Ciencias de la computación:

- Son aquellas que abarcan el estudio de las bases teóricas de la información y la computación y su aplicación en sistemas computacionales.



UNIVERSIDAD DEL BIO-BIO

- Teoría de Autómatas

  Estudia de máquinas de Computación abstractas.

- Teoría de la Computación.

  Estudia la computabilidad y la complejidad.

- Lenguajes y Gramáticas.

  Estudia la formalización de los lenguajes.

Introducción: Lenguajes

- El lenguaje es el medio que facilita la comunicación para los seres humanos.

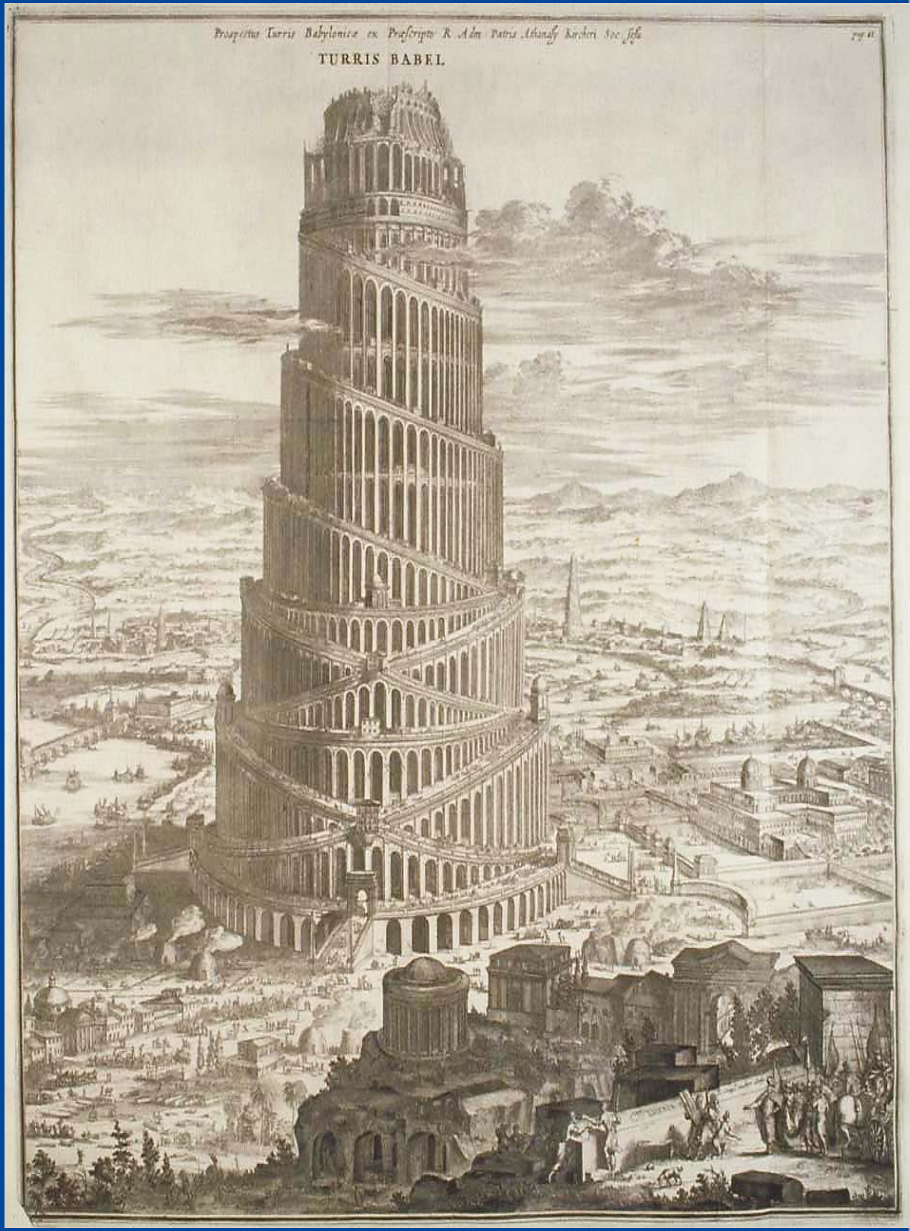
- Existen distintos lenguajes: corporal, de signos, de programación, oral, escrito….

- La clasificación inicial es entre lenguajes “naturales” y “artificiales”… aunque todos han sido creados por el hombre.

UNIVERSIDAD DEL BIO-BIO

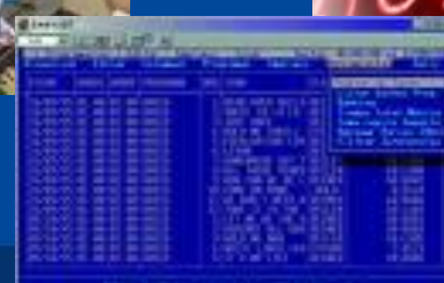
Lenguaje Natural:

comunicación en algún idioma específico, compartido por un grupo humano.



Teoría de Lenguajes Formales

Por contraposición a los “lenguajes naturales”, se llamará “lenguaje formal” a cada uno de los lenguajes “artificiales” propios de las matemáticas o de la informática, incluyendo a los lenguajes de programación.



Introducción: Teoría de Lenguajes

- Se estudiará a los **Lenguajes Formales** en sus **Aspectos Sintácticos**.

Sintaxis:
La forma correcta de construcción de palabras pertenecientes a un lenguaje.

UNIVERSIDAD DEL BIO-BIO

Lenguajes Formales

- **Los lenguajes permiten la comunicación con la máquina, parte de lo que puede hacer la máquina depende del poder descriptivo del lenguaje.**

  - Compiladores.
  - Traductores.
  - Diseño de lenguajes de alto nivel.

UNIVERSIDAD DEL BIO-BIO

Lenguajes Formales

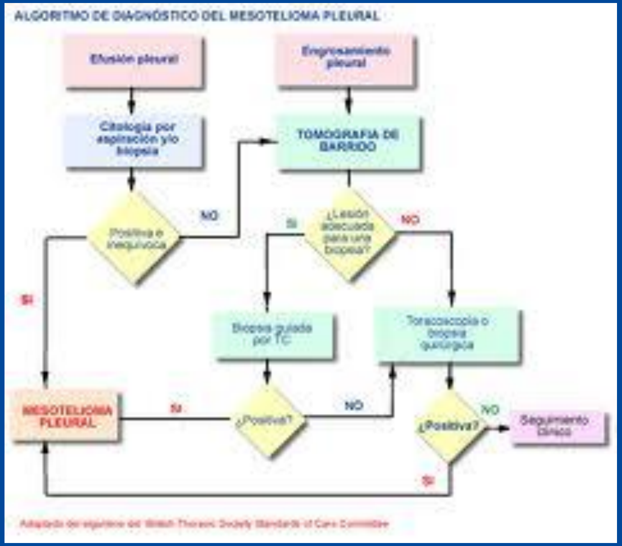
- Los idiomas (lenguajes naturales) son también un producto generado por los humanos, pero su creación a lo largo de siglos no fue plenamente consciente.

- la creación de lenguajes de programación ha sido un proceso consciente y racional, desarrollado formalmente.

Lenguajes Formales

- Los lenguajes formales no se hablan, son escritos. Permiten dar instrucciones a una máquina…

- Sin embargo, existe un área de estudio importante en el procesamiento de lenguaje natural…



UNIVERSIDAD DEL BIO-BIO

Lenguajes Formales

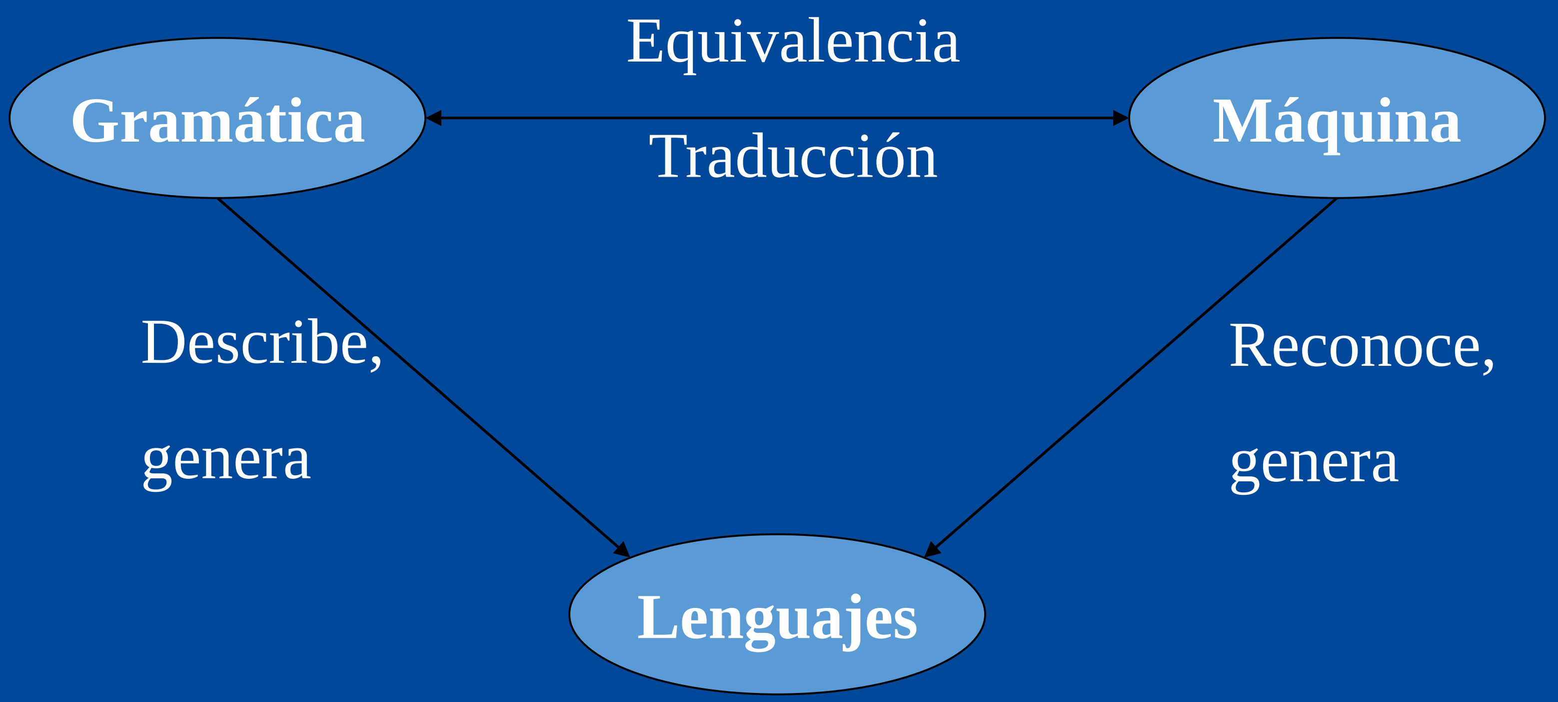
- En matemáticas, lógica y ciencias de computación, un **lenguaje formal** es un conjunto de palabras (palabras de caracteres) de longitud finita formadas a partir de un **alfabeto predefinido** (conjunto de caracteres) finito.

Lenguajes Formales

- Son lenguajes porque las estructuras que se forman tienen reglas (gramática, sintáxis) e interpretación semántica (significado) de manera similar a los lenguajes hablados.

Lenguajes Formales, Gramática, Máquinas de estados finitos.

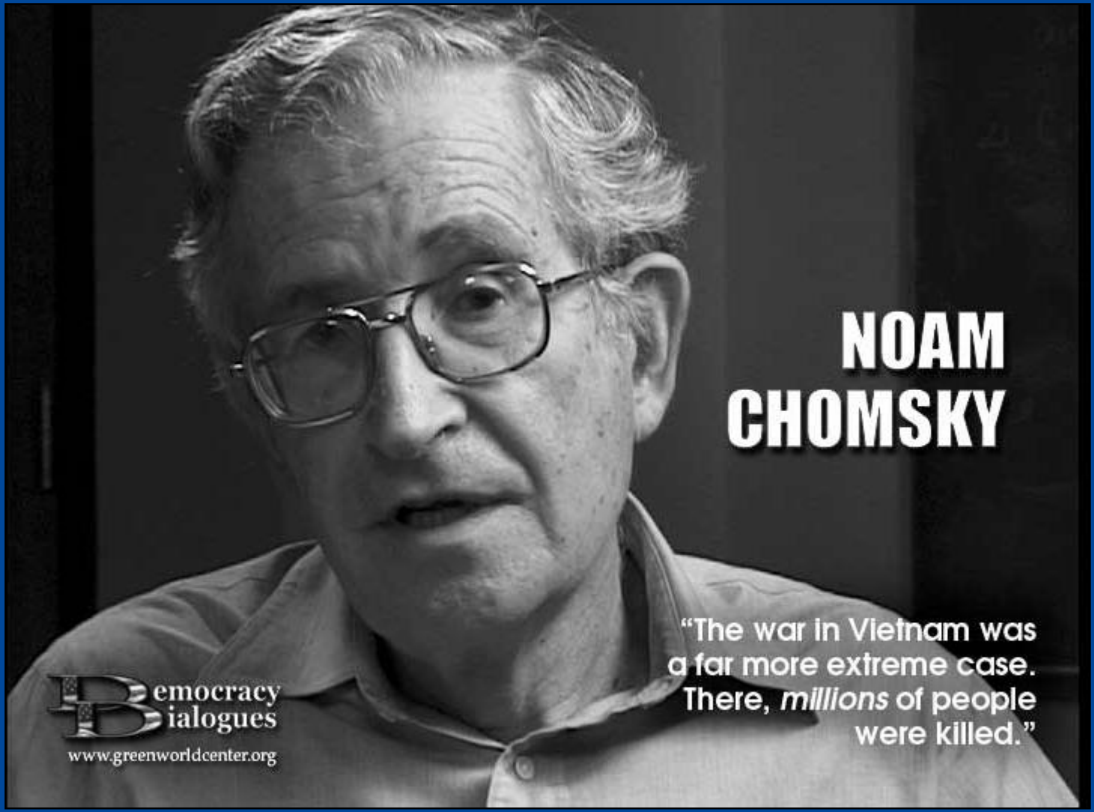
(Máquinas abstractas - Gramáticas Formales)



Clasificación de Lenguajes

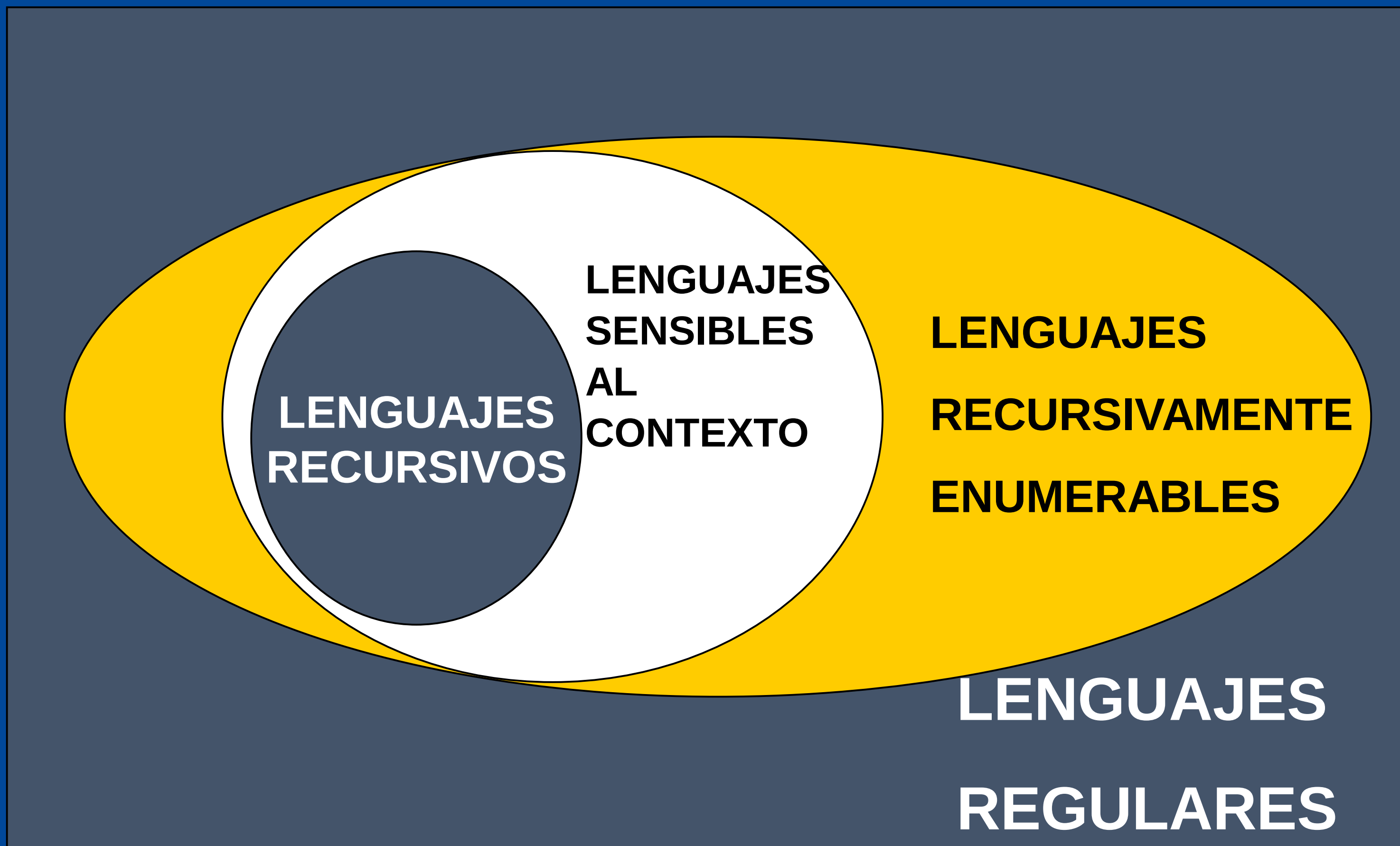
- Se llama "clase de lenguajes" a los conjuntos de lenguajes que comparten una cierta propiedad dada.

Jerarquía de Chomsky



En 1956, Noam Chomsky formuló una Jerarquía para organizar los distintos tipos de lenguaje formal.
Hoy en día, sigue vigente.

Jerarquía de Chomsky



Gramática

- Es el conjunto de reglas y principios que gobiernan el uso de un lenguaje determinado.

- Un área de interés para la gramática es la sintáxis del lenguaje.

- En lenguajes formales, las gramáticas estarán orientadas a la sintáxis del lenguaje.

Gramática

- Cada lenguaje tiene su propia gramática.
- Considerando la jerarquía de Chomsky, existirán gramáticas: regulares, libres del contexto y recursivamente enumerables.

Lenguajes Formales

(Máquinas abstractas - Gramáticas Formales)

Gramáticas	Lenguajes	Máquinas
Sin restricciones o de Tipo 0	Sin restricciones o de Tipo 0	Máquina de Turing
Sensible al contexto o de Tipo 1	Sensible al contexto o de Tipo 1	Autómata linealmente acotado
Recursivamente enumerables, Tipo 2	Libre de contexto o de Tipo 2	Autómata con pila
Regular o de Tipo 3	Regular o de Tipo 3	Autómata Finito Determinístico

Alfabetos.

Formalmente un a**lfabeto** es un conjunto *no vacío* y finito de **símbolos**.

A = {a, b} es definido como un alfabeto.

Una **palabra (string, cadena)** cualquiera sobre este alfabeto sería:

w1 = aaabbb

w2 = ababab

w3 = bbba, ….

Operaciones con palabras

- Si $\boldsymbol{w}$ es una palabra sobre cualquier alfabeto, su ***longitud*** se denota mediante $|\boldsymbol{w}|$ y corresponde al total de símbolos de la palabra.

Ej: Sea $\boldsymbol{w}$ = abaca sobre $\Sigma = \{a, b, c\}$; entonces

$$|\boldsymbol{w}| = 5$$

$$|\boldsymbol{w}|_a = 3$$

$$|\boldsymbol{w}|_b = 1$$

$$|\boldsymbol{w}|_c = 1$$

$$|\boldsymbol{\varepsilon}| = 0$$

Potencia de una palabra

$$w^n = \begin{cases} \varepsilon, & si \quad n = 0 \\ ww^{n-1}, si & n > 0 \end{cases}$$

Ej. Si ***w*** = 010 sobre Σ = {0,1}, entonces:

$w^0 = \varepsilon$

$w^1 = ww^0 = 010\varepsilon = 010$

$w^2 = ww^1 = 010010$

$w^3 = ww^2 = 010010010$

$w^i = ww^{i-1}$

Concatenación de palabras

- Si ***w*** y ***x*** son palabras, su ***concatenación*** **w.x** se logra al unir ***x*** a la palabra ***w***:

w = saca, x = corcho → w.x = sacacorcho

$$|w.x| = |w| + |x|$$

En el ejemplo:

$$|w|=4,\ |x| =6 \rightarrow |w.x| = |w| + |x| =10$$

UNIVERSIDAD DEL BIO-BIO

identidad en la concatenación

- ***ε*** se comporta como la ***identidad*** en la concatenación:

$$\boldsymbol{\varepsilon . w = w . \varepsilon = w}$$

UNIVERSIDAD DEL BIO-BIO

Concatenación de palabras

- Si ***w*** y ***v*** son palabras, la ***concatenación:***

$$\boldsymbol{w.v \neq v.w}$$

$$\boldsymbol{|w.v| = |w| + |v|}$$

Inversa, Reversa o Transpuesta de una palabra

$$w^R = \begin{cases} w, & si \quad w = \varepsilon \\ y^R a, & si\ w = ay,\ a \in \Sigma,\ y \in \Sigma^* \end{cases}$$

- ***Si*** $x = abc$, x^R ***será:*** $\quad x^R = (bc)^R a$

$$= (c)^R ba$$

$$= (\varepsilon)^R cba$$

$$= cba$$

- *Si* $x = wy$ entonces, $x^R = (wy)^R = y^R w^R$

Sufijo, prefijo y subpalabras

- Toda palabra podría considerarse prefijo de sí misma, luego se denominará ***prefijo*** a la palabra $\boldsymbol{x}$ que es prefijo de otra $\boldsymbol{w}$ pero que no es igual a ella.

- La palabra vacía ε es prefijo de cualquier palabra.

- Una palabra $\boldsymbol{w}$ es subpalabra de otra $\boldsymbol{z}$ si existen las palabras $\boldsymbol{x}$ e $\boldsymbol{y}$ tal que $\boldsymbol{z = xwy}$

Sufijo, prefijo y subpalabras

- Toda palabra podría considerarse sufijo de sí misma, luego se denominará ***sufijo*** a la palabra ***x*** que es sufijo de otra ***w*** pero que no es igual a ella.

- La palabra vacía ε es sufijo de cualquier palabra.

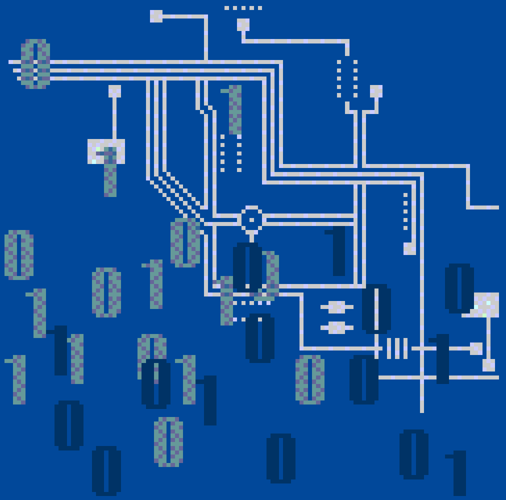
Cerradura o lenguaje universal de **A**

- El lenguaje compuesto por todas las palabras sobre el alfabeto **A** se denomina ***cerradura*** *de* **A** *o* ***lenguaje universal*** sobre **A** y se denota por **A***.

UNIVERSIDAD DEL BIO-BIO

Lenguajes

- Sea **A** un alfabeto; cualquier subconjunto L de **A*** se dice que es un **lenguaje L** con alfabeto **A**.



UNIVERSIDAD DEL BIO-BIO

Ejemplos de alfabetos:

- Alfabeto de dígitos decimales *D={0,1,2,3,4,5,6,7,8,9};*
- Alfabeto de dígitos binarios *B={0,1}*
- Alfabeto de dígitos hexadecimales

H={*0,1,2,3,4,5,6,7,8,9,A,B,C,D,E,F*}

- Alfabeto de las caracteres

  C={a,b,...z,A,...Z, ?,!...,,$}*

UNIVERSIDAD DEL BIO-BIO

Lenguajes Formales:

- Sea A = {a, b} y w = aaabbb

Así, un lenguaje sobre A que incluyera a “w”, sería:

- L= el conjunto de todas las palabras que contienen el mismo número de símbolos “a” que “b”.

Entonces, ¿es equivalente el siguiente lenguaje a la definición anterior?:

- L={w $\in$ A* / w=tiene igual número de a’s que de b’s}
- L={w $\in$ A* / w= an bn con n $\geq$ 0 }

Lenguajes Formales:

¿L=L'?

L={w $\in$ A* / w=tiene igual número de a's que de b's}

L'={w $\in$ A* / w= a^n b^n con n $\geq$ 0 }

No, porque palabras como: abab, baba, ba,…. No estarían consideradas en la definición.

Se debería decir entonces que

L'={w $\in$ A* / $|w|_a = |w|_b$}

Es una definición formal correcta para L.

Lenguajes Formales

- La **palabra vacía** (la palabra de longitud cero) se permite en este tipo de lenguajes, notándose frecuentemente mediante **ε** ó **λ**

- Un alfabeto es un conjunto finito, cada palabra tiene una longitud también finita.

- Un lenguaje puede estar compuesto por un número infinito (∞) de palabras.

Lenguajes Formales

- Si ***L*** es el lenguaje formado por algunas de las palabras sobre **A,** si ***w*** es una palabra sobre **A** y está en L, entonces se dice que ***w*** es un elemento de L.

$$w \in L$$

Concatenación de Lenguajes

- Sean A y B dos lenguajes sobre un alfabeto, el lenguaje concatenación de A y B será:

$$A.B = \{ w.x \,/\, w \in A \text{ y } x \in B \}$$

- A.B estará formado por todas las palabras que se forman concatenando cada palabra de A con todas las palabras de B.

Ej. Si A = {0,1} y B={00, 01, 11}

A.B = { }

Concatenación

- Los lenguajes a concatenar pueden ser sobre diferentes alfabetos:

Si A es un lenguaje sobre Σ_1 y B es un lenguaje sobre Σ_2, entonces A.B será un lenguaje sobre $\Sigma_1 \cup \Sigma_2$.

UNIVERSIDAD DEL BIO-BIO

Concatenación

- Para cualquier lenguaje A,

$$A.\{\varepsilon\} = \{\varepsilon\}.A = A$$

El lenguaje cuyo único elemento es la palabra vacía se comporta como el elemento identidad.

Potencia de un lenguaje

Sea A un lenguaje sobre un alfabeto Σ, se define:

$$A^n = \begin{cases} \{\varepsilon\}, & \text{si } n = 0 \\ A.A^{n-1}, & \text{si } n \geq 1 \end{cases}$$

Potencia de un lenguaje

- Ejemplo: Sea $\Sigma=\{0,1\}$, entonces:

- $\Sigma^{o}=\{\varepsilon\}$
- $\Sigma^{1}=\{0,1\}$
- $\Sigma^{2}=\{00,01,10,11\}$
- $\Sigma^{3}=\{000,001,010,011,100,101,110,111\}$

- $\Sigma^{n}=\{\varepsilon,0,1, 00,01,10,11,000,001,010,011,100,101,110,111,......\}$

Potencia de un lenguaje

Por ejemplo, si A = {01} , entonces:

$$A^0 = \{\varepsilon\}$$
$$A^1 = A.A^0 = \{01\}$$
$$A^2 = A.A^1 = \{0101\}$$
$$A^3 = A.A^2 = \{010101\}$$

Ejemplo: Sea $X = \{a,b,c\}$ e $Y = \{abb, ba\}$.
Entonces

$XY = \{aabb, babb, cabb, aba, bba, cba\}$

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

$X^1 = X = \{a,b,c\}$

$X^2 = XX = \{aa,ab,ac,ba,bb,bc,ca,cb,cc\}$

$X^3 = X^2X = \{aaa,aab,aac,aba,abb,abc,aca,acb,acc,$
$baa,bab,bac,bba,bbb,bbc,bca,bcb,bcc,$
$caa,cab,cac,cba,cbb,cbc,cca,ccb,ccc\}$

Unión e Intersección de Lenguajes

- Sean A y B dos lenguajes sobre alfabeto Σ, entonces la ***unión*** de A y B, **A** $\cup$ **B** está definida por:

$$A \cup B = \{ x / x \in A \;\; o \;\; x \in B \}$$

Unión e Intersección de Lenguajes

- Si A y B son dos lenguajes sobre alfabeto Σ, entonces la ***intersección*** de A y B, $A \cap B$ está definida por:

$$A \cap B = \{\, x \,/\, x \in A \;\; y \;\; x \in B \; simult\acute{a}neamente \}$$

Ejemplo:

Considere $\Sigma = \{0, 1\}$ y los lenguajes:

$A = \{\varepsilon, a, b, ab, ba\}$; $B = \{\varepsilon, a, ba, aba, bab\}$;

Entonces:

$$A \cup B = \{\varepsilon, a, b, ab, ba, aba, bab\}$$

$$A \cap B = \{\varepsilon, a, b, ba\}$$

Sublenguaje e igualdad de Lenguajes

- Si A y B son dos lenguajes sobre un alfabeto Σ, y si todas las palabras de A son también palabras de B, entonces A es un ***sublenguaje*** de B y se denota $\mathbf{A \subseteq B}$.

- Todo lenguaje L sobre el alfabeto Σ es un sublenguaje de Σ^*, es decir $\mathbf{L \subseteq \Sigma^*}$.

Sublenguaje e igualdad de Lenguajes

- Si A y B son dos lenguajes sobre alfabeto Σ, entonces se dice que son ***iguales*** si contienen exactamente las mismas palabras y se denota como A = B.

Teoremas

- Sean A y B dos lenguajes sobre un alfabeto Σ. Entonces **A = B** si y solo si **A** $\subseteq$ **B** y **B** $\subseteq$ **A**.
- Sean A, B y C lenguajes sobre un alfabeto Σ. Entonces:
  - **A.(B $\cup$ C) = A.B $\cup$ A.C**
  - **(B $\cup$ C).A = B.A $\cup$ C.A**

UNIVERSIDAD DEL BIO-BIO

Resumiendo:

- un a**lfabeto** es un conjunto *no vacío* y finito de símbolos.

- Cada ***símbolo*** de un alfabeto es una ***palabra*** (de un elemento) sobre dicho alfabeto.

UNIVERSIDAD DEL BIO-BIO

Resumiendo:

- Una **palabra** es una concatenación finita de símbolos de un alfabeto.

- La ***palabra vacía***, denotada por **ε**, es una palabra sobre cualquier alfabeto.

Resumiendo:

- Un ***lenguaje*** es un conjunto de palabras o palabras, puede ser infinito o vacío.

- El ***lenguaje vacío*** se denota por **Ø**.

- Cabe señalar que: $\emptyset \neq \{\varepsilon\}$

Lenguajes Formales

- **La palabra vacía (la palabra de longitud cero) se permite en este tipo de lenguajes, notándose frecuentemente mediante ε ó λ**

- **Un alfabeto es un conjunto finito, cada palabra tiene una longitud también finita**

- **Un lenguaje puede estar compuesto por un número infinito (∞) de palabras.**

Resumiendo:

- **un alfabeto es un conjunto *no vacío* y finito de símbolos.**
- **Cada *símbolo* de un alfabeto es una *palabra* (de un elemento) sobre dicho alfabeto.**

Resumiendo:

- **Una palabra es una concatenación finita de símbolos de un alfabeto.**

- **La *palabra vacía*, denotada por ε, es una palabra sobre cualquier alfabeto**.

Resumiendo:

- **Un *lenguaje* es un conjunto de palabras, puede ser infinito o vacío.**
- **El *lenguaje vacío* se denota por Ø.**
- **Cabe señalar que: Ø ≠ {ε}**

Lenguajes Formales

Si se considerara al alfabeto castellano, o el de los números enteros:

$$\Sigma_1 = \{a, b, c, \cdots., z\}$$

$$\Sigma_2 = \{0,1,2,3,\cdots..\}$$

Dado que $\Sigma 1$ y $\Sigma 2$ son alfabetos, entonces $\Sigma 1 \cup \Sigma 2$ también es un alfabeto.

- Si $\Sigma_1 \cap \Sigma_2$, $\Sigma_1 - \Sigma_2$, $\Sigma_2 - \Sigma_1$ son *no vacíos*, entonces también son alfabetos.

Concatenación de Lenguajes

- **Sean A y B dos lenguajes sobre un alfabeto, el lenguaje concatenación de A y B será:**

$$A.B = \{ w.x \,/\, w \in A \text{ y } x \in B\}$$

- **A.B estará formado por todas las palabras que se forman concatenando cada palabra de A con todas las palabras de B.**

Ej. Si A = {0,1} y B={00, 01, 11}

A.B = {000, 001, 011, 100, 101, 111}

BA={000, 001, 010, 011, 110, 111}

Concatenación

- **Los lenguajes a concatenar pueden ser sobre diferentes alfabetos:**

Si A es un lenguaje sobre Σ_1 y B es un lenguaje sobre Σ_2, entonces A.B será un lenguaje sobre $\Sigma_1 \cup \Sigma_2$.

Concatenación

- **Para cualquier lenguaje A,**

$$A.\{\varepsilon\} = \{\varepsilon\}.A = A$$

El lenguaje cuyo único elemento es la palabra vacía se comporta como el elemento identidad.

Potencia de un lenguaje

Sea A un lenguaje sobre un alfabeto Σ, se define:

$$A^n = \begin{cases} \{\varepsilon\}, & \text{si } n = 0 \\ A.A^{n-1}, & \text{si } n \geq 1 \end{cases}$$

Ejemplo: Sea $\Sigma=\{0,1\}$, entonces:

$\Sigma^o=\{\varepsilon\}$
$\Sigma^1=\{0,1\}$
$\Sigma^2=\{00,01,10,11\}$
$\Sigma^3= \Sigma\ \Sigma^2$
$= \{000,001,010,011,100,101,110,111\}$

**$\Sigma^n=\{\varepsilon,0,1,$
$00,01,10,11,000,001,010,011,100,101,110,111,$
$\ldots\ldots\}$**

Potencia de un lenguaje

Por ejemplo, si A = {01} , entonces:

$$A^0 = \{\varepsilon\}$$
$$A^1 = A.A^0 = \{01\}$$
$$A^2 = A.A^1 = \{0101\}$$
$$A^3 = A.A^2 = \{010101\}$$

Unión e Intersección de Lenguajes

- **Sean A y B dos lenguajes sobre alfabeto Σ, entonces la *unión* de A y B, $A \cup B$ está definida por:**

$$A \cup B = \{\, x \,/\, x \in A \;\; o \;\; x \in B \,\}$$

Intersección de Lenguajes

- **Si A y B son dos lenguajes sobre alfabeto Σ, entonces la *intersección* de A y B, $A \cap B$ está definida por:**

$$A \cap B = \{\, x \,/\, x \in A \ \ y \ \ x \in B \ \textit{simultáneamente} \}$$

Ejemplo:

Considere $\Sigma = \{a, b\}$ y los lenguajes:

$A = \{\varepsilon, a, b, ab, ba\}$; $B = \{\varepsilon, a, ba, aba, bab\}$;

Entonces:

$$A \cup B = \{\varepsilon, a, b, ab, ba, aba, bab\}$$

$$A \cap B = \{\varepsilon, a, ba\}$$

UNIVERSIDAD DEL BIO-BIO

Sublenguaje e igualdad de Lenguajes

- Si A y B son dos lenguajes sobre un alfabeto Σ, y si todas las palabras de A son también palabras de B, entonces A es un ***sublenguaje*** de B y se denota $\mathbf{A} \subseteq \mathbf{B}$.

- Todo lenguaje L sobre el alfabeto Σ es un sublenguaje de Σ^*, es decir $\mathbf{L} \subseteq \mathbf{\Sigma^*}$.

Sublenguaje e igualdad de Lenguajes

- Si A y B son dos lenguajes sobre alfabeto Σ, entonces se dice que son ***iguales*** si contienen exactamente las mismas palabras y se denota como $A = B$.

Teoremas

- Sean A y B dos lenguajes sobre un alfabeto Σ. Entonces A = B si y solo si $A \subseteq B$ y $B \subseteq A$.

- Sean A, B y C lenguajes sobre un alfabeto Σ. Entonces:
  - $A.(B \cup C) = A.B \cup A.C$
  - $(B \cup C).A = B.A \cup C.A$

Propiedades de los Lenguajes.

- Si **L1 ,L2 ,L3** son lenguajes definidos sobre ***A**, entonces:*

- **L1 .∅ = ∅ = ∅ .L1**

- La concatenación es asociativa: **(L1 . L2). L3 = L1 . (L2. L3)**

- La concatenación no es conmutativa: **L1 . L2 ≠ L2 . L1**

- Distributiva con respecto a la Unión: **L1 .(L2∪L3) = L1 . L2∪ L1 . L3**

- No Distributiva con respecto a la Intersección:

  L1 .(L2∩L3) ≠ L1 . L2∩ L1 . L3