

# Ciencias de la Computación I

## *Expresiones Regulares – BNF*

### *Diagramas Sintácticos*

---

Ciencias de la Computación I - Filminas de Clase – Facultad Cs. Exactas – UNCPBA - 2012

## Expresiones Regulares

---

- Las expresiones regulares describen lenguajes regulares.
- Una expresión regular (también llamada patrón) es una expresión que describe cadenas sobre un alfabeto.
- Las expresiones regulares se construyen utilizando los operadores de unión (+) concatenación (.) y Clausura (\*).

<u>Ejemplos</u>	$(a+b)^*$	describe	$L = \{a, b\}^*$
	$\varepsilon + a + b + aa$	describe	$L = \{\varepsilon, a, b, aa\}$
	$a(aa)^*$	describe	$L = \{a^{2k+1} / k \geq 0\}$

Ciencias de la Computación I - Filminas de Clase – Facultad Cs. Exactas – UNCPBA - 2012

## Expresiones Regulares

- La sintaxis de las expresiones regulares cambia (respecto de la notación convencional) según las herramientas y aplicaciones consideradas.
- Se utilizan en editores de texto, bases de datos, sistemas operativos y aplicaciones para buscar y reemplazar patrones de texto; también para validación de un formato específico como fechas, identificadores, nombres de usuario, direcciones de email, etc.
- Se incluyen en lenguajes de programación (por ej. en PERL es parte de su sintaxis y en otros como PHP y JAVA, existen librerías para el uso de expresiones regulares).

Ciencias de la Computación I - Filminas de Clase - Facultad Cs. Exactas - UNCPBA - 2012

## Expresiones Regulares- Notación especial

**El Punto "."** El punto es interpretado como cualquier símbolo (caracter alfanumérico, dígito, etc.), excepto el salto de línea.

Ej: patrón: a.n en la cadena "tiene amnesia" hay coincidencia  
"captando" no hay coincidencia

**Los corchetes "[ ]"** Los corchetes agrupan caracteres. Dentro de los corchetes es posible utilizar el guión "-" para especificar rangos de caracteres.

Ej. Patrón: [A-Z] coincide si la cadena contiene cualquier letra mayúscula  
Patrón: canci[oó]n coincide si contiene canción o cancion

**Los paréntesis "(")"** tienen el mismo significado que en la notación convencional de expresiones regulares

Ej: patrón: al (este|oeste|norte|sur) de

Ciencias de la Computación I - Filminas de Clase - Facultad Cs. Exactas - UNCPBA - 2012

## Expresiones Regulares-Notación Especial

La barra "|" Se utiliza como OR, es equivalente al + en notación convencional

Ej. Patrón: a|b coincide con cualquier símbolo a o b en una cadena

El signo "\$" Representa el final de la cadena de caracteres

Ej. Patrón: a\$" coincide con toda cadena que termine con a

El acento circunflejo "^" Este caracter tiene doble funcionalidad:

1) Al inicio "^" Significa que el inicio de la cadena debe coincidir con el símbolo que lo sucede

Ej. Patrón: ^[A-Z] Había una vez tiene coincidencia

Patrón: ^a[0-9]abc\$ a5abc tiene coincidencia

2) Cuando se utiliza en conjunto con los corchetes significa NOT

Ej. Patrón: [^a-z] casa no hay coincidencia

[^a-z] @012 tiene coincidencia

Ciencias de la Computación I - Filminas de Clase – Facultad Cs. Exactas – UNCPBA - 2012

## Expresiones Regulares- Notación Especial

La barra inversa "\" Se utiliza para "marcar" el siguiente caracter de forma que deje de tener o adquiriera un significado especial:

1) En combinación con caracteres especiales como ( ., \$, \*, +, etc.)

\. — Representa al punto como caracter literal

\\$ — Representa al \$ como caracter literal

\( — Representa al ( como caracter literal

Ej: Patrón: \([0-9][0-9]\.[0-9][0-9]\\$\) en la cadena (10.99\$) coincide

2) En otros casos da significado según la combinación:

\n — Representa la "nueva línea"

\d — Representa un dígito del 0 al 9.

\w — Representa cualquier caracter alfanumérico.

\s — Representa un espacio en blanco.

\D — Representa cualquier caracter que no sea un dígito del 0 al 9.

\S — Representa cualquier caracter que no sea un espacio en blanco.

Ej: Patrón: \d\.\d\d en la cadena costó 0.01\$ tiene coincidencia

Ciencias de la Computación I - Filminas de Clase – Facultad Cs. Exactas – UNCPBA - 2012

## Exp. Regulares- Notación Especial: Cuantificadores

### Las llaves "{}"

{nro1} ó  
{nro1,nro2} ó  
{nro1,}

Entre nro1 y nro2, es la cantidad de veces consecutivas que puede aparecer la expresión anterior. El nro2 puede no especificarse.

Ej: Patrón  $\wedge d\{2\}d\{2\}d\{4\}\$$       01/01/1999 coincide  
Patrón  $\wedge d\{2,4\}$       00, 865, 9999 y 84236 coinciden

**El asterisco "\*" Se usa para encontrar la expresión anterior al mismo 0 ó más veces.**

Ej: patrón  $\wedge [a-zA-Z]^*\$$       abcZ coincide

**El signo "+" Significa que la expresión anterior al mismo debe aparecer 1 ó más veces**

Ej:  $[a-z]d^+$       coincidirá en var1, var2, var10 ... pero no con var.

**El signo de interrogación "?" Significa que la expresión anterior al mismo debe aparecer 0 ó 1 vez en la cadena**

Ej: Patrón:  $a?bb$       con la cadenas bb, abb, bbb, cd**bb** tiene coincidencia

Ciencias de la Computación I - Filminas de Clase – Facultad Cs. Exactas – UNCPBA - 2012

## Expresiones Regulares

Algunos ejemplos

Notación convencional

(aa)\*a

(a+b)\*ab(a+b)\*

(0+1+2+3+4+5+6+7+8+9) (0+1+2+3+4+5+6+7+8+9)

Notación Especial

$\wedge (aa)^*a\$$

$\wedge (a|b)^*ab(a|b)^*\$$

$\wedge [0-9]\{2\}\$$

En la página web

<http://www.regular-expressions.info/javascriptexample.html> se pueden probar expresiones regulares con notación no convencional

Ciencias de la Computación I - Filminas de Clase – Facultad Cs. Exactas – UNCPBA - 2012

## BNF (Backus Naur Form)

• Notación utilizada frecuentemente para escribir gramáticas de tipo 2 o libres del contexto. Sigue las siguientes convenciones:

- no terminales se escriben entre  $\langle \rangle$
- terminales son cadenas de caracteres sin  $\langle \rangle$
- en lugar de  $\rightarrow$  se utiliza  $::=$  que se lee “se define como”
- varias reglas del tipo

$\langle A \rangle ::= \langle B_1 \rangle$

$\langle A \rangle ::= \langle B_2 \rangle$

...

$\langle A \rangle ::= \langle B_n \rangle$

Se pueden escribir como

$\langle A \rangle ::= \langle B_1 \rangle \mid \langle B_2 \rangle \mid \dots \mid \langle B_n \rangle$

Ciencias de la Computación I - Filminas de Clase – Facultad Cs. Exactas – UNCPBA - 2012

## BNF Extendido

• Usa nuevas reglas y símbolos.

Sintaxis	Significado
$::=$	se define como
$t$	el símbolo terminal $t$
$\langle nt \rangle$	el símbolo no terminal $nt$
$(\dots)$	usado para agrupar
$*$	cero o más repeticiones del elemento anterior
$+$	una o más repeticiones del elemento anterior
$[ \dots ]$	elemento opcional
$ $	alternativa de varias formas sintácticas válidas

**Nota:** Para distinguir los metasímbolos de los terminales con el mismo caracter se usa ´ (por ej., si el lenguaje que describe el BNF incluye los paréntesis deben estar precedidos por ´ )

Ciencias de la Computación I - Filminas de Clase – Facultad Cs. Exactas – UNCPBA - 2012

## BNF Extendido

### Ejemplos:

$L_1 = \{ x / x \in \{ \text{begin}, \text{end} \}^* \text{ y } x \text{ es una cadena de begin end balanceados} \}$

### BNF para $L_1$

$\langle \text{lista} \rangle ::= \langle \text{anidados} \rangle \mid \langle \text{anidados} \rangle \langle \text{lista} \rangle$

$\langle \text{anidados} \rangle ::= \text{begin } \langle \text{lista} \rangle \text{ end} \mid \text{begin end}$

### BNF Extendido para $L_1$

$\langle \text{lista} \rangle ::= \langle \text{anidados} \rangle \mid \langle \text{anidados} \rangle^* \quad \text{ó} \quad \langle \text{lista} \rangle ::= \langle \text{anidados} \rangle^+$

$\langle \text{anidados} \rangle ::= \text{begin } \langle \text{lista} \rangle \text{ end} \mid \text{begin end}$

Ciencias de la Computación I - Filminas de Clase - Facultad Cs. Exactas - UNCPBA - 2012

## Diagramas Sintácticos

- Son construcciones gráficas que permiten presentar la misma información sintáctica que las gramáticas libres del contexto y la notación BNF.
- Un diagrama sintáctico está formado por:
  - un conjunto de rectángulos etiquetados con nombres de símbolos no terminales
  - un conjunto de círculos o elipses etiquetados con nombres de símbolos terminales (tokens),
  - un conjunto de arcos que representan las secuencias en las que pueden combinarse los terminales y no terminales de la gramática.
- Cada diagrama representa un símbolo no terminal.
- Todo diagrama posee un punto de entrada (generalmente situado a la izquierda,) y un punto de salida (a la derecha), que están representados por un arco sin origen y un arco sin destino respectivamente.

Ciencias de la Computación I - Filminas de Clase - Facultad Cs. Exactas - UNCPBA - 2012

## Diagramas Sintácticos

- Al igual que en la notación BNF se debe indicar cual es el símbolo distinguido o axioma, a partir del cual comienza la derivación.

- Para reconocer una cadena del lenguaje, se parte del punto de entrada del diagrama correspondiente, y siguiendo algún camino en dicho diagrama, se llega al punto de salida.

Si el camino no existe, la cadena se rechaza.




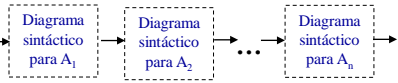
Cuando en el camino se encuentra un no terminal, el “flujo” continúa recursivamente a través del diagrama sintáctico correspondiente a ese no terminal (la aparición de un no terminal en un diagrama “explota” en otro diagrama que lo define).

- Cualquier camino entre el punto de entrada y el de salida representa una cadena generada por la regla que describe el diagrama.

Ciencias de la Computación I - Filminas de Clase – Facultad Cs. Exactas – UNCPBA - 2012

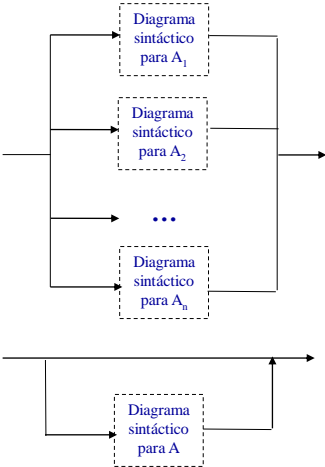
## Diagramas Sintácticos

Las reglas para la construcción de diagramas sintácticos a partir de una definición BNF son las siguientes:

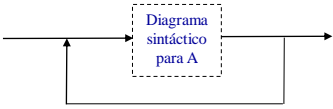
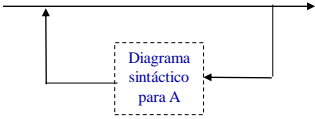
BNF	Diagrama sintáctico
<no terminal>	
terminal	
$\epsilon$	
Secuencia $A_1 A_2 \dots A_n$	

Ciencias de la Computación I - Filminas de Clase – Facultad Cs. Exactas – UNCPBA - 2012

# Diagramas Sintácticos

BNF	Diagrama sintáctico
Alternativa $A_1 \mid A_2 \mid \dots \mid A_n$  $[A] \quad (\varepsilon \mid A)$	

# Diagramas Sintácticos

BNF	Diagrama sintáctico
Repetición $A^+$ (1 ó más veces)	
$A^*$ (0 ó más veces)	



## Diagramas Sintácticos

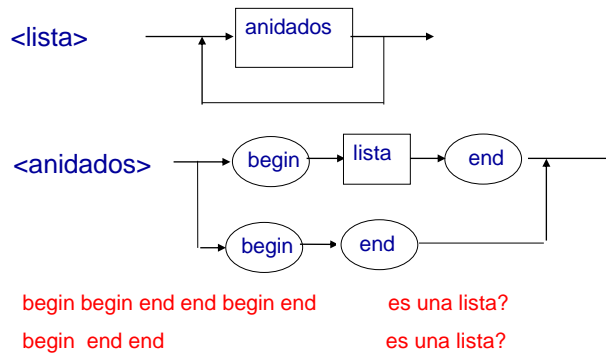
Ejemplos:

BNF Extendido para  $L_1$

$\langle \text{lista} \rangle ::= \langle \text{anidados} \rangle \langle \text{anidados} \rangle^*$

$\langle \text{anidados} \rangle ::= \text{begin} \langle \text{lista} \rangle \text{end} \mid \text{begin} \text{end}$

Diagrama sintáctico para  $L_1$



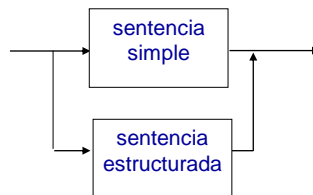
Ciencias de la Computación I - Filminas de Clase - Facultad Cs. Exactas - UNCPBA - 2012

## Diagramas Sintácticos

Ejemplos:

(De FreePascal, compilador usado en Introducción a la Programación I y II, [www.freepascal.org](http://www.freepascal.org))

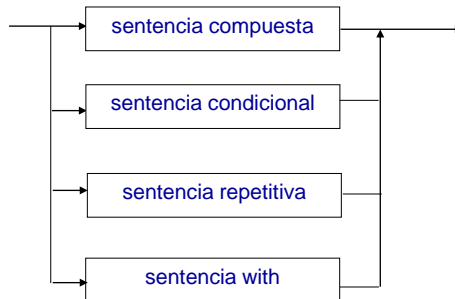
1)  $\langle \text{sentencia} \rangle ::= \langle \text{sentencia simple} \rangle \mid \langle \text{sentencia estructurada} \rangle$



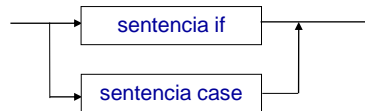
Ciencias de la Computación I - Filminas de Clase - Facultad Cs. Exactas - UNCPBA - 2012

## Diagramas Sintácticos

2)  $\langle \text{sentencia estructurada} \rangle ::= \langle \text{sentencia compuesta} \rangle \mid \langle \text{sentencia condicional} \rangle \mid \langle \text{sentencia repetitiva} \rangle \mid \langle \text{sentencia with} \rangle$



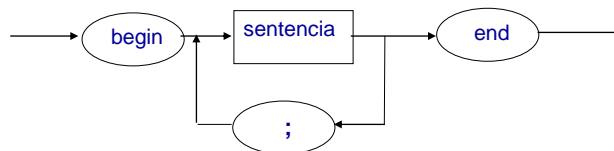
3)  $\langle \text{sentencia condicional} \rangle ::= \langle \text{sentencia if} \rangle \mid \langle \text{sentencia case} \rangle$



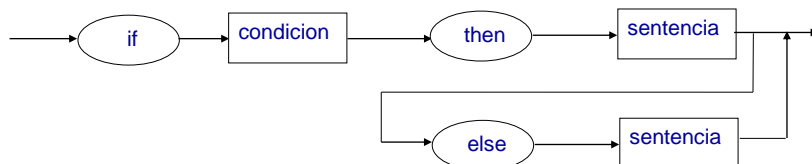
Ciencias de la Computación I - Filminas de Clase - Facultad Cs. Exactas - UNCPBA - 2012

## Diagramas Sintácticos

4)  $\langle \text{sentencia compuesta} \rangle ::= \text{begin } \langle \text{sentencia} \rangle \text{ ( } \langle \text{sentencia} \rangle \text{ )}^* \text{ end}$



5)  $\langle \text{sentencia if} \rangle ::= \text{if } \langle \text{condicion} \rangle \text{ then } \langle \text{sentencia} \rangle \text{ [else } \langle \text{sentencia} \rangle \text{]}$



Ciencias de la Computación I - Filminas de Clase - Facultad Cs. Exactas - UNCPBA - 2012