



Instituto Politécnico Nacional

Escuela Superior de Cómputo



Teoría computacional

Definiciones regulares

Prof. Rafael Aguilar



Contenido

- Expresiones regulares
- Definiciones regulares
- Ejemplos
- Ejercicios 01

Expresiones regulares

- Una **expresión regular** es una forma abreviada de representar cadenas de caracteres que se ajustan a un determinado patrón. Al **conjunto de cadenas representado por la expresión r** se le llama *lenguaje generado por la expresión regular r* y se escribe $L(r)$.
- Una expresión regular se define sobre un alfabeto Σ y **es una cadena formada por caracteres** de dicho alfabeto y por una serie de **operadores** también llamados **metacaracteres**.

Definiciones regulares

- Por conveniencia de notación se le da un nombre a algunas expresiones regulares, para así hacer más fácil su utilización en expresiones subsiguientes.

$$d_1 \rightarrow r_1$$

$$d_2 \rightarrow r_2$$

...

$$d_n \rightarrow r_n$$

- En donde:

- I. d_i es un nuevo símbolo, que no está en Σ y no es el mismo que cualquier otro d .
- II. Cada r_i es una expresión regular sobre el alfabeto $\Sigma \cup \{d_1, d_2, \dots, d_n\}$

Definiciones regulares

(Ejemplo 1)

- Los identificadores del lenguaje de programación C son cadenas de letras, dígitos y guiones bajos.

$letra_ \rightarrow A \mid B \mid \dots \mid Z \mid a \mid b \mid \dots \mid z \mid _$

$dígito \rightarrow 0 \mid 1 \mid \dots \mid 9$

$id \rightarrow letra_ (letra_ \mid dígito)^*$

Definiciones regulares

(Ejemplo 2)

- Los números sin signo (enteros o de punto flotante) son cadenas como 5280, 0.01234, 6.336E4 o 1.89E-4.

dígito $\rightarrow 0 \mid 1 \mid \dots \mid 9$

dígitos $\rightarrow \textit{dígito dígitos}^*$

fracción_opcional $\rightarrow \textit{.dígitos} \mid \lambda$

exponente_opcional $\rightarrow (E (+ \mid - \mid \lambda) \textit{dígitos}) \mid \lambda$

numero $\rightarrow \langle \textit{dígitos} \rangle \langle \textit{fracción_opcional} \rangle \langle \textit{exponente_opcional} \rangle$

Extensiones de las expresiones regulares

1. **Una o más instancias.** El operador unario posfijo $^+$ representa la cerradura positivo de una expresión regular y su lenguaje.
2. **Cero o una instancia.** El operador unario posfijo $?$ Significa “*cero o una ocurrencia*”.
3. **Clases de caracteres.** Una expresión regular $a_1/a_2|\dots|a_n$ en donde las a_i son cada una símbolos del alfabeto, puede sustituirse mediante la abreviación $[a_1a_2\dots a_n]$.

Definiciones regulares

(Ejemplo 1 Extensión)

- Los identificadores de C son cadenas de letras, dígitos y guiones bajos.

letra_ \rightarrow [A-Za-z_]

dígito \rightarrow [0-9]

id \rightarrow *letra_* (*letra_* | *dígito*)*

Definiciones regulares

(Ejemplo 2 extensión)

- Los números sin signo (enteros o de punto flotante) son cadenas como 5280, 0.01234, 6.336E4 o 1.89E-4.

dígito → [0-9]

dígitos → *dígito* *dígito* *

numero → *digitos* (*. digitos*)? (*E[+-]* ? *digitos*)?

Ejercicios 01 "Expresiones y definiciones regulares"

- *Sea* $\Sigma = \{a, b, \dots, c, z\}$.
- 1. **Describa** los lenguajes denotados por las siguientes expresiones regulares y **de al menos 10 ejemplos** de las palabras pertenecientes al lenguaje regular generado.
 - a) $a(a|b)^*a$
 - b) $e(ab)^*(ef)^+$
 - c) $((\lambda/a)b^*)^*$
 - d) $(a|b)^*a(a|b)(a|b)$
 - e) $a^*ba^*ba^*ba^+$
 - f) $(aa|bb)^*((ab|ba)(aa|bb)^*(ab|ba)(aa|bb)^*)^*$

Ejercicios 01 "Expresiones y definiciones regulares"

- **Sea** $\Sigma = \{\lambda, a, b, c, d, \dots, z, 0, 1, 2, \dots, 9, @, \$,), \%, \&, \#\}$.
- 2. **Liste 10 cadenas** de los lenguajes regulares generados por las siguientes expresiones regulares.
 - a) abc^+z^*
 - b) $((\lambda/1)e)^*$
 - c) $(1|2)^*z((e|u)(m|c))^+$
 - d) $(a^+bc^+z^{**})^+$
 - e) $(@^+\$%^+\&^{++})^+(\#^+@\"")^*$

Ejercicios 01 "Expresiones y definiciones regulares"

3. **Escriba expresiones regulares** que se apoyen de **definiciones regulares** para los siguientes lenguajes:
- a) Todas las cadenas de letras en minúsculas (a-z) que contengan las cinco vocales en orden. *Pg. "rtaeioujutf, artehjyibgfohgfdujhfd, aejhjhigghgougha"*
 - b) Comentarios largos que consistan de una cadena rodeada por /* y */ , sin un */ entre ellos, a menos que se encierre entre dobles comillas(").
 - c) Todas las cadenas de dígitos que tengan por lo menos un dígito repetido. *P.g. "12345672, 1286182, 1821"*
 - d) Los nombres que pueden darse a una función o variable en Java o C
 - e) Todos los posibles correos electrónicos (*revisar el RFC-822*)
 - f) Los números complejos (*Dar un formato de diseño propio y la expresión regular para validar*)
 - g) *RFC y CURP*