5) EXPRESSÕES REGULARES

Expressões regulares

→ São descrições algébricas de linguagens que usam as operações regulares (união, concatenação e estrela).

→ O valor de uma expressão regular é uma linguagem.
↳ Exemplo: $(\varepsilon \cup 1)0^*$ ⇒ linguagem das cadeias que podem ou não começar com um 1, seguido por zero ou mais 0s.
$= \{w \in \{0,1\}^* : \; w = 0^k \text{ ou } w = 10^k, \text{ com } k \geq 0\} = 0^* \cup 10^*$

→ Servem de entrada para muitos sistemas que processam strings (grep, lex, (bons) editores de texto). (cuidado com expressões regulares em outros contextos)

Expressões regulares

DEFINIÇÃO: R é uma expressão regular sobre um alfabeto Σ que descreve a linguagem $L(R)$ se

1) $R = x$ para algum $x \in \Sigma$. Nesse caso, $L(R) = \{x\}$
2) $R = \varepsilon$. Nesse caso, $L(R) = \{\varepsilon\}$.
3) $R = \emptyset$. Nesse caso, $L(R) = \emptyset$.

E sendo R_1 e R_2 duas expressões regulares,

4) $R = R_1 \cup R_2$. Nesse caso, $L(R) = L(R_1) \cup L(R_2)$.
5) $R = R_1 R_2$. Nesse caso, $L(R) = L(R_1)L(R_2)$
6) $R = R_1^*$. Nesse caso, $L(R) = L(R_1)^*$

É possível agrupar operadores usando parênteses (se $R = (R')$, então $L(R) = L(R')$).

→ Exemplo: $R = (\varepsilon \cup 1)0^* = R_1 R_2$
- $R_1 = (\varepsilon \cup 1) = R_3 \cup R_4$
  - $R_3 = \varepsilon$
  - $R_4 = 1$
- $R_2 = 0^* = R_5^*$
  - $R_5 = 0$

Precedência dos operadores

→ Têm maior precedência, a menos de parênteses:
   estrela
   concatenação
   união

→ Exemplos: $01 \cup 20^*$
   $0(1 \cup 20^*)$
   $0(1 \cup 2)0^*$

Equivalência de Expressões Regulares

→ Duas expressões regulares R_1 e R_2 são equivalentes, denotado por $R_1 \equiv R_2$, se $L(R_1) \equiv L(R_2)$

→ Exemplos: $(0 \cup \varepsilon)1 \equiv 01 \cup 1$
   $(0 \cup 1)^* \equiv (0^*1^*)^*$

Abuso de notação

→ Às vezes vamos denotar uma ER como sendo a linguagem descrita por ela:

$$0^*10^* = \{w \in \{0,1\}^* : w \text{ contém um único } 1\}$$

onde o correto seria

$$L(0^*10^*) = \{w \in \{0,1\}^* : w \text{ contém um único } 1\}$$

Açúcar sintático

→ $R^+ \equiv RR^*$

→ $\Sigma \equiv x_1 \cup x_2 \cup \cdots \cup x_k$, onde $x_i \in \Sigma$

Propriedades das expressões regulares

1) $A \cup B = B \cup A$ (comut.)
2) $A \cup (B \cup C) = (A \cup B) \cup C$ (assoc.)
3) $A(BC) = (AB)C$ (assoc.)
4) $A(B \cup C) = AB \cup AC$ (distr.)
5) $(A \cup B)C = AC \cup BC$ (distr.)
6) $\emptyset A = A\emptyset = \emptyset$ (aniq.)
7) $\varepsilon A = A\varepsilon = A$ (ident.)
8) $A \cup \emptyset = \emptyset \cup A = A$ (ident.)
9) $A \cup A = A$ (idemp.)
10) $(A^*)^* = A^*$ (fecham.)
11) $\emptyset^* = \varepsilon$ (fecham.)
12) $\varepsilon^* = \varepsilon$ (fecham.)

Exemplos

Qual a linguagem descrita por $\Sigma^* 001 \Sigma^*$?

$= \{w \in \{0,1\}^* :$ w contém 001 como subcadeia$\}$

Qual a linguagem descrita por $1^*(01^+)^*$?

$= \{w \in \{0,1\}^* :$ todo 0 de w é seguido de pelo menos um 1$\}$

ou : w não contém 00 como subcadeia

Qual a linguagem descrita por $01 \cup 10$?

$= \{01, 10\}$

Qual a linguagem descrita por $0\Sigma^*0 \cup 1\Sigma^*1 \cup 0 \cup 1$?

$= \{0, 1, 0010, 1110, 11, 10001, \dots\}$

$= \{w \in \{0,1\}^* :$ w começa e termina com o mesmo símbolo$\}$

Exemplos

Forneça uma expressão regular que descreva as seguintes linguagens

1) $\{w \in \{a,b,c\}^* : |w| \text{ é par}\} = ((a \cup b \cup c)(a \cup b \cup c))^*$

2) $\{w \in \{0,1\}^* :$ todo 0 em w é seguido de um único $1\}$
$= 1^*(01)^*$

3) $\emptyset = \emptyset = \emptyset 1 = \Sigma^* \emptyset$

4) $\{\varepsilon\} = \varepsilon = \emptyset^* = \varepsilon^*$

5) $\{w \in \{0,1\}^* :$ toda posição ímpar de w é $1\} = 1(\Sigma 1)^* \cup (1\Sigma)^*$

6) $\{w \in \{a,b\}^* : |w|_a \geq 1 \text{ e } |w|_b \geq 1\}$
$= \Sigma^* a \Sigma^* b \Sigma^* \cup \Sigma^* b \Sigma^* a \Sigma^*$

7) $\{w \in \{0,1\}^* :$ w tem no máximo um par de 1s consecutivos$\}$
$= (0 \cup 10^+)^* (11 \cup \varepsilon)(0^+1)^* 0^*$

8) $\{w \in \{0,1\}^* : |w|_0 \text{ é divisível por } 3\}$
$= (1^* 0 1^* 0 1^* 0 1^*)^*$

9) $\{w \in \{0,1\}^* :$ w não contém 110 como subcadeia$\}$
$= (0 \cup 10)^* 1^*$

Linguagens e expressões regulares

LEMA: A linguagem descrita por uma expressão regular é regular.

DEMONSTRAÇÃO: Seja R uma expressão regular qualquer.
Vamos provar por indução na quantidade de operações presentes em R que $L(R)$ é regular.

BASE: R não tem operações. Então temos três casos:

1) $R = \mathbf{x}$, com $x \in \Sigma$. Nesse caso, $L(R) = \{x\}$ e $\rightarrow \bigcirc \xrightarrow{x} \circledcirc$ é um AFN que reconhece $L(R)$.
2) $R = \boldsymbol{\varepsilon}$. Nesse caso, $L(R) = \{\varepsilon\}$ e $\rightarrow \circledcirc$ a reconhece.
3) $R = \boldsymbol{\emptyset}$. Nesse caso, $L(R) = \emptyset$ e $\rightarrow \bigcirc$ a reconhece.

Logo, o caso base vale.

Agora suponha que R tem $n \geq 1$ operações e que para qualquer expressão regular R' com $< n$ operações, vale que $L(R')$ é regular.

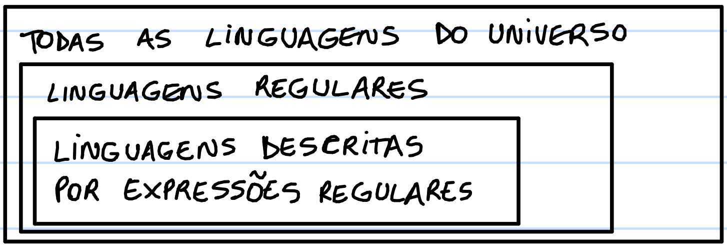
Como R tem ao menos uma operação, temos três casos:

1) $R = R_1 \cup R_2$. Como R_1 e R_2 têm menos operações, vale, por HI, que $L(R_1)$ e $L(R_2)$ são regulares. Como já visto em aula, a união de duas linguagens regulares é regular $\therefore$ $L(R_1) \cup L(R_2)$ é regular. Como $L(R) = L(R_1) \cup L(R_2)$, o resultado segue.
2) $R = R_1 R_2$. Por HI, $L(R_1)$ e $L(R_2)$ são regulares. Como $L(R) = L(R_1) L(R_2)$, o resultado segue porque ling. regulares são fechados sob concat.
3) $R = R_1^*$. Por HI, $L(R_1)$ é regular. Por fechamento, $L(R) = L(R_1)^*$ é regular.

CQD

Entendendo o resultado

LEMA: A linguagem descrita por uma expressão regular é regular.



→ A demonstração do lema em conjunto com as demonstrações dos teoremas sobre as propriedades de fechamento fornece um algoritmo para transformar expressões regulares em AFN.

Transformando ERs em AFNs

$R = (a \cup b)^* aba = R_1 \cdot R_2$

$\cancel{R_1 = (a \cup b)^* = R_3^*}$

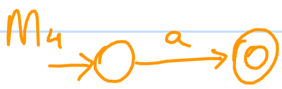
$\cancel{R_2 = aba = R_4 R_5}$

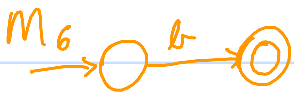
$\cancel{R_3 = a \cup b = R_4 \cup R_6}$

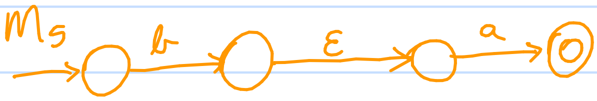
$\cancel{R_4 = a}$

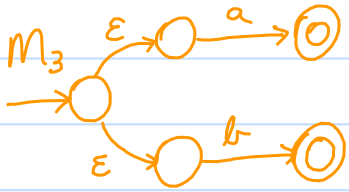
$\cancel{R_5 = ba} = R_6 R_4$

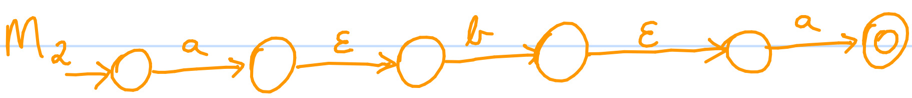
$\cancel{R_6 = b}$

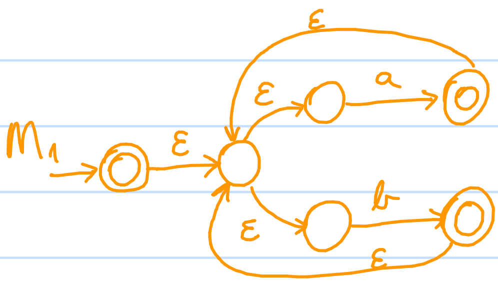


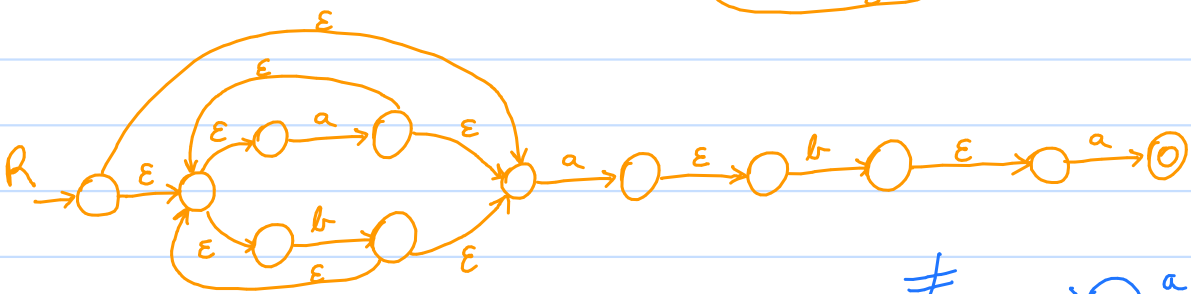




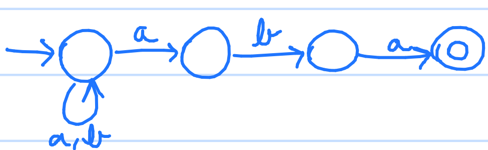








$\neq$



LEMA: Toda linguagem regular pode ser descrita por uma expressão regular

Ideia: Ling regular ⇒ AFD/AFN M

Vamos transformar M em um AFNG equivalente e então modificá-lo para que fique apenas com dois estados, o que significa que o rótulo da transição restante é a expressão regular equivalente.

COROLÁRIO: Uma linguagem é regular se e somente se alguma expressão regular a descreve.

DEMONSTRAÇÃO: Seja A uma linguagem.

($\Rightarrow$) Suponha que A é regular.

Então pelo lema anterior existe uma expressão regular que a descreve.

($\Leftarrow$) Suponha que uma ER R descreve A.

Então pelo primeiro lema, A é regular.

CQD