



LINGUAGENS FORMAIS E AUTÓMATOS / COMPILADORES

LINGUAGENS REGULARES, EXPRESSÕES REGULARES E
GRAMÁTICAS REGULARES

Artur Pereira <artur@ua.pt>,
Miguel Oliveira e Silva <mos@ua.pt>

DETI, Universidade de Aveiro

SUMÁRIO

- 1 LINGUAGENS REGULARES
- 2 EXPRESSÕES REGULARES
- 3 GRAMÁTICAS REGULARES
- 4 EQUIVALÊNCIA ENTRE EXPRESSÕES REGULARES E GRAMÁTICAS REGULARES

DEFINIÇÃO DE LINGUAGEM REGULAR

A classe das **linguagens regulares** sobre o alfabeto A define-se indutivamente da seguinte forma:

- 1 O conjunto vazio, \emptyset , é uma linguagem regular (LR).
- 2 Qualquer que seja o $a \in A$, o conjunto $\{a\}$ é uma LR.
- 3 Se L_1 e L_2 são linguagens regulares, então $L_1 \cup L_2$ é uma LR.
- 4 Se L_1 e L_2 são linguagens regulares, então $L_1.L_2$ é uma LR.
- 5 Se L_1 é uma linguagem regular, então $(L_1)^*$ é uma LR.
- 6 Nada mais é linguagem regular.

-
- Note que $\{\varepsilon\}$ é uma LR, uma vez que $\{\varepsilon\} = \emptyset^*$.
 - Q Qualquer linguagem finita é uma LR. *Mostre-o com base nesta definição*
 - Q *Com base nesta definição, mostre que o conjunto dos números binários começados em 1 e terminados em 0 é uma LR sobre o alfabeto $A = \{0, 1\}$*

DEFINIÇÃO DE EXPRESSÃO REGULAR

O conjunto das **expressões regulares** sobre o alfabeto A define-se indutivamente da seguinte forma:

- ① $()$ é uma expressão regular (ER) que representa a LR $\{\}$.
 - ② Qualquer que seja o $a \in A$, a é uma ER que representa a LR $\{a\}$.
 - ③ Se e_1 e e_2 são ER representando respectivamente as LR L_1 e L_2 , então $(e_1|e_2)$ é uma ER representando a LR $L_1 \cup L_2$.
 - ④ Se e_1 e e_2 são ER representando respectivamente as LR L_1 e L_2 , então $(e_1 e_2)$ é uma ER representando a LR $L_1.L_2$.
 - ⑤ Se e_1 é uma ER representando a LR L_1 , então e_1^* é uma ER representando a LR $(L_1)^*$.
 - ⑥ Nada mais é expressão regular.
-
- É habitual representar-se por ε a ER $()^*$. Representa a linguagem $\{\varepsilon\}$.

PROJETO DE UMA EXPRESSÃO REGULAR

Q Determine uma ER que representa o conjunto dos números binários começados em 1 e terminados em 0.

R $1(0|1)^*0$

Q Determine uma ER que represente as sequências definidas sobre o alfabeto $A = \{a, b, c\}$ que satisfazem o requisito de qualquer b ter um a imediatamente à sua esquerda e um c imediatamente à sua direita.

R $(a|abc|c)^*$

Q Determine uma ER que represente as sequências binárias com um número par de zeros.

R $1^*(01^*01^*)^*$

PROPRIEDADES DAS EXPRESSÕES REGULARES (1)

OPERAÇÃO DE ESCOLHA — |

- comutativa: $e_1 | e_2 = e_2 | e_1$
- associativa: $e_1 | (e_2 | e_3) = (e_1 | e_2) | e_3 = e_1 | e_2 | e_3$
- existência de elemento neutro: $e_1 | () = () | e_1 = e_1$
- idempotência: $e_1 | e_1 = e_1$

OPERAÇÃO DE CONCATENAÇÃO — .

- associativa: $e_1(e_2e_3) = (e_1e_2)e_3 = e_1e_2e_3$
- existência de elemento neutro: $e_1\varepsilon = \varepsilon e_1 = e_1$
- existência de elemento absorvente: $e_1() = ()e_1 = ()$
- não goza da propriedade comutativa

PROPRIEDADES DAS EXPRESSÕES REGULARES (2)

OPERAÇÕES DE ESCOLHA E CONCATENAÇÃO

- distributiva à esquerda da concatenação em relação à escolha:
$$e_1(e_2 \mid e_3) = e_1 e_2 \mid e_1 e_3$$
- distributiva à direita da concatenação em relação à escolha:
$$(e_1 \mid e_2)e_3 = e_1 e_3 \mid e_2 e_3$$

OPERAÇÃO DE FECHO

- $(e^*)^* = e^*$
- $(e_1^* \mid e_2^*)^* = (e_1 \mid e_2)^*$
- $(e_1 \mid e_2^*)^* = (e_1 \mid e_2)^*$
- $(e_1 \mid e_2)^* \neq e_1^* \mid e_2^*$
- $(e_1 e_2)^* \neq e_1^* e_2^*$

SIMPLIFICAÇÃO NOTACIONAL

- Na escrita de expressões regulares assume-se que a operação de fecho (*) tem precedência em relação à operação de concatenação e que esta tem precedência em relação à operação de escolha (|).
- O uso destas precedências em conjunto com as propriedades associativas da concatenação e da escolha permite a queda de alguns parêntesis e consequentemente uma notação simplificada.

Exemplo:

$$e_1 | e_2 \cdot e_3^* = e_1 | (e_2 \cdot (e_3^*))$$

SIMPLIFICAÇÃO NOTACIONAL

Q Determine uma ER que representa o conjunto dos números binários começados em 1 e terminados em 0.

R $1(0|1)^*0 = (1((0|1)^*))0$

Q Determine uma ER que represente as sequências definidas sobre o alfabeto $A = \{a, b, c\}$ que satisfazem o requisito de qualquer b ter um a imediatamente à sua esquerda e um c imediatamente à sua direita.

R $(a|abc|c)^* = ((a|((ab)c))|c)^*$

Q Determine uma ER que represente as sequências binárias com um número par de zeros.

R $1^*(01^*01^*)^* = (1^*)(((((0(1^*))0)(1^*)))^*)$

SIMPLIFICAÇÃO NOTACIONAL

Q Sobre o alfabeto $A = \{0, 1\}$ construa uma expressão regular que reconheça a linguagem

$$L = \{\omega \in A^* : \#(0, \omega) = 2\}$$

R $1^*01^*01^*$

Q Sobre o alfabeto $A = \{a, b, \dots, z\}$ construa uma expressão regular que reconheça a linguagem

$$L = \{\omega \in A^* : \#(a, \omega) = 3\}$$

R $(b|c|\dots|z)^*a(b|c|\dots|z)^*a(b|c|\dots|z)^*a(b|c|\dots|z)^*$

EXTENSÕES NOTACIONAIS (1)

- uma ou mais ocorrências:

$$e^+ = e.e^*$$

- uma ou nenhuma ocorrência:

$$e? = (e|\varepsilon)$$

- um símbolo do sub-alfabeto dado:

$$[a_1 a_2 a_3 \cdots a_n] = (a_1 \mid a_2 \mid a_3 \mid \cdots \mid a_n)$$

- um símbolo do sub-alfabeto dado:

$$[a_1 - a_n] = (a_1 \mid \cdots \mid a_n)$$

- um símbolo do alfabeto fora do conjunto dado:

$$[\hat{a}_1 a_2 a_3 \cdots a_n]$$

- um símbolo do alfabeto fora do conjunto dado:

$$[\hat{a}_1 - a_n]$$

EXTENSÕES NOTACIONAIS (2)

- n ocorrências de:

$$e\{n\} = \underbrace{e.e.\cdots.e}_n$$

- de n_1 a n_2 ocorrências:

$$e\{n_1, n_2\} = \underbrace{e.e.\cdots.e}_{n_1, n_2}$$

- n ou mais ocorrências:

$$e\{n, \} = \underbrace{e.e.\cdots.e}_{n,}$$

EXTENSÕES NOTACIONAL

Q Sobre o alfabeto $A = \{0, 1\}$ construa uma expressão regular que reconheça a linguagem

$$L = \{\omega \in A^* : \#(0, \omega) = 2\}$$

R $1^*01^*01^* = (1^*0)\{2\}1^*$

Q Sobre o alfabeto $A = \{a, b, \dots, z\}$ construa uma expressão regular que reconheça a linguagem

$$L = \{\omega \in A^* : \#(a, \omega) = 3\}$$

R $(b|c|\dots|z)^*a(b|c|\dots|z)^*a(b|c|\dots|z)^*a(b|c|\dots|z)^* = ([b-z]^*a)\{3\}[b-z]^*$

EXTENSÕES NOTACIONAIS (3)

EXPRESSÕES REGULARES ESPECIAIS FREQUENTES

- . um símbolo qualquer diferente de $\backslash n$
- ^ palavra vazia no início de linha
- \$ palavra vazia no fim de linha
- $\backslash <$ palavra vazia no início de palavra
- $\backslash >$ palavra vazia no fim de palavra

GRAMÁTICAS REGULARES

DEFINIÇÃO DE GRAMÁTICA

Uma gramática é um quádruplo $G = (T, N, P, S)$, onde

- T é um conjunto finito não vazio de símbolos terminais;
- N , sendo $N \cap T = \emptyset$, é um conjunto finito não vazio de símbolos não terminais;
- P é um conjunto de produções (ou regras de rescrita), cada uma da forma $\alpha \rightarrow \beta$;
- $S \in N$ é o símbolo inicial.
- $\alpha \in (N \cup T)^* N (N \cup T)^*$
- $\beta \in (N \cup T)^*$
- As restrições a α e β definem uma taxonomia das linguagens (gramáticas)

GRAMÁTICAS REGULARES

- \mathcal{D} Uma gramática $G = (T, N, P, S)$ diz-se **regular** se, para qualquer produção $(\alpha \rightarrow \beta) \in P$, as duas condições seguintes são satisfeitas

$$\alpha \in N$$

$$\beta \in T^* \cup T^*N$$

- A linguagem gerada por uma gramática regular é regular
 - Logo, é possível converter uma gramática regular numa expressão regular que represente a mesma linguagem e vice-versa
- As gramáticas regulares são fechadas sob as operações de reunião, concatenação, fecho, intersecção e complementação.

OPERAÇÕES SOBRE GRAMÁTICAS REGULARES

REUNIÃO DE GRAMÁTICAS REGULARES

\mathcal{D} Sejam $G_1 = (T_1, N_1, P_1, S_1)$ e $G_2 = (T_2, N_2, P_2, S_2)$ duas gramáticas regulares quaisquer, com $N_1 \cap N_2 = \emptyset$. A gramática $G = (T, N, P, S)$ onde

$$T = T_1 \cup T_2$$

$$N = N_1 \cup N_2 \cup \{S\} \quad \text{com} \quad S \notin (N_1 \cup N_2)$$

$$P = \{S \rightarrow S_1, S \rightarrow S_2\} \cup P_1 \cup P_2$$

é regular e gera a linguagem $L = L(G_1) \cup L(G_2)$.

- Para $i = 1, 2$, a nova produção $S \rightarrow S_i$ permite que G gere a linguagem $L(G_i)$

OPERAÇÕES SOBRE GRAMÁTICAS REGULARES

\mathcal{Q} Sobre o conjunto de terminais $T = \{a, b, c\}$, determine uma gramática regular que represente a linguagem

$$L = L_1 \cup L_2$$

sabendo que

$$L_1 = \{a\omega : \omega \in T^*\}$$

$$L_2 = \{\omega a : \omega \in T^*\}$$

Comece por obter as gramáticas regulares que representam L_1 e L_2 .

\mathcal{R}

$$S_1 \rightarrow a X_1$$

$$X_1 \rightarrow a X_1$$

$$X_1 \rightarrow b X_1$$

$$X_1 \rightarrow c X_1$$

$$X_1 \rightarrow \varepsilon$$

$$S_2 \rightarrow a S_2$$

$$S_2 \rightarrow b S_2$$

$$S_2 \rightarrow c S_2$$

$$S_2 \rightarrow a$$

$$S \rightarrow S_1 \mid S_2$$

$$S_1 \rightarrow a X_1$$

$$X_1 \rightarrow a X_1 \mid b X_1 \mid c X_1$$

$$X_1 \rightarrow \varepsilon$$

$$S_2 \rightarrow a S_2 \mid b S_2 \mid c S_2$$

$$S_2 \rightarrow a$$

OPERAÇÕES SOBRE GRAMÁTICAS REGULARES

CONCATENAÇÃO DE GRAMÁTICAS REGULARES

\mathcal{D} Sejam $G_1 = (T_1, N_1, P_1, S_1)$ e $G_2 = (T_2, N_2, P_2, S_2)$ duas gramáticas regulares quaisquer, com $N_1 \cap N_2 = \emptyset$. A gramática $G = (T, N, P, S)$ onde

$$T = T_1 \cup T_2$$

$$N = N_1 \cup N_2$$

$$P = \{A \rightarrow \omega S_2 : (A \rightarrow \omega) \in P_1 \wedge \omega \in T_1^*\} \\ \cup \{A \rightarrow \omega : (A \rightarrow \omega) \in P_1 \wedge \omega \in T_1^* N_1\} \\ \cup P_2$$

$$S = S_1$$

é regular e gera a linguagem $L = L(G_1) \cdot L(G_2)$.

- As produções da primeira gramática do tipo $\beta \in T^*$ ganham o símbolo inicial da segunda gramática no fim
- As produções da primeira gramática do tipo $\beta \in T^* N$ mantêm-se inalteradas
- As produções da segunda gramática mantêm-se inalteradas

OPERAÇÕES SOBRE GRAMÁTICAS REGULARES

Q Sobre o conjunto de terminais $T = \{a, b, c\}$, determine uma gramática regular que represente a linguagem

$$L = L_1 \cdot L_2$$

sabendo que

$$L_1 = \{a\omega : \omega \in T^*\}$$

$$L_2 = \{\omega a : \omega \in T^*\}$$

R

$$S_1 \rightarrow a X_1$$

$$X_1 \rightarrow a X_1$$

$$X_1 \rightarrow b X_1$$

$$X_1 \rightarrow c X_1$$

$$X_1 \rightarrow \varepsilon$$

$$S_2 \rightarrow a S_2$$

$$S_2 \rightarrow b S_2$$

$$S_2 \rightarrow c S_2$$

$$S_2 \rightarrow a$$

$$S_1 \rightarrow a X_1$$

$$X_1 \rightarrow a X_1 \mid b X_1 \mid c X_1$$

$$X_1 \rightarrow S_2$$

$$S_2 \rightarrow a S_2 \mid b S_2 \mid c S_2$$

$$S_2 \rightarrow a$$

OPERAÇÕES SOBRE GRAMÁTICAS REGULARES

FECHO DE KLEENE DE GRAMÁTICAS REGULARES

\mathcal{D} Seja $G_1 = (T_1, N_1, P_1, S_1)$ uma gramática regular qualquer. A gramática $G = (T, N, P, S)$ onde

$$T = T_1$$

$$N = N_1 \cup \{S\} \quad \text{com} \quad S \notin N_1$$

$$P = \{S \rightarrow \varepsilon, S \rightarrow S_1\} \\ \cup \{A \rightarrow \omega S : (A \rightarrow \omega) \in P_1 \wedge \omega \in T_1^*\} \\ \cup \{A \rightarrow \omega : (A \rightarrow \omega) \in P_1 \wedge \omega \in T_1^* N_1\}$$

é regular e gera a linguagem $L = (L(G_1))^*$.

- As produções que terminam num não terminal mantêm-se inalteradas
- As produções que só têm terminais ganham o símbolo inicial no fim
- As novas produções $S \rightarrow \varepsilon$ e $S \rightarrow S_1$ garante que $(L(G_1))^n \subseteq L(G)$, para qualquer $n \geq 0$

OPERAÇÕES SOBRE GRAMÁTICAS REGULARES

- Q Sobre o conjunto de terminais $T = \{a, b, c\}$, determine uma gramática regular que represente a linguagem

$$L = L_1^*$$

sabendo que

$$L_1 = \{a\omega : \omega \in T^*\}$$

Comece por obter a gramática regular que representa L_1 .

R

$$S_1 \rightarrow a X_1$$

$$X_1 \rightarrow a X_1$$

$$X_1 \rightarrow b X_1$$

$$X_1 \rightarrow c X_1$$

$$X_1 \rightarrow \varepsilon$$

$$S \rightarrow \varepsilon \mid S_1$$

$$S_1 \rightarrow a X_1$$

$$X_1 \rightarrow a X_1 \mid b X_1 \mid c X_1$$

$$X_1 \rightarrow S$$

OPERAÇÕES SOBRE GRAMÁTICAS REGULARES

INTERSECÇÃO DE GRAMÁTICAS REGULARES

\mathcal{D} Sejam $G_1 = (T_1, N_1, P_1, S_1)$ e $G_2 = (T_2, N_2, P_2, S_2)$ duas gramáticas regulares quaisquer, com $N_1 \cap N_2 = \emptyset$. A gramática $G = (T, N, P, S)$ onde

$$T = T_1 \cup T_2$$

$$N = N_1 \cup N_2$$

$$P = \dots$$

$$S = (S_1, S_2)$$

é regular e gera a linguagem $L = L(G_1) \cap L(G_2)$.

- $(N_1, N_2) \rightarrow a \ (N'_1, N'_2) \in P$ sse

$$N_1 \rightarrow a \ N'_1 \in P_1 \quad \wedge \quad N_2 \rightarrow a \ N'_2 \in P_2$$

- $(N_1, N_2) \rightarrow a \in P$ sse

$$N_1 \rightarrow a \in P_1 \quad \wedge \quad N_2 \xRightarrow{*} a \quad \vee$$

$$N_2 \rightarrow a \in P_2 \quad \wedge \quad N_1 \xRightarrow{*} a$$

OPERAÇÕES SOBRE GRAMÁTICAS REGULARES

Q Sobre o conjunto de terminais $T = \{a, b, c\}$, determine uma gramática regular que represente a linguagem

$$L = L_1 \cap L_2$$

sabendo que

$$L_1 = \{a\omega : \omega \in T^*\}$$

$$L_2 = \{\omega a : \omega \in T^*\}$$

R

$$S_1 \rightarrow a X_1$$

$$X_1 \rightarrow a X_1$$

$$X_1 \rightarrow b X_1$$

$$X_1 \rightarrow c X_1$$

$$X_1 \rightarrow \varepsilon$$

$$S_2 \rightarrow a S_2$$

$$S_2 \rightarrow b S_2$$

$$S_2 \rightarrow c S_2$$

$$S_2 \rightarrow a$$

$$S \rightarrow a X \mid a$$

$$X \rightarrow a X \mid b X \mid c X \mid a$$

where $S = (S_1, S_2)$

and $X = (X_1, S_2)$.

EQUIVALÊNCIA COM AS EXPRESSÕES REGULARES

CONVERSÃO DE UMA ER EM UMA GR

- Basta obter GR para as expressões regulares primitivas e aplicar as operações regulares sobre GR
- A GR para a ER ε é dada por

$$S \rightarrow \varepsilon$$

- A GR para a ER a , qualquer que seja o a , é dada por

$$S \rightarrow a$$

Q Obtenha uma GR equivalente à ER $e = (a|b|c)^*(bb|cc)(a|b|c)^*$

R

$$S_1 \rightarrow a S_1 \mid b S_1 \mid c S_1 \mid \varepsilon$$

$$S_1 \rightarrow a S_1 \mid b S_1 \mid c S_1 \mid S_2$$

$$S_2 \rightarrow b b \mid c c$$

$$S_2 \rightarrow b b S_3 \mid c c S_3$$

$$S_3 \rightarrow a S_3 \mid b S_3 \mid c S_3 \mid \varepsilon$$

$$S_3 \rightarrow a S_3 \mid b S_3 \mid c S_3 \mid \varepsilon$$

EQUIVALÊNCIA COM AS EXPRESSÕES REGULARES

CONVERSÃO DE UMA GR EM UMA ER

Seja $G = (T, N, P, S)$ uma gramática regular qualquer. Uma ER que represente a mesma linguagem que a gramática G pode ser obtida por um processo de transformação de equivalência.

ALGORITMO DE CONVERSÃO

- 1 Converte-se a gramática $G = (T, N, P, S)$ no conjunto de triplos seguinte:

$$\begin{aligned}\mathcal{E} &= \{(E, \varepsilon, S)\} \\ &\cup \{(A, \omega, B) : (A \rightarrow \omega B) \in P \wedge B \in N\} \\ &\cup \{(A, \omega, \varepsilon) : (A \rightarrow \omega) \in P \wedge \omega \in T^*\}\end{aligned}$$

com $E \notin N$.

- 2 Removem-se, por transformações de equivalência, um a um, todos os símbolos de N , até se obter um único triplo da forma (E, e, ε) . O valor de e é a expressão regular pretendida.

EQUIVALÊNCIA COM AS EXPRESSÕES REGULARES

REMOÇÃO DOS SÍMBOLOS DE N

- Para cada símbolo $B \in N$
 - Ⓐ Substituir todos os triplos da forma (A, β_i, B) por um único (A, ω_1, B) , onde $\omega_1 = \beta_1 \mid \beta_2 \mid \cdots \mid \beta_n$
 - Ⓑ Substituir todos os triplos da forma (B, α_i, B) por um único (B, ω_2, B) , onde $\omega_2 = \alpha_1 \mid \alpha_2 \mid \cdots \mid \alpha_m$
 - Ⓒ Substituir todos os triplos da forma (B, γ_i, C) por um único (B, ω_3, C) , onde $\omega_3 = \gamma_1 \mid \gamma_2 \mid \cdots \mid \gamma_k$
 - Ⓓ Substituir o triplo de triplos $((A, \omega_1, B), (B, \omega_2, B), (B, \omega_3, C))$ pelo triplo $(A, \omega_1 \omega_2^* \omega_3, C)$

EQUIVALÊNCIA COM AS EXPRESSÕES REGULARES

\mathcal{Q} Obtenha uma ER equivalente à gramática regular seguinte

$$S \rightarrow a S \mid b S \mid c S \mid aba X$$

$$X \rightarrow a X \mid b X \mid c X \mid \varepsilon$$

\mathcal{R}

$$\begin{aligned}\mathcal{E} &= \{(E, \varepsilon, S), (S, a, S), (S, b, S), (S, c, S), (S, aba, X), \\ &\quad (X, a, X), (X, b, X), (X, c, X), (X, \varepsilon, \varepsilon)\} \\ &= \{(E, \varepsilon, S), (S, a|b|c, S), (S, aba, X), \\ &\quad (X, a, X), (X, b, X), (X, c, X), (X, \varepsilon, \varepsilon)\} \\ &= \{(E, (a|b|c)^* aba, X), \\ &\quad (X, a, X), (X, b, X), (X, c, X), (X, \varepsilon, \varepsilon)\} \\ &= \{(E, (a|b|c)^* aba, X), (X, (a|b|c), X), (X, \varepsilon, \varepsilon)\} \\ &= \{(E, (a|b|c)^* aba(a|b|c)^*, \varepsilon)\}\end{aligned}$$