Aula 11 – Expressão Regular

Toda linguagem regular pode ser descrita por uma expressão denominada *Expressão Regular.* Trata-se de um formalismo denotacional, também considerado gerador, pois se pode inferir como construir ("gerar") as palavras de uma linguagem.

Uma expressão regular é definida a partir de conjuntos (símbolos das linguagens) básicos e operações de concatenação e de união. As expressões regulares são consideradas adequadas para a comunicação Humano x Humano e, principalmente, para a comunicação Humano x Máquina.

10.1 Expressão Regular, Linguagem Gerada

Uma Expressão Regular (frequentemente abreviada por ER) sobre um alfabeto Σ é indutivamente definida como segue:

a) Base de Indução:

a1) A expressão: $\emptyset$

é expressão regular e denota a linguagem vazia: $\emptyset$

a2) A expressão: ε

é uma expressão regular e denota a linguagem que contém exclusivamente a palavra vazia: $\{\varepsilon\}$

a3) Para qualquer símbolo $x \in \Sigma$, a expressão: x

é uma expressão regular e denota a linguagem que contém exclusivamente a palavra constituída pelo símbolo x: $\{x\}$

b) Passo de Indução: Se r e s são expressões regulares e denotam as linguagens R e S, respectivamente, então:

b1) *União*. A expressão: $(r + s)$

é uma expressão regular e denota a linguagem $R \cup S$

b2) *Concatenação*. A expressão: (rs)

é expressão regular e denota a linguagem: $RS = \{rs \mid r \in R \text{ e } s \in S\}$

b3) *Concatenação Sucessiva*. A expressão: (r^*)

é expressão regular e denota a linguagem: R^*

Se r é uma expressão regular, a correspondente linguagem denotada é dita a *Linguagem Gerada* por r, sendo representada por:

L(r) ou GERA(r)

A omissão de parênteses em uma ER é usual, respeitando as seguintes convenções:

- a concatenação sucessiva tem precedência sobre a concatenação e a união;
- a concatenação tem precedência sobre a união.

Exemplo: Na tabela abaixo são apresentadas expressões regulares e as correspondentes linguagens geradas.

Tabela 1 - Expressões Regulares e correspondentes Linguagens Geradas

Expressão Regular	Linguagem Gerada
aa	Somente a palavra aa
ba*	Todas as palavras que iniciam por b, seguindo por zero ou mais a's (b, ba, baa, ...)
(a + b)*	Todas as palavras sobre os símbolos {a,b}
(a + b)*aa(a + b)*	Todas as palavras contendo aa como subpalavra
a*ba*ba	Todas as palavras contendo exatamente dois b's
(a + b)*(aa + bb)	Todas as palavras que terminam com aa ou bb

Fonte: Dados gerados a partir das expressões regulares

Detalhando a linguagem gerada pela expressão (a + b)* (aa + bb), vale que:

Exercício:

1. Descreva no formalismo da representação (palavras) as linguagens geradas pelas seguintes expressões regulares (desenvolva todos os passos de indução):

   a) (aa + ba)*(aa + bb)*
   b) (b + ab)*(ε + a)
   c) (aa + bb + (aa + bb)(ab + ba)(aa + bb))*

Teorema 1: Expressão Regular → Linguagem Regular

Se r é uma expressão regular, então GERA(r) é uma linguagem regular.

Prova: (por indução):

Por definição, uma linguagem é regular se, e somente se, é possível construir um autômato finito (Determinístico, Não Determinístico ou Não Determinístico com movimentos vazios) que reconheça essa linguagem. Assim, é necessário mostrar que, dada uma expressão regular r qualquer, é possível construir um autômato finito M tal que:

ACEITA(M) = GERA(r)

Na construção do correspondente autômato finito com movimentos vazios M apresentado abaixo, a demonstração de que ACEITA(M) = GERA(r) é por indução no número de operadores.

a) ***Base de Indução****:* Seja r uma expressão regular com zero operador. Então r só pode ser da forma:

$r = \emptyset$

$r = \varepsilon$

$r = x$ (x pertence a Σ)

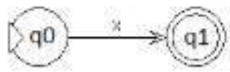
Os Autômatos Finitos:

$M1 = \{\emptyset, \{q0\}, \delta1, q0, \emptyset\}$

$M2 = \{\emptyset, \{q0\}, \delta2, q0, \{q0\}\}$

$M3 = \{\{x\}, \{q0,q1\}, \delta3, q0, \{q1\}\}$

Tabela 2 – Autômatos finitos correspondentes às expressões regulares com zero operadores

ER	Autômato Finito Correspondente
r = ø	
r = ε	
r = x (x pertence a Σ)	

Fonte: Dados a partir da descrição e base de indução

b) ***Hipótese de Indução****:* Seja r uma expressão regular com até n > 0 operadores. Suponha que é possível definir um autômato finito que aceita a linguagem gerada por r;

c) ***Passo de Indução****:* Seja r uma expressão regular com n + 1 operadores. Então r pode ser representada por um dos seguintes casos, onde r1 e r2 possuem conjuntamente no máximo n operadores:

r = r1 + r2 r = r1r2 r = r1*

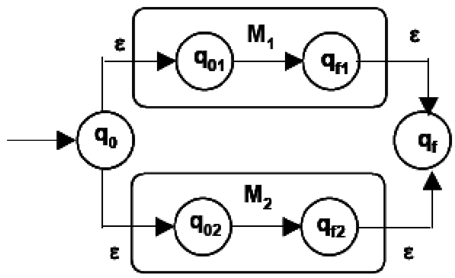
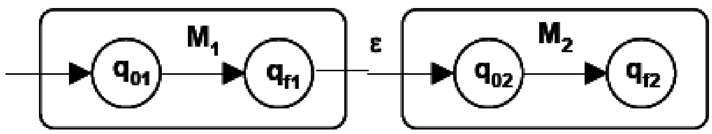
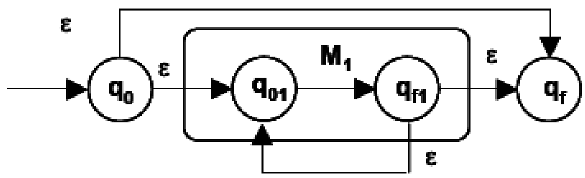
Portanto, por hipótese de indução, é possível construir os autômatos:

M1 = (Q1, Σ1, δ1, q01, {qf1}) e M2 = (Q2, Σ2, δ2, q01, {qf2}), tais que:

ACEITA(M1) = GERA(r1) e ACEITA(M2) = GERA(r2)

Nota-se que, sem perda de generalidade, é possível assumir que M1 e M2 possuem exatamente um estado final.

Tabela 3 – Autômatos finitos correspondentes às expressões regulares com n+1 operadores

ER	Autômato Finito Correspondente
r = r1 + r2	
r = r1r2	
r = r1*	

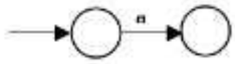
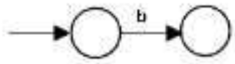
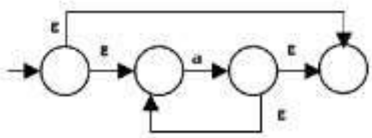
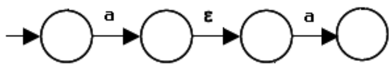
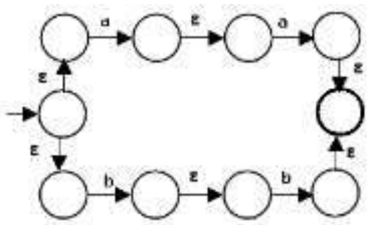
Fonte: Construção dos Autômatos a partir das Expressões Regulares

Teorema 2: Linguagem Regular → Expressão Regular

Se L é uma Linguagem Regular, então existe uma Expressão Regular r tal que:

GERA(r) = L(r)

Tabela 4 – Expressões Regulares e os correspondentes autômatos finitos

ER	Autômato Finito Correspondente
a	
b	
a*	
aa	
bb	
(aa + bb)	

Fonte: Autômatos construídos a partir das Expressões Regulares