Linguagens Formais e Autômatos

Aula 08 - Expressões regulares

Prof. Dr. Daniel Lucrédio
Departamento de Computação / UFSCar
Última revisão: ago/2015

Referências bibliográficas

- **Introdução à teoria dos autômatos, linguagens e computação / John E. Hopcroft, Rajeev Motwani, Jeffrey D. Ullman** ; tradução da 2.ed. original de Vandenberg D. de Souza. - Rio de Janeiro : Elsevier, 2002 (Tradução de: Introduction to automata theory, languages, and computation - ISBN 85-352-1072-5)
  - Capítulo 3 - Seções 3.1 e 3.4
- **Introdução à teoria da computação / Michael Sipser** ; tradução técnica Ruy José Guerra Barretto de Queiroz ; revisão técnica Newton José Vieira. -- São Paulo : Thomson Learning, 2007 (Título original : Introduction to the theory of computation. "Tradução da segunda edição norte-americana" - ISBN 978-85-221-0499-4)
  - Capítulo 1 - Seção 1.3

Expressões regulares

- Autômatos denotam linguagens
- Autômatos possuem duas notações
  - Diagrama de estados
  - Tabela de transições
- Vimos apenas uma notação para linguagens
  - Notações de conjuntos / formadores de conjuntos
  - Ex: {w | regra sobre w}
- Existe outra notação
  - Expressões regulares

Expressões regulares

- Notação algébrica
- Definição de álgebra (by wikipedia)
  - Um conjunto A e uma coleção de operações sobre A
  - Operações k-árias
    - 0-árias: ex: constantes, 2, x, y
    - 1-árias: ex: -10
    - 2-árias: ex: 2+2, 3*y

Expressões regulares

- Na teoria da computação
- Álgebra envolve
  - Conjunto A = alfabeto
  - Operações = operações regulares
- Operações regulares
  - Operações sobre membros de um alfabeto
  - Linguagens regulares são fechadas sob as operações regulares

Conjuntos fechados sob uma operação

- Exemplo:
  - N = {1,2,3,...} (conjunto de números naturais)
  - N é fechado sob multiplicação
  - Ou seja: para quaisquer x e y em N
    - x * y também está em N
  - N não é fechado sob divisão
    - Contra-exemplo: 1 e 2 estão em N, mas ½ não está!
- Definição:
  - Uma coleção de objetos é fechada sob alguma operação se, aplicando-se essa operação a membros da coleção, recebe-se um objeto ainda na coleção

Operações regulares

- Operações que:
  - Aplicadas sobre elementos de linguagens regulares
  - Resultam em linguagens regulares
- Em outras palavras:
  - Sejam L1 e L2 duas linguagens regulares
  - L1 op_{reg} L2 é regular

Operações regulares

- São 3 as operações regulares:
  - União: $A \cup B = \{x \mid x \in A \text{ ou } x \in B\}$
  - Concatenação: $A.B = \{xy \mid x \in A \text{ e } y \in B\}$
  - Estrela (ou fechamento, ou fechamento de Kleene): $A^* = \{x_1x_2...x_k \mid k \geq 0 \text{ e cada } x_i \in A\}$
- União e concatenação são operações binárias
- Estrela é uma operação unária

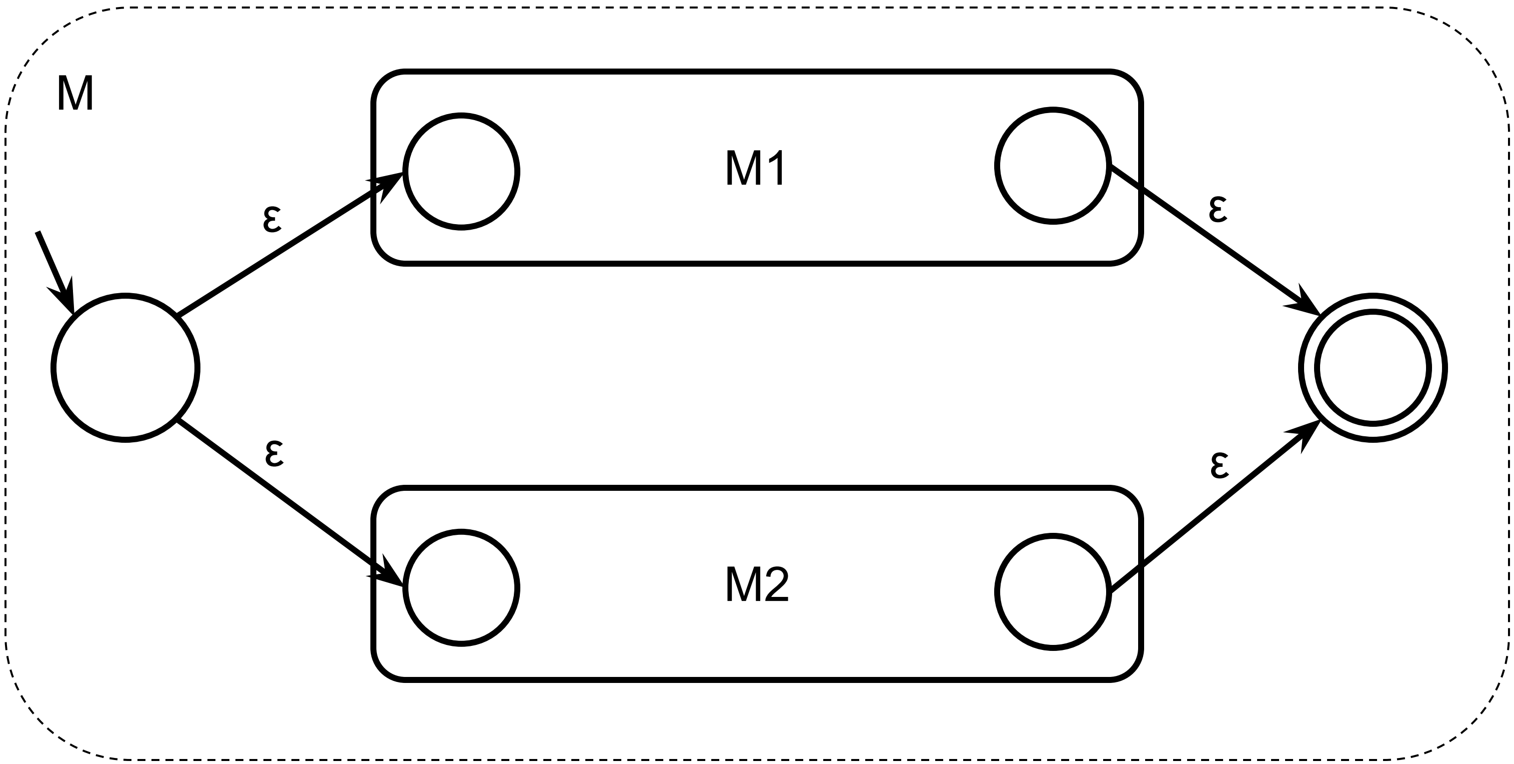
Operações regulares

- União
  - L = {001, 10, 111} e M = {ε, 001}
  - L ∪ M = {ε, 10, 001, 111}
- Concatenação
  - L = {001, 10, 111} e M = {ε, 001}
  - L.M (ou LM) = {001,10,111,001001,10001,111001}
- Estrela
  - L = {0, 11}
  - L* = {ε, 0, 00, 000, 000, 11, 011, 1111, 00011011, ... (não há uma ordem lógica aqui)}

Operações regulares

- Teorema: A classe de linguagens regulares é fechada sob a operação de união
  - Em outras palavras: se A1 e A2 são linguagens regulares, A1 $\cup$ A2 é regular
- Prova:
  - Construção
  - A1 é regular, existe um autômato M1
  - A2 é regular, existe um autômato M2
  - Construímos um autômato M que simula M1 e M2, aceitando se uma das simulações aceita

Operações regulares

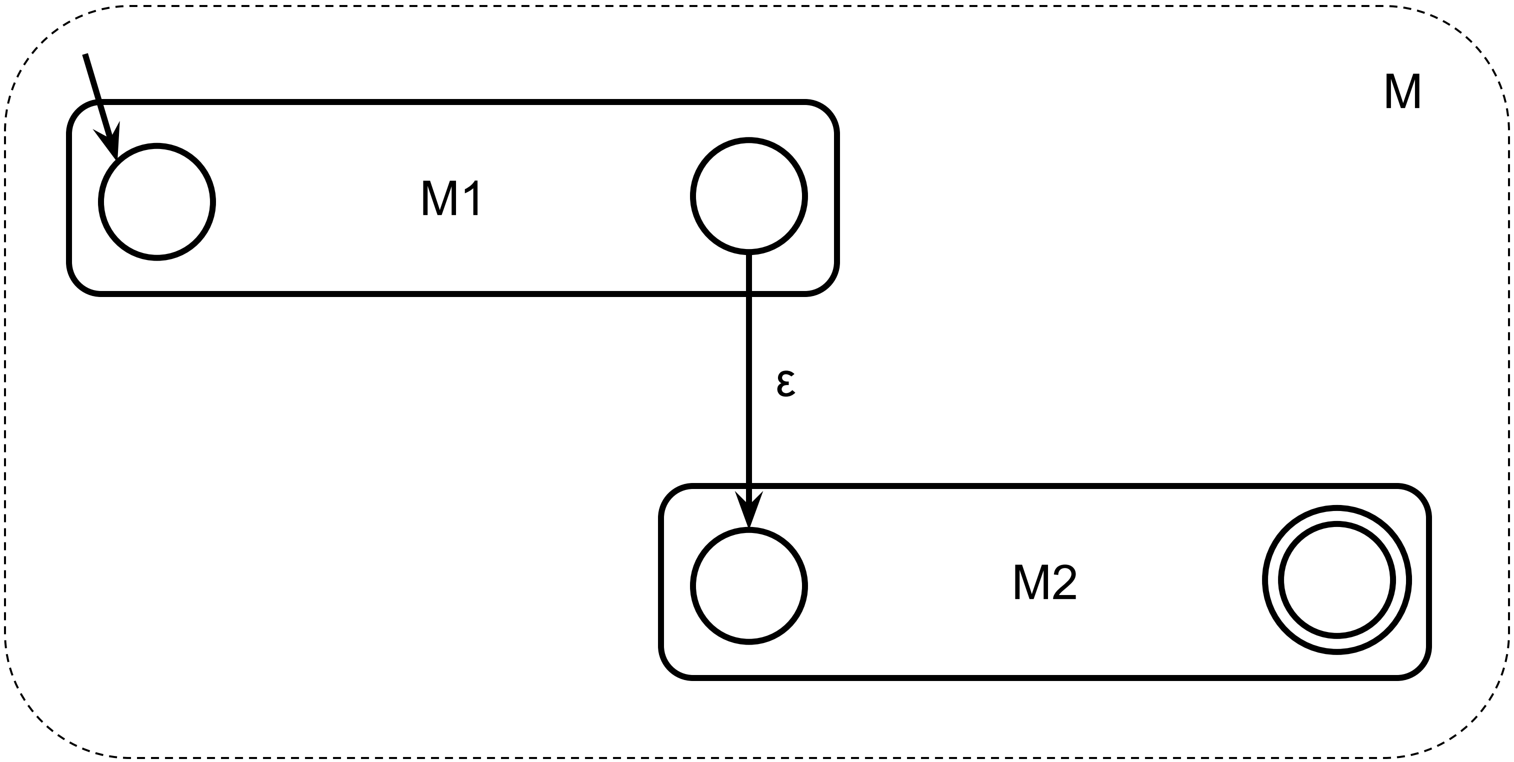


$$L(M) = L(M1) \cup L(M2)$$

Operações regulares

- Teorema: A classe de linguagens regulares é fechada sob a operação de concatenação
  - Em outras palavras, se A1 e A2 são linguagens regulares, A1.A2 é regular
- Prova:
  - Construção
  - A1 é regular, existe um autômato M1
  - A2 é regular, existe um autômato M2
  - Construímos um autômato M que simula M1 e em seguida M2, passando de M1 para M2 quando M1 aceita, e aceitando quando M2 aceita

Operações regulares

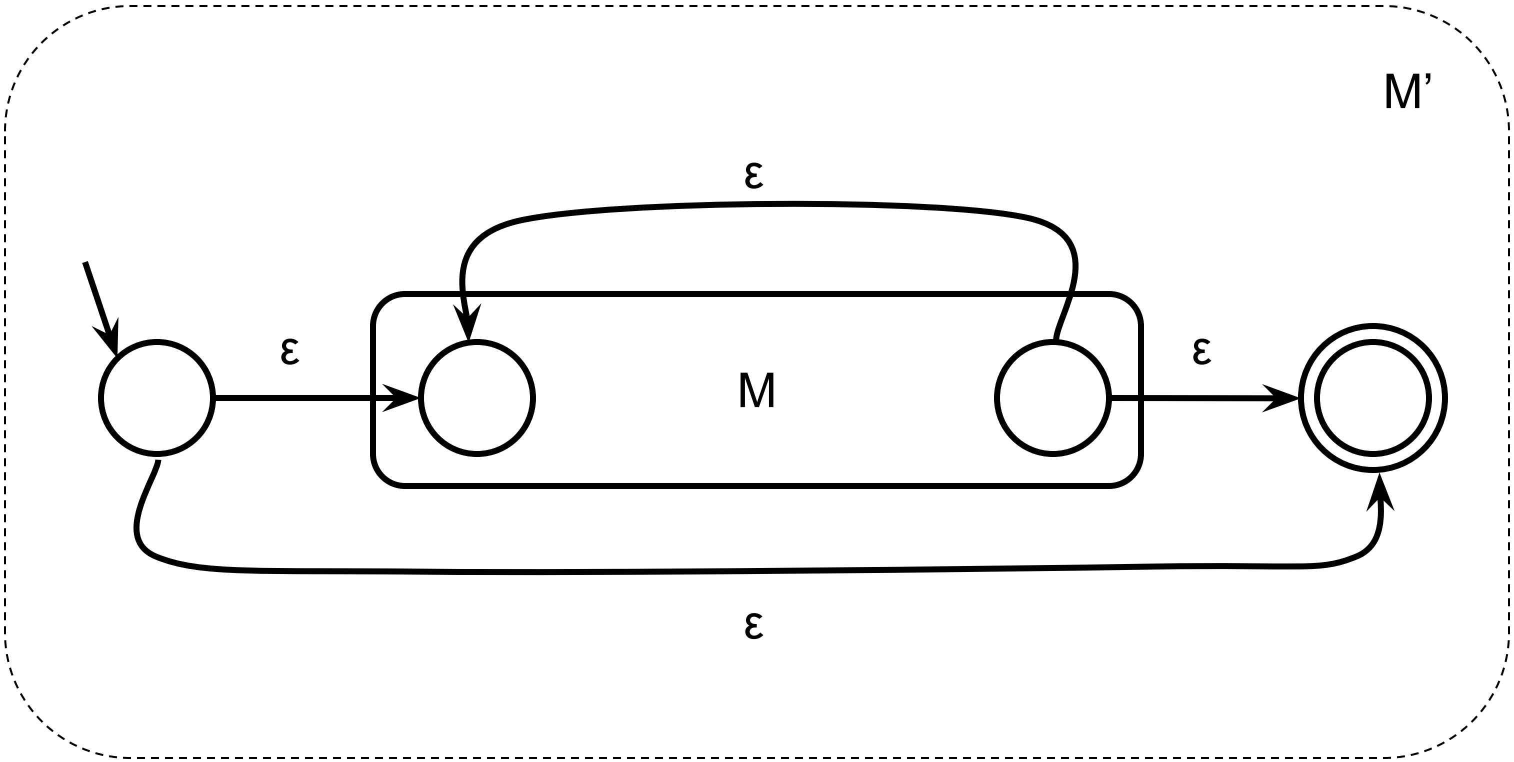


$$L(M) = L(M1) \,.\, L(M2)$$

Operações regulares

- Teorema: A classe de linguagens regulares é fechada sob a operação de estrela
  - Em outras palavras, se A é uma linguagem regular, A* é regular
- Prova:
  - Construção
  - A é regular, existe um autômato M
  - Construímos um autômato M’ que simula M, com a possibilidade de ir direto para o estado de aceitação, e a possibilidade de voltar de um estado de aceitação para o inicial

Operações regulares



L(M') = L(M)*

Expressões regulares

- Existe ainda uma “quarta operação regular”
  - Constantes (operações 0-árias)
  - A diferença é sutil
- As operações operam sobre conjuntos
  - Mas muitas vezes as usaremos com símbolos
  - Ou seja, em uma concatenação, ao invés de dizer {a}.{b}, queremos dizer a.b, ou simplesmente ab
- Por isso, iremos considerar que uma constante é uma operação
  - Toma como entrada um símbolo
  - E o resultado é uma linguagem (conjunto), cujo único elemento é aquele símbolo

Expressões regulares

- Conceito de “variáveis”
  - Símbolos que representam linguagens regulares
  - Ex: a.b, onde a = {w | w contém número par de 0s} e b = {0,11,101}

Expressões regulares

- Construindo expressões regulares
  - Assim como aritmética
  - Expressões elementares (constantes e/ou variáveis)
  - Operadores que formam expressões mais complexas
- Operadores possuem precedência
  - Ex: multiplicação sobre adição, etc
- Métodos para agrupar operadores
  - Ex: parêntesis, colchetes, chaves, etc

Expressões regulares

- Constantes:
  - ε e Ø são expressões regulares
    - L(ε) = {ε}
    - L(Ø) = Ø
  - Se a é um símbolo qualquer, a é uma expressão regular
    - L(a) = {a}
- União
  - Se E e F são expressões regulares, E + F é uma expressão regular
    - L(E+F) = L(E) ∪ L(F)

Expressões regulares

- Concatenação
    - Se E e F são expressões regulares, EF é uma expressão regular
        - L(EF) = L(E).L(F)
- Estrela
    - Se E é uma expressão regular, E* é uma expressão regular
        - L(E*) = (L(E))*
- Parêntesis
    - Se E é uma expressão regular, (E) é uma expressão regular
        - L((E)) = L(E)

Expressões regulares

- Precedência
  - Estrela → concatenação → união
  - Ex: 01*+1
- Parêntesis
  - Mudam a precedência
  - Ex: (01)*+1 ou 0(1*+1)

Exemplos de expressões regulares

- Alfabeto = {0,1}
  - 0^*10^* = {w | w contém um único 1}
  - $01 + 10$ = {01,10}
  - $(\varepsilon + 0)1^*$ = {w | w é uma sequência de zero ou mais 1s, começando opcionalmente com 0}
  - $(0+1)^*$ = Conjunto das partes do alfabeto ou conjunto de todas as cadeias possíveis sobre o alfabeto, incluindo a cadeia vazia ($|w| \geq 0$)
  - $(0+1)(0+1)^*$ = Idem ao exemplo acima, mas sem a cadeia vazia ($|w| \geq 1$)

Exercícios

- Escreva expressões regulares correspondentes às seguintes linguagens:
  - {w | w começa com um 1 e termina com um 0}
    - Resp: 1(0+1)*0
  - {w | w contém pelo menos três 1s}
    - Resp: 0*10*10*1(0+1)*
  - {w | o comprimento de w é no máximo 5}
    - Resp: (0+1+ε)(0+1+ε)(0+1+ε)(0+1+ε)(0+1+ε)
  - {w | toda posição ímpar de w é um 1}
    - Resp: (1(0+1))* + 1((0+1)1)* ou (1(0+1))*(1+ε)

Leis algébricas para expressões regulares

Leis algébricas para expressões regulares

- É possível (e muitas vezes necessário) simplificar expressões regulares
  - Ex: $1^*0 + 1^*0(\varepsilon+0+1)^*(\varepsilon+0+1) = 1^*0(0+1)^*$
- Existem algumas leis algébricas que facilitam esse processo

Leis algébricas para expressões regulares

- Associatividade e comutatividade
  - L+M=M+L
    - Ex: 0+1 = 1+0
  - (L+M)+N=L+(M+N)
    - Ex: (a* + bc) + a = a* + (bc + a)
  - (LM)N=L(MN)
    - Ex: (00(1+0))111=00((1+0)111)

Leis algébricas para expressões regulares

- Identidades (elemento neutro) e aniquiladores
  - $\emptyset+L=L+\emptyset=L$ ($\emptyset$ é identidade para união)
  - $\varepsilon L=L\varepsilon=L$ (ε é identidade para concatenação)
  - $\emptyset L=L\emptyset=\emptyset$ ($\emptyset$ é aniquilador para concatenação)
- Exs:
  - $\varepsilon a(b+c)+aa\varepsilon = a(b+c)+aa$
  - $\emptyset(\varepsilon+1)^*(1+0(01^*10(0+1))) + 01 = \emptyset + 01 = 01$

Leis algébricas para expressões regulares

- Leis distributivas
  - L(M+N)=LM+LN
  - (M+N)L=ML+NL
    - Exs:
      - 0(0+1) = 00 + 01
      - (0+1)(0+1) = (0+1)0 + (0+1)1
      - (0+1)(0+1) = 0(0+1) + 1(0+1)

Leis algébricas para expressões regulares

- Lei da idempotência
  - L+L=L
    - Exs:
      - (0+1+ε) + (0+1+ε)=(0+1+ε)
      - (0+1+ε)+01*0+(ε+0+1)+(ε+1+0)=(0+1+ε)+01*0

Leis algébricas para expressões regulares

- Leis envolvendo fechamentos
  - $(L^*)^*=L^*$
    - Ex: $((01)^*)^*=(01)^*$
  - $\emptyset^*=\varepsilon$
  - $\varepsilon^*=\varepsilon$
- Operadores de fechamento adicionais
  - $L^+=LL^*=L^*L$
  - $L^*=L^++\varepsilon$
  - $L?=\varepsilon+L$

Exemplos de simplificação

0+010

0**ε**+010

0**(ε+10)**

0**(10)?**

ab*b(a+ε)

ab*b**a?**

a**b**$^{+}$a?

a+(b+c+ε)a(b+c)+ca+ba

a+**(b+c)?**a(b+c)+ca+ba

a+(b+c)?a(b+c)+**(c+b)**a

εa+(b+c)?a(b+c)+(c+b)a

εa+**(c+b)a+(b+c)?a(b+c)**

(ε+c+b)a+(b+c)?a(b+c)

(b+c)?a+(b+c)?a(b+c)

(b+c)?a**ε**+(b+c)?a(b+c)

(b+c)?a**(ε+b+c)**

(b+c)?a**(b+c)?**

Fim

Aula 08 - Expressões regulares