Linguagens Formais e Autômatos (CC5220/CCM420)

Aula 02 - Conceitos relacionados a Gramática e Linguagens Formais

Prof. Luciano Rossi

Ciência da Computação
Centro Universitário FEI

2º Semestre de 2025

- Um **alfabeto** é um conjunto finito de elementos chamados **símbolos**.

Exemplo

- $\Sigma_1 = \{0, 1\}$;
- $\Sigma_2 = \{a, b, \ldots, z\}$;
- $\Gamma_1 = \{0, 1, a, b\}$;
- $\Gamma_2 = \{if, while, for, =\}$;

- Geralmente os alfabetos são representados por letras gregas maiúsculas.

- Dado um alfabeto Γ, uma **cadeia** (sobre um alfabeto) é uma sequência $w_1, w_2, \ldots, w_n$ onde $w_i \in \Gamma$ para $1 \leqslant i \leqslant n$.
- Uma cadeia é uma sequência finita de símbolos de um alfabeto;

Exemplo

- 01001 é uma cadeia sobre o alfabeto $\{0, 1\}$;
- $abracadabra$ é uma cadeia sobre o alfabeto $\{a, b, \ldots, z\}$;

- Cadeias geralmente são denotadas por letras gregas minúsculas $(\omega, \alpha, \beta, \gamma)$;
- Cadeias também são chamadas de strings ou palavras.

- O **comprimento** de uma cadeia ω, denotado por $|\omega|$, é o número de elementos na sequência.

Exemplo

- $\omega = maycon$ sobre o alfabeto $\{a, b, \ldots, z\}$;
    - $|\omega| = 6$
- $\omega = 011011$ sobre o alfabeto $\{0, 1\}$;
    - $|\omega| = 6$
- $\omega = 011011$ sobre o alfabeto $\{1, 01, 11\}$;
    - $|\omega| = 4$

- A **concatenação** da cadeia $\alpha = a_1 a_2 \ldots a_n$ com a cadeia $\beta = b_1 b_2 \ldots b_m$, denotada por $\alpha\beta$, é a cadeia $a_1 a_2 \ldots a_n b_1 b_2 \ldots b_m$

Exemplo

- Sejam $\alpha = vovo$ e $\beta = juju$, então:

$$\alpha\beta = vovojuju$$

- Dado uma cadeia α, definimos:

$$\alpha^k = \underbrace{\alpha\alpha\ldots\alpha}_{k}$$

Exemplo

- Se $\alpha = aba$, então $\alpha^3 = \underbrace{aba}_{\alpha}\underbrace{aba}_{\alpha}\underbrace{aba}_{\alpha}$
- Se $\beta = 01110$, então $\beta^2 = \underbrace{01110}_{\beta}\underbrace{01110}_{\beta}$

- Quando $\alpha = a$, onde a é um símbolo do alfabeto, escrevemos a^k por brevidade.

Exemplo

- $ab^2ac^3b = a\underbrace{bb}_{b^2}a\underbrace{ccc}_{c^3}b$

- A **cadeia vazia**, denotada por ϵ, é a cadeia de comprimento 0.

Note

- Dada uma cadeia ω sobre um alfabeto Σ

$$\omega\epsilon = \epsilon\omega = \omega$$

- O **reverso** da cadeia $\omega = w_1 w_2 \ldots w_n$, denotado por ω^R, é a cadeia $w_n w_{n-1} \ldots w_1$

Exemplo

- Se $\alpha = abcde$, então $\alpha^R = edcba$

- Uma cadeia β é **subcadeia** de uma cadeia α se existem cadeias α e γ tais que $\omega = \alpha\beta\gamma$.

Exemplo

- $\beta = ab$ é subcadeia de $\omega = aaaabbb$, pois $\omega = \underbrace{aaa}_{\alpha}\underbrace{ab}_{\beta}\underbrace{bb}_{\gamma}$
- $\beta = 01$ é subcadeia de $\omega = 01101$, pois $\omega = \underbrace{\epsilon}_{\alpha}\underbrace{01}_{\beta}\underbrace{101}_{\gamma}$
- $\beta = ba$ é subcadeia de $\omega = aaabba$, pois $\omega = \underbrace{aaab}_{\alpha}\underbrace{ba}_{\beta}\underbrace{\epsilon}_{\gamma}$

Dado um alfabeto Σ,

- o **fecho de Kleene** de Σ, denotado por Σ^*, é $\Sigma^* = \bigcup_{i=0}^{\infty} \Sigma^i$
- o **fecho positivo** de Σ, denotado por Σ^+, é $\Sigma^+ = \bigcup_{i=1}^{\infty} \Sigma^i = \Sigma^* \setminus \{\epsilon\}$

Exemplo

Seja $\Sigma = \{0, 1\}$, então: Seja $\Sigma = \{0, 1\}$, então:

- $\Sigma^* = \{\epsilon, 0, 1, 00, 01, 10, 11, 100, \dots\}$
- $\Sigma^+ = \{0, 1, 00, 01, 10, 11, 100, \dots\}$

Definição

Uma **linguagem** L sobre um alfabeto Σ é um subconjunto de Σ^*, ou seja:

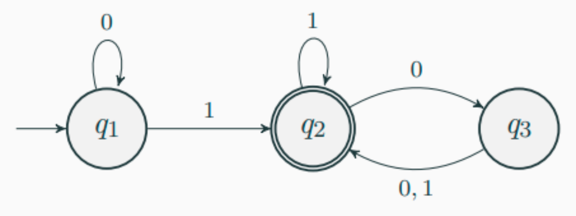
$$L \subseteq \Sigma^*$$

- Por exemplo:
  - Seja $\Sigma^* = \{\epsilon, 0, 1, 00, 01, 10, 11, 100, \ldots\}$
  - Seja $L = \{000, 001, 010, 100, 011, 101, 110, 111\}$
  - Veja que $L \subseteq \Sigma^*$
  - Desse modo, L é um linguagem.

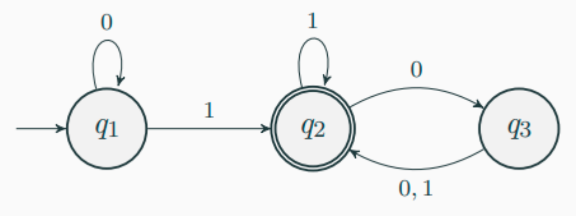
Exemplos

- $L_1 = \{1, 10, 11, 100\}$;
- $L_2 = \{w \in \{0,1\} : w \text{ contém o mesmo número de 0's e 1's}\}$;
- $L_3 = \{0^n 1^n \in \{0,1\}^* : n \geqslant 1\}$;
- $L_4 = \{a^i b^j \in \{a,b\}^* : 1 \leqslant i \leqslant j\}$;
- $L_5 = \{\epsilon\}$;
- $L_6 = \{\}$;
- $L_7 = \{w \in \{0,1,2\}^* : \text{ o quinto símbolo de } w \text{ é } 2\}$;
- $L_8 = \{w : w \text{ é um programa sintaticamente correto em C}\}$.

- É um **modelo** computacional com uma quantidade limitada (finita) de memória.
- Modelo computacional **mais simples** que estudaremos no curso.
- Aplicações
  - Modelagem de controladores simples: estabelecem uma terminologia e técnica padrão.
  - Usado na fase de analise léxica dos compiladores.
  - É um dispositivo reconhecedor de linguagem.



- Três estados: q_1, q_2, q_3.
- O estado inicial (q_1) é indicado por uma flecha vinda de lugar algum.
- Um estado final (q_2) é indicado por um círculo com aro duplo.
- As flechas ligando estados são chamadas de transições.
- Quando o autômato recebe uma cadeia de entrada, ele processa a cadeia e aceita ou rejeita ela.



Exemplo

Vamos processar as seguintes cadeias com o autômato M:

- 010101
- 011000
- 100

Qual linguagem é aceita pelo autômato M?

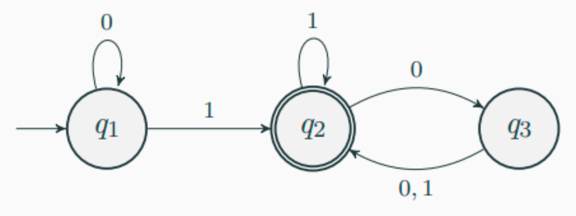
Perguntas

- Um autômato pode ter mais do que um estado inicial?
- Um autômato pode ter mais de uma flecha saindo do mesmo estado com o mesmo símbolo do alfabeto?
- Um autômato precisa ter um estado final?
- Um autômato pode ter mais do que um estado final?
- Cada estado precisa ter uma flecha saindo com cada símbolo do alfabeto?

Definição

Uma autômato finito determinístico (AFD) é uma 5-upla (Q, Σ, δ, q_0, F), onde:

- Q é um conjunto finito de elementos chamados estados;
- Σ é um alfabeto
- $\delta : Q \times \Sigma \to Q$ é a função de transição
- $q_0 \in Q$ é o estado inicial
- $F \subseteq Q$ é o conjunto de estados finais (ou de aceitação)



$M = (Q, \Sigma, \delta, q_1, F)$, onde $Q = \{q_1, q_2, q_3\}$, $\Sigma = \{0, 1\}$, $F = \{q_2\}$ e δ é definido como:

δ	0	1
q_1	q_1	q_2
q_2	q_3	q_2
q_3	q_2	q_2

- A memória do AFD = seus estados = Finita
- Determinismo: para cada símbolo da entrada existe exatamente um estado para o qual o autômato pode transitar do estado atual

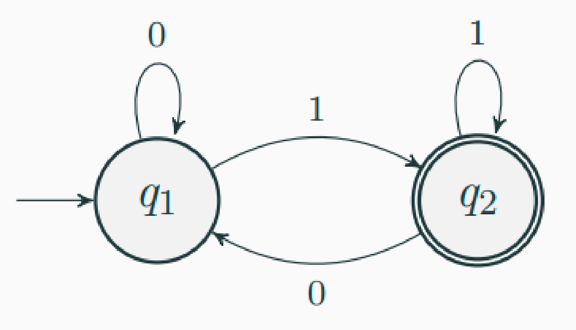
- Como um AFD aceita uma cadeia?
- Seja $M = (Q, \Sigma, \delta, q_1, F)$ um AFD e seja $\omega = w_1 w_2 \ldots w_n$ uma cadeia sobre Σ. Dizemos que M aceita ω se existe uma sequência de estados $(r_1, r_2, \ldots, r_n)$ tal que
  - $r_1 = q_1$
  - $\delta(r_i, w_i) = r_{i+1}, \forall i = 1, \ldots, n-1$
  - $r_n \in F$

- Se X é o conjunto de todas as cadeias que um AFD M aceita, então dizemos que:
  - X é a linguagem de M
  - $L(M) = X$
  - M reconhece X

Exemplo

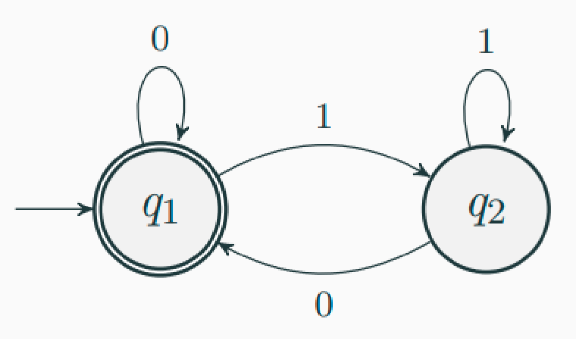
$L(M^*) = \{w \in \{0,1\}^* : w$ contém ao menos um símbolo 1 e contém um número par de zeros após o último 1 $\}$.

Um AFD aceita várias cadeias mas reconhece apenas uma linguagem!



Qual linguagem M' reconhece:

$$L(M') = \{w : w \text{ termina em } 1\}$$



Qual linguagem M'' reconhece:

$$L(M') = \{w : w = \epsilon \text{ ou termina em } 0\}$$

Definição

Uma linguagem é regular se algum AFD a reconhece.

Construa AFD's que reconheçam as seguintes linguagens e represente-os formalmente:

- $L_1 = \{w \in \{0,1\}^* : w \text{ termina em } 01\}$ (com 3 estados)
- $L_2 = \{w \in \{a,b\}^* : w \text{ começa e termina com o mesmo símbolo}\}$ (com 5 estados)
- $L_3 = \{w \in \{0,1\}^* : \text{ toda posição ímpar de w é } 1\}$ (com 4 estados)
- $L_4 = \{w \in \{0,1\}^* : w \text{ não contém } 110 \text{ como subcadeia}\}$ (com 4 estados)
- $L_5 = \{w \in \{0,1\}^* : \text{ todo } 0 \text{ em w é seguido de pelo menos um } 1\}$ (com 4 estados)
- $L_6 = \{w \in \{0,1\}^* : w \text{ contém ao menos dois } 0\text{s e no máximo um } 1\}$ (com 7 estados)

Linguagens Formais e Autômatos (CC5220/CCM420)

Aula 02 – Conceitos relacionados a Gramática e Linguagens Formais

Prof. Luciano Rossi

Ciência da Computação
Centro Universitário FEI

2° Semestre de 2025