CC1004 - Modelos de Computação
Teóricas 3 e 4

Ana Paula Tomás

Departamento de Ciência de Computadores
Faculdade de Ciências da Universidade do Porto

Fevereiro 2021

Introdução às Linguagens Formais

- **Alfabeto** Σ
- **Palavra** $x, y, z, w, \ldots$
- **Palavra vazia** ε
- $\Sigma^\star$ é o conjunto das palavras de alfabeto Σ
- **Linguagem** $L, M, S, \ldots,$ $L \subseteq \Sigma^\star,$ $x \in L$
- **Linguagem vazia** $\{\} = \emptyset \neq \{\varepsilon\}$
- **Comprimento** de uma palavra x denota-se por $|x|$

Operações: União, Interseção, Complementar, Diferença, Concatenação e Fecho de Kleene

União: $L \cup M = \{x \mid x \in L \text{ ou } x \in M\}$

Interseção: $L \cap M = \{x \mid x \in L \text{ e } x \in M\}$

Complementar: $\overline{L} = \{x \mid x \notin L\} = \Sigma^\star \setminus L$

Diferença: $L \setminus M = \{x \mid x \in L \text{ e } x \notin M\}$

Concatenação: $LM = \{xy \mid x \in L \text{ e } y \in M\}$ S

Fecho de Kleene: $L^\star = \bigcup_{n\in\mathbb{N}} L^n = L^0 \cup L^1 \cup L^2 \cup L^3 \cup \cdots = \{x \mid x \text{ é sequência finita de palavras de } L\}$

onde $L^0 = \{\varepsilon\}$, $L^1 = L$, ; $L^n = \{x \mid x \text{ é concatenação de } n \text{ palavras de } L\}$

Propriedades algébricas

Proposição:
Quaisquer que sejam as linguagens R, S e T de alfabeto Σ tem-se:

$$
\begin{array}{ll}
R \cup S = S \cup R & (R \cup S) \cup T = R \cup (S \cup T) \\
R \cap S = S \cap R & (R \cap S) \cap T = R \cap (S \cap T) \\
R(ST) = (RS)T & R(S \cup T) = RS \cup RT \\
(R \cup S)T = RT \cup ST & \emptyset^\star = \{\varepsilon\} \\
(R^\star)^\star = R^\star & (\{\varepsilon\} \cup R)^\star = R^\star \\
(R^\star S^\star)^\star = (R \cup S)^\star & \{\varepsilon\}R = R = R\{\varepsilon\} \\
\emptyset R = \emptyset = R\emptyset &
\end{array}
$$

Exemplos

- A linguagem das palavras de $\{\mathtt{a},\mathtt{b}\}^\star$ que **têm** aa **como prefixo e** bbb **como sufixo** é
$$\{\mathtt{aa}\}\{\mathtt{a},\mathtt{b}\}^\star\{\mathtt{bbb}\}$$
pois as suas palavras são da forma $\mathtt{aa}w\mathtt{bbb}$, com $w \in \Sigma^\star$.

- A linguagem das palavras de $\{0,1\}^\star$ que **têm** 00 **como subpalavra** é
$$\{0,1\}^\star\{00\}\{0,1\}^\star$$
pois as suas palavras são da forma $y00w$, com $y, w \in \Sigma^\star$.

- A linguagem das palavras em $\{0,1\}^\star$ que **têm comprimento par** é
$$\{01,10,00,11\}^\star$$
ou, equivalentemente, $(\{0,1\}\{0,1\})^\star$.

Ideia: se x tem comprimento par, $x = \varepsilon$ ou x é uma sequência finita de palavras de comprimento 2, sem restrição.

Exemplos

- A linguagem das palavras de $\{\mathtt{a},\mathtt{b}\}^\star$ que **têm** $\mathtt{aa}$ **como prefixo e** $\mathtt{bbb}$ **como sufixo** é
$$\{\mathtt{aa}\}\{\mathtt{a},\mathtt{b}\}^\star\{\mathtt{bbb}\}$$
pois as suas palavras são da forma $\mathtt{aa}w\mathtt{bbb}$, com $w \in \Sigma^\star$.

- A linguagem das palavras de $\{0,1\}^\star$ que **têm** 00 **como subpalavra** é
$$\{0,1\}^\star\{00\}\{0,1\}^\star$$
pois as suas palavras são da forma $y00w$, com $y, w \in \Sigma^\star$.

- A linguagem das palavras em $\{0,1\}^\star$ que **têm comprimento par** é
$$\{01,10,00,11\}^\star$$
ou, equivalentemente, $(\{0,1\}\{0,1\})^\star$.

Ideia: se x tem comprimento par, $x = \varepsilon$ ou x é uma sequência finita de palavras de comprimento 2, sem restrição.

Exemplos

- A linguagem das palavras de $\{\mathtt{a},\mathtt{b}\}^\star$ que **têm** $\mathtt{aa}$ **como prefixo e** $\mathtt{bbb}$ **como sufixo** é
$$\{\mathtt{aa}\}\{\mathtt{a},\mathtt{b}\}^\star\{\mathtt{bbb}\}$$
pois as suas palavras são da forma $\mathtt{aa}w\mathtt{bbb}$, com $w \in \Sigma^\star$.

- A linguagem das palavras de $\{0,1\}^\star$ que **têm** 00 **como subpalavra** é
$$\{0,1\}^\star\{00\}\{0,1\}^\star$$
pois as suas palavras são da forma $y00w$, com $y, w \in \Sigma^\star$.

- A linguagem das palavras em $\{0,1\}^\star$ que **têm comprimento par** é
$$\{01,10,00,11\}^\star$$
ou, equivalentemente, $(\{0,1\}\{0,1\})^\star$.

Ideia: se x tem comprimento par, $x = \varepsilon$ ou x é uma sequência finita de palavras de comprimento 2, sem restrição.

Autómatos Finitos Determinísticos (AFDs)

Um **autómato finito determinístico** A é um modelo abstrato de uma máquina, sendo definido por $A = (S, \Sigma, \delta, s_0, F)$, em que

- S é **o conjunto de estados** e é **finito**;
- Σ é o **alfabeto** de símbolos de entrada;
- δ é uma função de $S \times \Sigma$ em S, designada por **função de transição**;
- s_0 é o **estado inicial**;
- $F \subseteq S$ é **o conjunto de estados finais** (estados de aceitação)

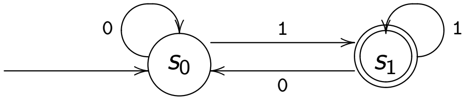
A **linguagem reconhecida pelo autómato** é o conjunto das palavras de $\Sigma^\star$ que o levam do estado s_0 a algum estado $s \in F$, sendo completamente processadas.

Determinístico: para cada par (s, a) existe um e um só s' tal que $\delta(s, a) = s'$.

Diagrama de transição de um AFD

Um autómato finito pode ser representado esquematicamente por um multigrafo dirigido com símbolos associados aos ramos. Esse multigrafo designa-se por **diagrama de transição** do autómato. Os **nós** correspondem aos estados do autómato. Um **ramo** de s para s' etiquetado por a indica que $\delta(s, a) = s'$.

Exemplo: AFD que reconhece $\{0, 1\}^\star\{1\}$, i.e., as palavras que terminam em 1.



Convenção: no diagrama, **os estados finais** são representados por duas circunferências e o **estado inicial** é apontado por uma seta.

Exemplos de AFDs

AFD para decidir se uma sequência de alfabeto $\{0,1\}$ **representa em binário um inteiro positivo múltiplo de 3**. Admitimos que possa ter 0's não significativos, ou seja, que o bit mais significativo possa ser 0 sem que o valor seja 0, como 0011.
Recordar que

$$b_{n-1}b_{n-2}\ldots b_2 b_1 b_{0_{(2)}} = \sum_{k=0}^{n-1} b_k 2^k$$

Ideia: Se X representa k então Xb é $2k+b$. Logo, $X1$ é $2k+1$ e $X0$ é $2k$.
Como é que a alteração afeta o resto da divisão de k por 3?

$$\begin{array}{ll} 2(3m)+0 = 3m' & 2(3m)+1 = 3m'+1 \\ 2(3m+1)+0 = 3m'+2 & 2(3m+1)+1 = 3m' \\ 2(3m+2)+0 = 3m'+1 & 2(3m+2)+1 = 3m'+2 \end{array}$$

Isto quer dizer que, se X é $3m$ então $X0$ é $3m'$, para $m' = 2m$. Se X é $3m+2$ então $X0$ é $3m'+1$, com $m' = 2m+1$, etc.

Exemplos de AFDs

AFD para decidir se uma sequência de alfabeto $\{0,1\}$ **representa em binário um inteiro positivo múltiplo de 3**. Admitimos que possa ter 0's não significativos, ou seja, que o bit mais significativo possa ser 0 sem que o valor seja 0, como 0011. Recordar que

$$b_{n-1}b_{n-2}\ldots b_2 b_1 b_{0_{(2)}} = \sum_{k=0}^{n-1} b_k 2^k$$

Ideia: Se X representa k então Xb é $2k+b$. Logo, $X1$ é $2k+1$ e $X0$ é $2k$.

Como é que a alteração afeta o resto da divisão de k por 3?

$$\begin{array}{ll}
2(3m)+0 = 3m' & 2(3m)+1 = 3m'+1 \\
2(3m+1)+0 = 3m'+2 & 2(3m+1)+1 = 3m' \\
2(3m+2)+0 = 3m'+1 & 2(3m+2)+1 = 3m'+2
\end{array}$$

Isto quer dizer que, se X é $3m$ então $X0$ é $3m'$, para $m' = 2m$. Se X é $3m+2$ então $X0$ é $3m'+1$, com $m' = 2m+1$, etc.

Exemplos de AFDs

AFD para decidir se uma sequência de alfabeto $\{0,1\}$ **representa em binário um inteiro positivo múltiplo de 3**. Admitimos que possa ter 0's não significativos, ou seja, que o bit mais significativo possa ser 0 sem que o valor seja 0, como 0011. Recordar que

$$b_{n-1}b_{n-2}\ldots b_2 b_1 b_{0\,(2)} = \sum_{k=0}^{n-1} b_k 2^k$$

Ideia: Se X representa k então Xb é $2k+b$. Logo, $X1$ é $2k+1$ e $X0$ é $2k$. Como é que a alteração afeta o resto da divisão de k por 3?

$$\begin{array}{ll} 2(3m)+0 = 3m' & 2(3m)+1 = 3m'+1 \\ 2(3m+1)+0 = 3m'+2 & 2(3m+1)+1 = 3m' \\ 2(3m+2)+0 = 3m'+1 & 2(3m+2)+1 = 3m'+2 \end{array}$$

Isto quer dizer que, se X é $3m$ então $X0$ é $3m'$, para $m' = 2m$. Se X é $3m+2$ então $X0$ é $3m'+1$, com $m' = 2m+1$, etc.

Exemplos de AFDs

AFD para decidir se uma sequência de alfabeto $\{0,1\}$ **representa em binário um inteiro positivo múltiplo de 3**. Admitimos que possa ter 0's não significativos, ou seja, que o bit mais significativo possa ser 0 sem que o valor seja 0, como 0011. Recordar que

$$b_{n-1}b_{n-2}\ldots b_2 b_1 b_{0_{(2)}} = \sum_{k=0}^{n-1} b_k 2^k$$

Ideia: Se X representa k então Xb é $2k+b$. Logo, $X1$ é $2k+1$ e $X0$ é $2k$. Como é que a alteração afeta o resto da divisão de k por 3?

$$\begin{array}{ll} 2(3m)+0 = 3m' & \quad 2(3m)+1 = 3m'+1 \\ 2(3m+1)+0 = 3m'+2 & \quad 2(3m+1)+1 = 3m' \\ 2(3m+2)+0 = 3m'+1 & \quad 2(3m+2)+1 = 3m'+2 \end{array}$$

Isto quer dizer que, se X é $3m$ então $X0$ é $3m'$, para $m' = 2m$. Se X é $3m+2$ então $X0$ é $3m'+1$, com $m' = 2m+1$, etc.

Exemplos de AFDs

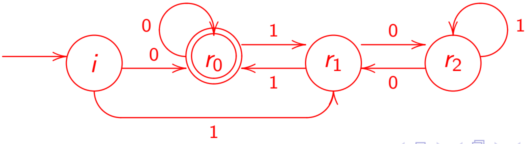
AFD para decidir se uma sequência de alfabeto $\{0, 1\}$ **representa em binário um inteiro positivo múltiplo de 3**. Admitimos que possa ter 0's não significativos, ou seja, que o bit mais significativo possa ser 0 sem que o valor seja 0, como 0011. Recordar que

$$b_{n-1}b_{n-2}\ldots b_2b_1b_{0_{(2)}} = \sum_{k=0}^{n-1} b_k 2^k$$

Ideia: Se X representa k então Xb é $2k+b$. Logo, $X1$ é $2k+1$ e $X0$ é $2k$. Como é que a alteração afeta o resto da divisão de k por 3?

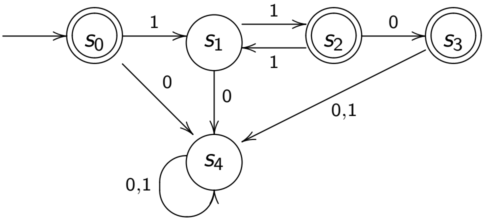
$$\begin{array}{ll} 2(3m)+0 = 3m' & 2(3m)+1 = 3m'+1 \\ 2(3m+1)+0 = 3m'+2 & 2(3m+1)+1 = 3m' \\ 2(3m+2)+0 = 3m'+1 & 2(3m+2)+1 = 3m'+2 \end{array}$$

Isto quer dizer que, se X é $3m$ então $X0$ é $3m'$, para $m' = 2m$. Se X é $3m+2$ então $X0$ é $3m'+1$, com $m' = 2m+1$, etc.



Exemplos de AFDs

- A linguagem $\{1^{2n} \mid n \geq 0\} \cup \{1^{2n}0 \mid n \geq 1\}$, de alfabeto $\{0, 1\}$, é a linguagem reconhecida, por exemplo, pelo AFD



Qual é o conjunto das palavras que levam o AFD de s_0 a cada estado?

De s_0 a s_0: $\{\varepsilon\}$

De s_0 a s_1: $\{1^{2n+1} \mid n \in \mathbb{N}\} = \{11\}^\star\{1\}$

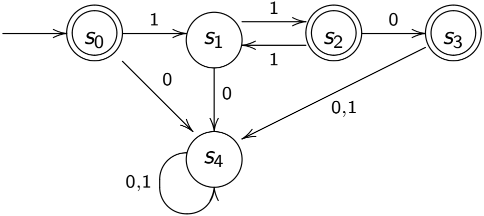
De s_0 a s_2: $\{1^{2n} \mid n \geq 1\} = \{11\}^\star\{11\}$

De s_0 a s_3: $\{1^{2n}0 \mid n \geq 1\} = \{11\}^\star\{110\}$

De s_0 a s_4: $\{x \mid x \text{ tem algum 0 e não é da forma } 1^{2n}0, \text{ para } n \geq 1\}$

Exemplos de AFDs

- A linguagem $\{1^{2n} \mid n \geq 0\} \cup \{1^{2n}0 \mid n \geq 1\}$, de alfabeto $\{0, 1\}$, é a linguagem reconhecida, por exemplo, pelo AFD



Qual é o conjunto das palavras que levam o AFD de s_0 a cada estado?

De s_0 a s_0: $\{\varepsilon\}$

De s_0 a s_1: $\{1^{2n+1} \mid n \in \mathbb{N}\} = \{11\}^\star\{1\}$

De s_0 a s_2: $\{1^{2n} \mid n \geq 1\} = \{11\}^\star\{11\}$

De s_0 a s_3: $\{1^{2n}0 \mid n \geq 1\} = \{11\}^\star\{110\}$

De s_0 a s_4: $\{x \mid x \text{ tem algum } 0 \text{ e não é da forma } 1^{2n}0, \text{ para } n \geq 1\}$

Descrição de Linguagens por Expressões Regulares

Os conjuntos das **expressões regulares sobre Σ** e das linguagens que descrevem podem ser assim definidos indutivamente:

(i) ε é uma expressão regular e descreve $\{\varepsilon\}$

(ii) $\emptyset$ é uma expressão regular e descreve $\emptyset$

(iii) Se $\mathsf{a} \in \Sigma$ então a é uma expressão regular e descreve $\{\mathsf{a}\}$

(iv) Se r e s são expressões regulares sobre Σ que descrevem R e S, então $(r + s)$, (rs) e $(r^\star)$ são expressões regulares e descrevem $R \cup S$, RS e $R^\star$ respetivamente.

(v) As expressões regulares sobre Σ e as linguagens por elas descritas são as obtidas por (i)–(iv).

Notação: $\mathcal{L}(r)$ denota a linguagem descrita pela expressão regular r.

Linguagem regular: se e só se pode ser descrita por uma expressão regular.

Descrição de Linguagens por Expressões Regulares

Os conjuntos das **expressões regulares sobre Σ** e das linguagens que descrevem podem ser assim definidos indutivamente:

(i) ε é uma expressão regular e descreve $\{\varepsilon\}$

(ii) $\emptyset$ é uma expressão regular e descreve $\emptyset$

(iii) Se $\mathrm{a} \in \Sigma$ então a é uma expressão regular e descreve $\{\mathrm{a}\}$

(iv) Se r e s são expressões regulares sobre Σ que descrevem R e S, então $(r+s)$, (rs) e $(r^\star)$ são expressões regulares e descrevem $R \cup S$, RS e $R^\star$ respetivamente.

(v) As expressões regulares sobre Σ e as linguagens por elas descritas são as obtidas por (i)–(iv).

Notação: $\mathcal{L}(r)$ denota a linguagem descrita pela expressão regular r.

Linguagem regular: se e só se pode ser descrita por uma expressão regular.

Descrição de Linguagens por Expressões Regulares

Os conjuntos das **expressões regulares sobre Σ** e das linguagens que descrevem podem ser assim definidos indutivamente:

(i) ε é uma expressão regular e descreve $\{\varepsilon\}$

(ii) $\emptyset$ é uma expressão regular e descreve $\emptyset$

(iii) Se $\mathtt{a} \in \Sigma$ então $\mathtt{a}$ é uma expressão regular e descreve $\{\mathtt{a}\}$

(iv) Se r e s são expressões regulares sobre Σ que descrevem R e S, então $(r + s)$, (rs) e $(r^\star)$ são expressões regulares e descrevem $R \cup S$, RS e $R^\star$ respetivamente.

(v) As expressões regulares sobre Σ e as linguagens por elas descritas são as obtidas por (i)–(iv).

Notação: $\mathcal{L}(r)$ denota a linguagem descrita pela expressão regular r.

Linguagem regular: se e só se pode ser descrita por uma expressão regular.

Descrição de Linguagens por Expressões Regulares

Os conjuntos das **expressões regulares sobre Σ** e das linguagens que descrevem podem ser assim definidos indutivamente:

(i) ε é uma expressão regular e descreve $\{\varepsilon\}$

(ii) $\emptyset$ é uma expressão regular e descreve $\emptyset$

(iii) Se $\mathrm{a} \in \Sigma$ então a é uma expressão regular e descreve $\{\mathrm{a}\}$

(iv) Se r e s são expressões regulares sobre Σ que descrevem R e S, então $(r + s)$, (rs) e $(r^\star)$ são expressões regulares e descrevem $R \cup S$, RS e $R^\star$ respetivamente.

(v) As expressões regulares sobre Σ e as linguagens por elas descritas são as obtidas por (i)–(iv).

Notação: $\mathcal{L}(r)$ denota a linguagem descrita pela expressão regular r.

Linguagem regular: se e só se pode ser descrita por uma expressão regular.

Descrição de Linguagens por Expressões Regulares

Os conjuntos das **expressões regulares sobre Σ** e das linguagens que descrevem podem ser assim definidos indutivamente:

(i) ε é uma expressão regular e descreve $\{\varepsilon\}$

(ii) $\emptyset$ é uma expressão regular e descreve $\emptyset$

(iii) Se $\mathtt{a} \in \Sigma$ então $\mathtt{a}$ é uma expressão regular e descreve $\{\mathtt{a}\}$

(iv) Se r e s são expressões regulares sobre Σ que descrevem R e S, então $(r + s)$, (rs) e $(r^\star)$ são expressões regulares e descrevem $R \cup S$, RS e $R^\star$ respetivamente.

(v) As expressões regulares sobre Σ e as linguagens por elas descritas são as obtidas por (i)–(iv).

Notação: $\mathcal{L}(r)$ denota a linguagem descrita pela expressão regular r.

Linguagem regular: se e só se pode ser descrita por uma expressão regular.

Exemplos: descrição de linguagens por expressões regulares

Assumindo que $\Sigma = \{\mathtt{a}, \mathtt{b}\}$

- Linguagem das palavras que não têm b's.

$$(\mathtt{a}^\star)$$

- Linguagem das palavras que começam por a.

$$(\mathtt{a}((\mathtt{a}+\mathtt{b})^\star))$$

Na prática, **omitimos alguns parentesis**, *considerando que a união, concatenação e fecho de Kleene teriam precedências idênticas às das soma, produto e potência em expressões numéricas.*

Linguagem das palavras que não têm b's: $\mathtt{a}^\star$
Linguagem das palavras que começam por a: $\mathtt{a}(\mathtt{a}+\mathtt{b})^\star$

Exemplos: descrição de linguagens por expressões regulares

Para $\Sigma = \{\mathtt{a}, \mathtt{b}\}$

- Linguagem das palavras que têm bb como subpalavra.

$$(\mathtt{a}+\mathtt{b})^\star\mathtt{bb}(\mathtt{a}+\mathtt{b})^\star$$

- Linguagem das palavras que terminam em aba.

$$(\mathtt{a}+\mathtt{b})^\star\mathtt{aba}$$

- Linguagem das palavras que têm comprimento ímpar.

$$((\mathtt{a}+\mathtt{b})(\mathtt{a}+\mathtt{b}))^\star(\mathtt{a}+\mathtt{b})$$

  ou, em alternativa, $(\mathtt{aa}+\mathtt{ab}+\mathtt{bb}+\mathtt{ba})^\star(\mathtt{a}+\mathtt{b})$.

- Linguagem das palavras que não têm a's ou terminam ba

$$\mathtt{b}^\star + (\mathtt{a}+\mathtt{b})^\star\mathtt{ba}$$

====== Fim da aula teórica 3 =====

Questões?

Para $\Sigma = \{0, 1\}$ descrever informalmente (mas com rigor) a linguagem descrita por cada uma das expressões regulares (abreviadas) seguintes.

- $01 + 100 + \varepsilon + 1001$
- $0^\star + 1^\star$
- $(0^\star + 1^\star)^\star$
- $(001^\star + \varepsilon)1$
- $(01^\star0 + 1)^\star$
- $(01^\star0 + 1)^\star01^\star$

Exemplos

Para $\Sigma = \{0, 1\}$ descrever informalmente (mas com rigor) a linguagem descrita por cada uma das expressões regulares (abreviadas) seguintes.

- $01 + 100 + \varepsilon + 1001$

  Linguagem constituída por apenas quatro palavras: 01, 100, ε, e 1001.

- $0^\star + 1^\star$

  Linguagem constituída pelas palavras que não têm simultaneamemte 0's e 1's, ou seja, são as que não têm 1's ou não têm 0's.

- $(0^\star + 1^\star)^\star$

  Esta expressão é equivalente a $(0 + 1)^\star$.

  Ambas descrevem a linguagem $\{0, 1\}^\star$, que é o conjunto de todas as palavras de alfabeto $\{0, 1\}$.

Exemplos

- $(001^\star + \varepsilon)1$

  Descreve a linguagem constituída pela palavra 1 e pelas palavras que começam por 001 e não têm mais 0's.

  - As palavras são da forma $w1$, com $w \in \mathcal{L}(001^\star + \varepsilon)$, isto é, com $w \in \{\varepsilon, 00, 001, 0011, 00111, \ldots\}$. Assim, $w1 \in \{1, 001, 0011, 00111, 001111, \ldots\}$.

  - Pela distributividade da concatenação relativamente à união (e associatividade), a expressão é equivalente a $001^\star 1 + \varepsilon 1$.

    Como ε é o elemento neutro da concatenação, $\varepsilon 1$ e 1 denotam a mesma linguagem, o que nos permite simplificar para $001^\star 1 + 1$. .
    E, concluimos que a expressão é equivalente a $0011^\star + 1$, pois, por definição de $1^\star$, as expressões $1^\star 1$ e $11^\star$ são equivalentes (descrevem a linguagem das palavras que têm algum 1 e não têm 0's).

Exemplos

- $(001^\star + \varepsilon)1$

  Descreve a linguagem constituída pela palavra 1 e pelas palavras que começam por 001 e não têm mais 0's.

  - As palavras são da forma $w1$, com $w \in \mathcal{L}(001^\star + \varepsilon)$, isto é, com $w \in \{\varepsilon, 00, 001, 0011, 00111, \ldots\}$. Assim, $w1 \in \{1, 001, 0011, 00111, 001111, \ldots\}$.

  - Pela distributividade da concatenação relativamente à união (e associatividade), a expressão é equivalente a $001^\star 1 + \varepsilon 1$.
    Como ε é o elemento neutro da concatenação, $\varepsilon 1$ e 1 denotam a mesma linguagem, o que nos permite simplificar para $001^\star 1 + 1$. .
    E, concluimos que a expressão é equivalente a $0011^\star + 1$, pois, por definição de $1^\star$, as expressões $1^\star 1$ e $11^\star$ são equivalentes (descrevem a linguagem das palavras que têm algum 1 e não têm 0's).

Exemplos

- $(001^\star + \varepsilon)1$

  Descreve a linguagem constituída pela palavra 1 e pelas palavras que começam por 001 e não têm mais 0's.

  - As palavras são da forma $w1$, com $w \in \mathcal{L}(001^\star + \varepsilon)$, isto é, com $w \in \{\varepsilon, 00, 001, 0011, 00111, \ldots\}$. Assim, $w1 \in \{1, 001, 0011, 00111, 001111, \ldots\}$.

  - Pela distributividade da concatenação relativamente à união (e associatividade), a expressão é equivalente a $001^\star 1 + \varepsilon 1$.

    Como ε é o elemento neutro da concatenação, $\varepsilon 1$ e 1 denotam a mesma linguagem, o que nos permite simplificar para $001^\star 1 + 1$. .
    E, concluimos que a expressão é equivalente a $0011^\star + 1$, pois, por definição de $1^\star$, as expressões $1^\star 1$ e $11^\star$ são equivalentes (descrevem a linguagem das palavras que têm algum 1 e não têm 0's).

Exemplos

- $(001^\star + \varepsilon)1$

  Descreve a linguagem constituída pela palavra 1 e pelas palavras que começam por 001 e não têm mais 0's.

  - As palavras são da forma $w1$, com $w \in \mathcal{L}(001^\star + \varepsilon)$, isto é, com $w \in \{\varepsilon, 00, 001, 0011, 00111, \ldots\}$. Assim, $w1 \in \{1, 001, 0011, 00111, 001111, \ldots\}$.

  - Pela distributividade da concatenação relativamente à união (e associatividade), a expressão é equivalente a $001^\star 1 + \varepsilon 1$.

    Como ε é o elemento neutro da concatenação, $\varepsilon 1$ e 1 denotam a mesma linguagem, o que nos permite simplificar para $001^\star 1 + 1$. .
    E, concluimos que a expressão é equivalente a $0011^\star + 1$, pois, por definição de $1^\star$, as expressões $1^\star 1$ e $11^\star$ são equivalentes (descrevem a linguagem das palavras que têm algum 1 e não têm 0's).

Exemplos

- $(01^\star 0 + 1)^\star$ descreve a linguagem constituída pelas palavras que têm **número par de** 0**’s**.
- $(01^\star 0 + 1)^\star 01^\star$ descreve a linguagem formada pelas palavras que têm **número ímpar de** 0**’s**.

Ideia: Por exemplo, $111101111111010100010001110111111$, com número par de 0’s, resulta da concatenação de blocos de 1’s com palavras da linguagem $\mathcal{L}(01^\star 0) = \{00, 010, 0110, 01110, 011110, \ldots\}$, isto é, com palavras que começam e terminam em 0 e têm exatamente dois zeros.

$$1111 \quad \underbrace{011111110}_{\substack{\in \\ \mathcal{L}(01^\star 0)}} \quad 1 \quad \underbrace{010}_{\substack{\in \\ \mathcal{L}(01^\star 0)}} \quad \underbrace{00}_{\substack{\in \\ \mathcal{L}(01^\star 0)}} \quad 1 \quad \underbrace{00}_{\substack{\in \\ \mathcal{L}(01^\star 0)}} \quad \underbrace{01110}_{\substack{\in \\ \mathcal{L}(01^\star 0)}} \quad 111111$$

$$\boxed{\mathcal{L}((01^\star 0 + 1)^\star) = (\,\mathcal{L}(01^\star 0 + 1)\,)^\star = (\,\mathcal{L}(01^\star 0) \cup \mathcal{L}(1)\,)^\star = \{1, 00, 010, 0110, \ldots\}^\star}$$

Exemplos

- $(01^\star 0 + 1)^\star$ descreve a linguagem constituída pelas palavras que têm **número par de** 0**'s**.
- $(01^\star 0 + 1)^\star 01^\star$ descreve a linguagem formada pelas palavras que têm **número ímpar de** 0**'s**.

Ideia: Por exemplo, 1111011111110101000100011101111111, com número par de 0's, resulta da concatenação de blocos de 1's com palavras da linguagem $\mathcal{L}(01^\star 0) = \{00, 010, 0110, 01110, 011110, \ldots\}$, isto é, com palavras que começam e terminam em 0 e têm exatamente dois zeros.

$$1111\ \underbrace{011111110}_{\in\ \mathcal{L}(01^\star 0)}\ 1\ \underbrace{010}_{\in\ \mathcal{L}(01^\star 0)}\ \underbrace{00}_{\in\ \mathcal{L}(01^\star 0)}\ 1\ \underbrace{00}_{\in\ \mathcal{L}(01^\star 0)}\ \underbrace{01110}_{\in\ \mathcal{L}(01^\star 0)}\ 111111$$

$$\mathcal{L}((01^\star 0 + 1)^\star) = (\,\mathcal{L}(01^\star 0 + 1)\,)^\star = (\,\mathcal{L}(01^\star 0) \cup \mathcal{L}(1)\,)^\star = \{1, 00, 010, 0110, \ldots\}^\star$$

Exemplos

- $(01^\star 0 + 1)^\star$ descreve a linguagem constituída pelas palavras que têm **número par de** 0**'s**.
- $(01^\star 0 + 1)^\star 01^\star$ descreve a linguagem formada pelas palavras que têm **número ímpar de** 0**'s**.

Ideia: Por exemplo, 1111011111110101000100011101111111, com número par de 0's, resulta da concatenação de blocos de 1's com palavras da linguagem $\mathcal{L}(01^\star 0) = \{00, 010, 0110, 01110, 011110, \ldots\}$, isto é, com palavras que começam e terminam em 0 e têm exatamente dois zeros.

$$1111 \; \underbrace{011111110}_{\in \mathcal{L}(01^\star 0)} \; 1 \; \underbrace{010}_{\in \mathcal{L}(01^\star 0)} \; \underbrace{00}_{\in \mathcal{L}(01^\star 0)} \; 1 \; \underbrace{00}_{\in \mathcal{L}(01^\star 0)} \; \underbrace{01110}_{\in \mathcal{L}(01^\star 0)} \; 111111$$

$$\boxed{\mathcal{L}((01^\star 0 + 1)^\star) = (\, \mathcal{L}(01^\star 0 + 1)\,)^\star = (\, \mathcal{L}(01^\star 0) \cup \mathcal{L}(1)\,)^\star = \{1, 00, 010, 0110, \ldots\}^\star}$$

Equivalência de expressões regulares

Expressões regulares equivalentes

Sejam r e s expressões regulares sobre Σ. Dizemos que r e s são equivalentes, e escrevemos $r \equiv s$, se $\mathcal{L}(r) = \mathcal{L}(s)$, isto é, r e s definem a mesma linguagem.

Formalmente, $\equiv$ *é uma* **relação de equivalência** *definida no conjunto das expressões regulares (i.e., uma relação binária que é reflexiva, simétrica e transitiva).*

Exemplos

Vimos anteriormente alguns casos de equivalência de expressões regulares (abreviadas, i.e., com omissão de parentesis). Em particular

$$(0^\star + 1^\star)^\star \equiv (0+1)^\star$$

que é uma instância de uma propriedade mais geral

$$(r^\star + s^\star)^\star \equiv (r+s)^\star$$

Equivalência de expressões regulares

Expressões regulares equivalentes

Sejam r e s expressões regulares sobre Σ. Dizemos que r e s são equivalentes, e escrevemos $r \equiv s$, se $\mathcal{L}(r) = \mathcal{L}(s)$, isto é, r e s definem a mesma linguagem.

Formalmente, $\equiv$ *é uma* **relação de equivalência** *definida no conjunto das expressões regulares (i.e., uma relação binária que é reflexiva, simétrica e transitiva).*

Exemplos

Vimos anteriormente alguns casos de equivalência de expressões regulares (abreviadas, i.e., com omissão de parentesis). Em particular

$$(0^\star + 1^\star)^\star \equiv (0 + 1)^\star$$

que é uma instância de uma propriedade mais geral

$$(r^\star + s^\star)^\star \equiv (r + s)^\star$$

Propriedades Algébricas

Quaisquer que sejam as expressões regulares r, s e t de alfabeto Σ tem-se:

$$r + s \equiv s + r \qquad (r+s)+t \equiv r+(s+t)$$

$$r(st) \equiv (rs)t \qquad r(s+t) \equiv rs + rt$$

$$(r+s)t \equiv rt + st \qquad \emptyset^\star \equiv \varepsilon$$

$$(r^\star)^\star \equiv r^\star \qquad (\varepsilon + r)^\star \equiv r^\star$$

$$(r^\star s^\star)^\star \equiv (r+s)^\star \qquad \varepsilon r \equiv r \equiv r\varepsilon$$

$$\emptyset r \equiv \emptyset \equiv r\emptyset$$

$$(r^\star + s^\star)^\star \equiv (r+s)^\star$$

NB: para facilitar a leitura/escrita, usamos as precedências para omitir parentesis-

Propriedades Algébricas

$$(r^\star + s^\star)^\star \equiv (r+s)^\star$$

Justificação:

Para concluir que $\mathcal{L}((r^\star + s^\star)^\star) = \mathcal{L}((r+s)^\star)$ notamos que:

- Por definição de fecho de Kleene, as palavras de $\mathcal{L}((r^\star + s^\star)^\star)$ são sequências finitas de palavras de $\mathcal{L}(r^\star)$ ou $\mathcal{L}(s^\star)$. Além de ε, podem ser da forma $x_1 \ldots x_k$, com $k \geq 1$, e $x_i \in L(r^\star) \cup L(s^\star)$, para $1 \leq i \leq k$.
- Também por definição de fecho de Kleene, cada um dos x_i é uma sequência finita de palavras de $L(r)$ ou é uma sequência finita de palavras de $L(s)$.
- Portanto, se $y \in \mathcal{L}((r^\star + s^\star)^\star)$, então y é uma sequência finita de palavras de $\mathcal{L}(r) \cup \mathcal{L}(s)$, o que implica que $\mathcal{L}((r^\star + s^\star)^\star) \subseteq \mathcal{L}((r+s)^\star)$.
- Reciprocamente, se $z \in \mathcal{L}((r+s)^\star)$ e $z \neq \varepsilon$ então $z = z_1 \ldots z_p$, com $p \geq 1$ e $z_i \in \mathcal{L}(r) \cup \mathcal{L}(s)$, para todo i. Logo, z_i pertence a $\mathcal{L}(r)^\star \cup \mathcal{L}(s)^\star$, pois $\mathcal{L}(r) \subseteq \mathcal{L}(r)^\star$ e $\mathcal{L}(s) \subseteq \mathcal{L}(s)^\star$. Portanto, $\mathcal{L}((r+s)^\star) \subseteq \mathcal{L}((r^\star + s^\star)^\star)$. □

Propriedades Algébricas

$$(r^\star + s^\star)^\star \equiv (r + s)^\star$$

Justificação:

Para concluir que $\mathcal{L}((r^\star + s^\star)^\star) = \mathcal{L}((r + s)^\star)$ notamos que:

- Por definição de fecho de Kleene, as palavras de $\mathcal{L}((r^\star + s^\star)^\star)$ são sequências finitas de palavras de $\mathcal{L}(r^\star)$ ou $\mathcal{L}(s^\star)$. Além de ε, podem ser da forma $x_1 \ldots x_k$, com $k \geq 1$, e $x_i \in L(r^\star) \cup L(s^\star)$, para $1 \leq i \leq k$.
- Também por definição de fecho de Kleene, cada um dos x_i é uma sequência finita de palavras de $L(r)$ ou é uma sequência finita de palavras de $L(s)$.
- Portanto, se $y \in \mathcal{L}((r^\star + s^\star)^\star)$, então y é uma sequência finita de palavras de $\mathcal{L}(r) \cup \mathcal{L}(s)$, o que implica que $\mathcal{L}((r^\star + s^\star)^\star) \subseteq \mathcal{L}((r + s)^\star)$.
- Reciprocamente, se $z \in \mathcal{L}((r + s)^\star)$ e $z \neq \varepsilon$ então $z = z_1 \ldots z_p$, com $p \geq 1$ e $z_i \in \mathcal{L}(r) \cup \mathcal{L}(s)$, para todo i. Logo, z_i pertence a $\mathcal{L}(r)^\star \cup \mathcal{L}(s)^\star$, pois $\mathcal{L}(r) \subseteq \mathcal{L}(r)^\star$ e $\mathcal{L}(s) \subseteq \mathcal{L}(s)^\star$. Portanto, $\mathcal{L}((r + s)^\star) \subseteq \mathcal{L}((r^\star + s^\star)^\star)$. $\square$

Propriedades Algébricas

$$(r^\star + s^\star)^\star \equiv (r+s)^\star$$

Justificação:

Para concluir que $\mathcal{L}((r^\star + s^\star)^\star) = \mathcal{L}((r+s)^\star)$ notamos que:

- Por definição de fecho de Kleene, as palavras de $\mathcal{L}((r^\star + s^\star)^\star)$ são sequências finitas de palavras de $\mathcal{L}(r^\star)$ ou $\mathcal{L}(s^\star)$. Além de ε, podem ser da forma $x_1 \ldots x_k$, com $k \geq 1$, e $x_i \in L(r^\star) \cup L(s^\star)$, para $1 \leq i \leq k$.
- Também por definição de fecho de Kleene, cada um dos x_i é uma sequência finita de palavras de $L(r)$ ou é uma sequência finita de palavras de $L(s)$.
- Portanto, se $y \in \mathcal{L}((r^\star + s^\star)^\star)$, então y é uma sequência finita de palavras de $\mathcal{L}(r) \cup \mathcal{L}(s)$, o que implica que $\mathcal{L}((r^\star + s^\star)^\star) \subseteq \mathcal{L}((r+s)^\star)$.
- Reciprocamente, se $z \in \mathcal{L}((r+s)^\star)$ e $z \neq \varepsilon$ então $z = z_1 \ldots z_p$, com $p \geq 1$ e $z_i \in \mathcal{L}(r) \cup \mathcal{L}(s)$, para todo i. Logo, z_i pertence a $\mathcal{L}(r)^\star \cup \mathcal{L}(s)^\star$, pois $\mathcal{L}(r) \subseteq \mathcal{L}(r)^\star$ e $\mathcal{L}(s) \subseteq \mathcal{L}(s)^\star$. Portanto, $\mathcal{L}((r+s)^\star) \subseteq \mathcal{L}((r^\star + s^\star)^\star)$. $\square$

Propriedades Algébricas

$$(r^\star + s^\star)^\star \equiv (r + s)^\star$$

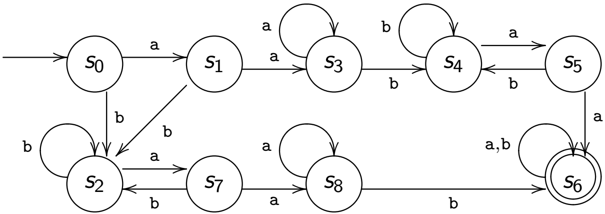
Justificação:

Para concluir que $\mathcal{L}((r^\star + s^\star)^\star) = \mathcal{L}((r + s)^\star)$ notamos que:

- Por definição de fecho de Kleene, as palavras de $\mathcal{L}((r^\star + s^\star)^\star)$ são sequências finitas de palavras de $\mathcal{L}(r^\star)$ ou $\mathcal{L}(s^\star)$. Além de ε, podem ser da forma $x_1 \ldots x_k$, com $k \geq 1$, e $x_i \in L(r^\star) \cup L(s^\star)$, para $1 \leq i \leq k$.
- Também por definição de fecho de Kleene, cada um dos x_i é uma sequência finita de palavras de $L(r)$ ou é uma sequência finita de palavras de $L(s)$.
- Portanto, se $y \in \mathcal{L}((r^\star + s^\star)^\star)$, então y é uma sequência finita de palavras de $\mathcal{L}(r) \cup \mathcal{L}(s)$, o que implica que $\mathcal{L}((r^\star + s^\star)^\star) \subseteq \mathcal{L}((r + s)^\star)$.
- Reciprocamente, se $z \in \mathcal{L}((r + s)^\star)$ e $z \neq \varepsilon$ então $z = z_1 \ldots z_p$, com $p \geq 1$ e $z_i \in \mathcal{L}(r) \cup \mathcal{L}(s)$, para todo i. Logo, z_i pertence a $\mathcal{L}(r)^\star \cup \mathcal{L}(s)^\star$, pois $\mathcal{L}(r) \subseteq \mathcal{L}(r)^\star$ e $\mathcal{L}(s) \subseteq \mathcal{L}(s)^\star$. Portanto, $\mathcal{L}((r + s)^\star) \subseteq \mathcal{L}((r^\star + s^\star)^\star)$. $\square$

Exemplo AFD - exercício 1f) - Folha prática 1

Linguagem das palavras de $\{\mathtt{a},\mathtt{b}\}^\star$ que **têm** aab **e** baa **como subpalavra**.



s_0: apenas ε

s_1: apenas a

s_3: palavras da forma $\mathtt{a}^k$, com $k \geq 2$

s_4: termina em b e tem aab como subpalavra mas não baa.

s_5: termina em ba e tem aab como subpalavra mas não baa.

s_6: tem aab e baa como subpalavra

s_2: não tem aab nem baa como subpalavra e termina em b

s_7: não tem aab nem baa como subpalavra e termina em ba

s_8: tem baa como subpalavra mas não aab e termina em aa