Formální jazyky a překladače

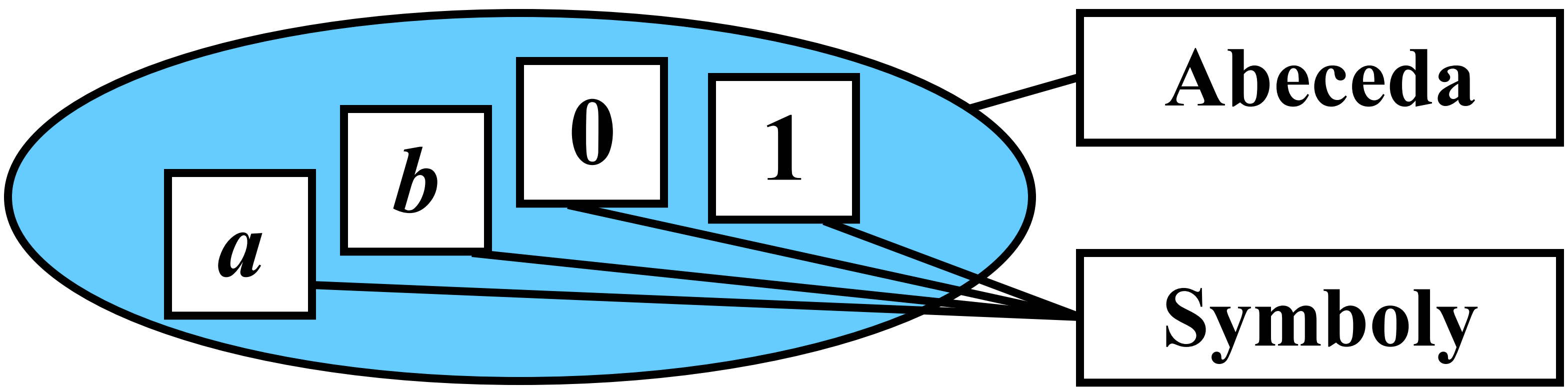
Alexander Meduna
&
Roman Lukáš

Kapitola I. Abecedy, řetězce a jazyky

Abecedy a symboly

Definice: *Abeceda* je konečná, neprázdná množina elementů, které nazýváme *symboly*.

Příklad:



Pokud označíme abecedu $\mathbf{\Sigma}$, potom $\mathbf{\Sigma} = \{\boldsymbol{a}, \boldsymbol{b}, \mathbf{0}, \mathbf{1}\}$

Řetězec

Myšlenka: $x = a_1a_2\ldots a_n$

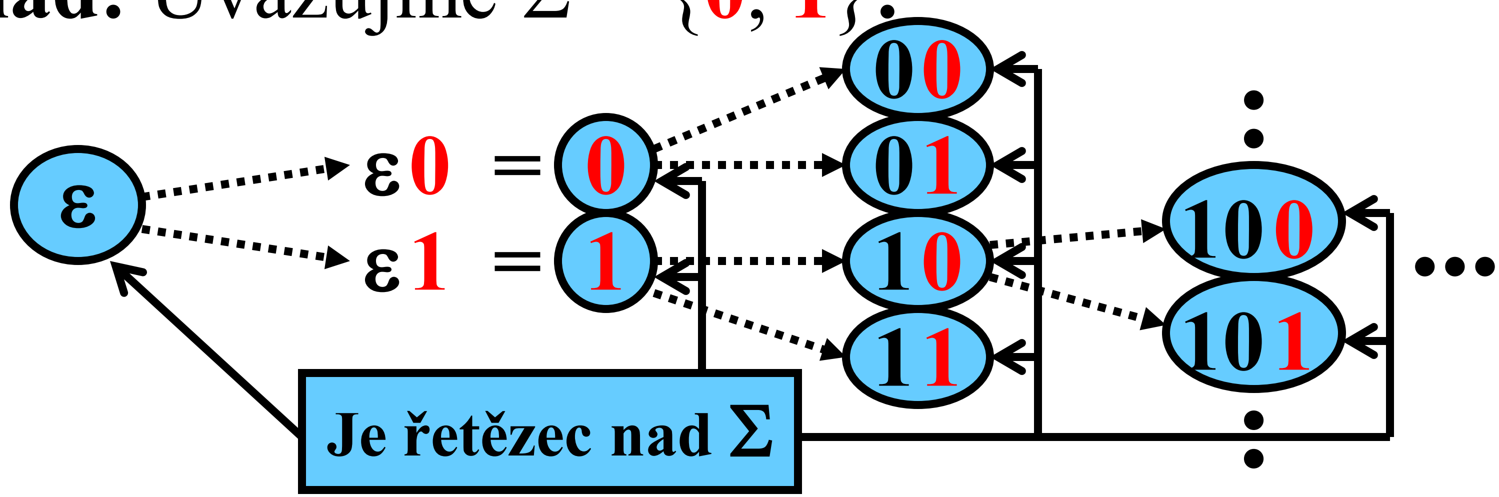
Definice: Nechť Σ je abeceda.

1) ε je řetězec nad abecedou Σ

2) pokud x je řetězec nad Σ a $a \in \Sigma$, potom xa je řetězec nad abecedou Σ

Pozn.: **ε** značí tzv. *prázdný řetězec* = neobsahuje žádný symbol.

Příklad: Uvažujme Σ = {0, 1}:



Délka řetězce

Myšlenka: $|a_1a_2\ldots a_n| = n$

Definice: Nechť x je řetězec nad abecedou Σ.
Délka řetězce x, $|x|$, je definována:

1) pokud $x = \varepsilon$, pak $|x| = 0$

2) pokud $x = a_1\ldots a_n$, pak $|x| = n$
pro $n \geq 1$ a $a_i \in \Sigma$ pro všechna $i = 1,\ldots,n$

Pozn.: Délka řetězce x je celkový počet symbolů v řetězci x.

Příklad: Uvažujme x =1010

Určeme: $|x|$

$$\boldsymbol{x = 1\ 0\ 1\ 0}$$
$$a_1a_2a_3a_4 \rightarrow n = 4, \text{ tedy } |\boldsymbol{x}| = \mathbf{4}$$

Konkatenace (zřetězení) řetězců

Myšlenka: ***xy***

Definice: Nechť x a y jsou dva řetězce nad abecedou Σ. *Konkatenace* x a y je řetězec xy.

Pozn.: $x\varepsilon = \varepsilon x = x$

Příklady:

Konkatenace **101** a **001** je řetězec **101001**

Konkatenace **ε** a **001** je řetězec **ε001** = **001**

Mocnina řetězce

Myšlenka: $x^i = \underbrace{xx\ldots x}_{i\text{-krát}}$

Definice: Nechť x je řetězec nad abecedou Σ.
Pro $i \geq 0$, i-tá *mocnina* řetězce x, x^i, je definována:
1) $x^0 = \varepsilon$ **2)** pro $i \geq 1$: $x^i = xx^{i-1}$

Pozn.: $x^i x^j = x^j x^i = x^{i+j}$, kde $i, j \geq 0$

Příklad: Uvažujme $x = 10$

Určeme: x^3

$x^3 = xx^2 = 10x^2$

$x^2 = xx^1 = 10x^1$

$x^1 = xx^0 = 10x^0$

$x^0 = \varepsilon$

$x^1 = 10\varepsilon = 10$

$x^2 = 1010$

$\boxed{x^3 = 101010}$

Reverzace řetězce

Myšlenka: reversal($a_1 \ldots a_n$) = $a_n \ldots a_1$

Definice: Nechť x je řetězec nad abecedou Σ.
Reverzace řetězce x, *reversal*(x), je definována:
1) pokud $x = \varepsilon$ pak reversal(ε) = ε
2) pokud $x = a_1 \ldots a_n$ pak reversal($a_1 \ldots a_n$) = $a_n \ldots a_1$
pro $n \geq 1$ a $a_i \in \Sigma$ pro všechna $i = 1,\ldots,n$

Příklad: Uvažujme x =1010

Určeme: reversal(x)

reversal($a_1a_2a_3a_4$) = $a_4a_3a_2a_1$, tedy

reversal(1 0 1 0) = 0 1 0 1

Prefix řetězce

Myšlenka: *x* je prefixem řetězce *xz*

Definice: Nechť x a y jsou dva řetězce nad abecedou Σ; x je *prefixem* y, pokud existuje řetězec z nad abecedou Σ, přičemž platí $\boldsymbol{xz = y}$.

Pozn.: pokud $x \notin \{\varepsilon, y\}$ pak x je ***vlastní prefix*** řetězce y.

Příklad: Uvažujme řetězec 1010

Určeme: Všechny prefixy **1010**

Prefixy 1010

- **ε**
- **1**
- **10**
- **101**
- **1010**

Vlastní prefixy 1010: **1**, **10**, **101**

Sufix řetězce

Myšlenka: *x* je sufix řetězce *zx*

Definice: Nechť x a y jsou dva řetězce nad abecedou Σ; x je *sufixem* y, pokud existuje řetězec z nad abecedou Σ, přičemž platí $zx = y$.

Pozn.: Pokud $x \notin \{\varepsilon, y\}$ pak x je ***vlastním sufix*** řetězce y.

Příklad: Uvažujme řetězec 1010

Určeme: Všechny sufixy **1010**

Sufixy 1010

ε
0
10
010
1010

Vlastní sufixy 1010

Podřetězec

Myšlenka: *x* je podřetězec řetězce *zxz'*

Definice: Nechť x a y jsou dva řetězce nad abecedou Σ. x je *podřetězec* y, pokud existují řetězce z, z' nad abecedou Σ přičemž platí $\boldsymbol{zxz' = y}$.

Pozn: Pokud $x \notin \{\varepsilon, y\}$, pak x je ***vlastní podřetězec*** řetězce y.

Příklad: Uvažujme řetězec 1010

Určeme: Všechny podřetězce **1 0 1 0**

Podřetězce 1010

ε

1, 0

10, 01

101, 010

1010

Vlastní podřetězce 1010

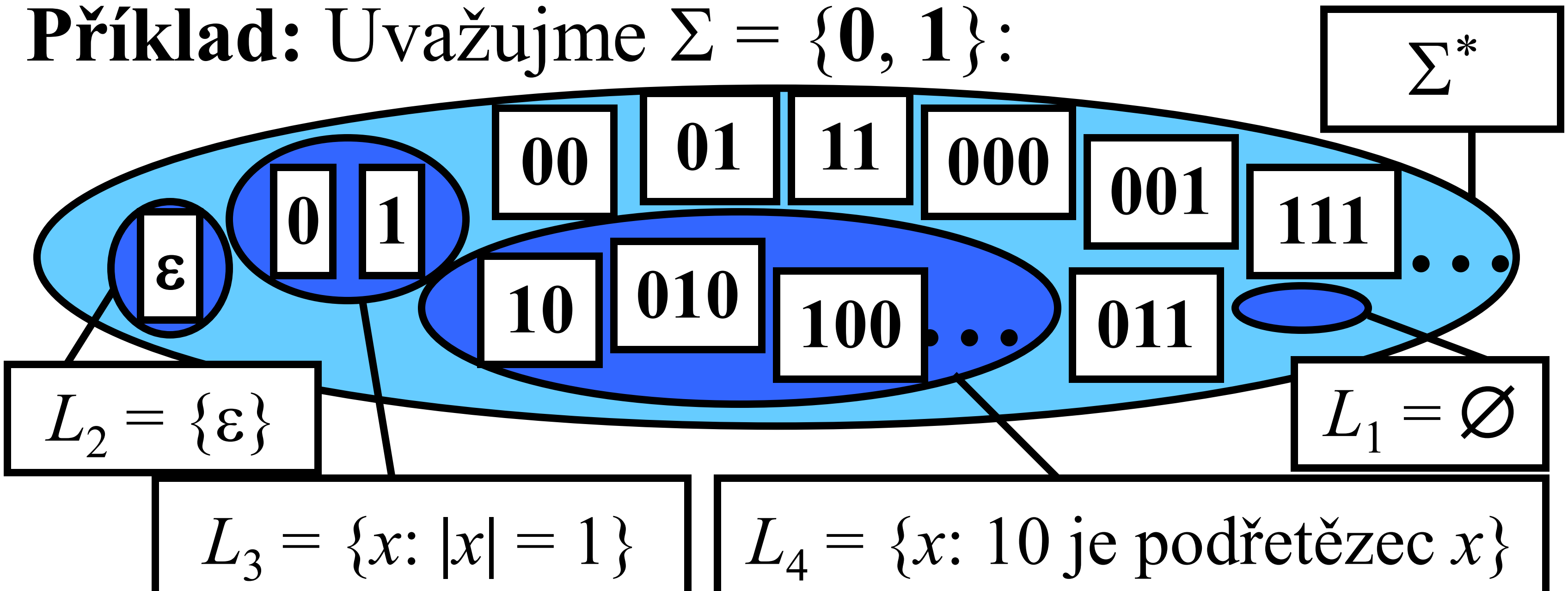
Jazyky

Myšlenka: $L \subseteq \Sigma^*$

Definice: Nechť Σ^* značí množinu všech řetězců nad Σ. Každá podmnožina $L \subseteq \Sigma^*$ je *jazyk* nad Σ.

Pozn.: Σ^+ znační množinu $\Sigma^* - \{\varepsilon\}$.

Příklad: Uvažujme $\Sigma = \{\mathbf{0}, \mathbf{1}\}$:



L_1, L_2, L_3, L_4 jsou jazyky nad Σ

Konečné a nekonečné jazyky

Myšlenka: Konečný jazyk obsahuje konečný počet řetězců

Definice: Jazyk L je *konečný*, pokud L obsahuje konečný počet řetězců, jinak je *nekonečný*.

Pozn.: Nechť S je množina; card(S) značí počet prvků v S

Příklad:

- $L_1 = \emptyset$ je **konečný jazyk**, protože card(L_1) = **0**
- $L_2 = \{\varepsilon\}$ je **konečný jazyk**, protože card(L_2) = **1**
- $L_3 = \{x: |x| = 1\} = \{0, 1\}$ je **konečný jazyk**, protože card(L_3) = **2**
- $L_4 = \{x$: 10 je podřetězec $x\} = \{10, 010, 100, \dots\}$ je **nekonečný jazyk**

Sjednocení jazyků

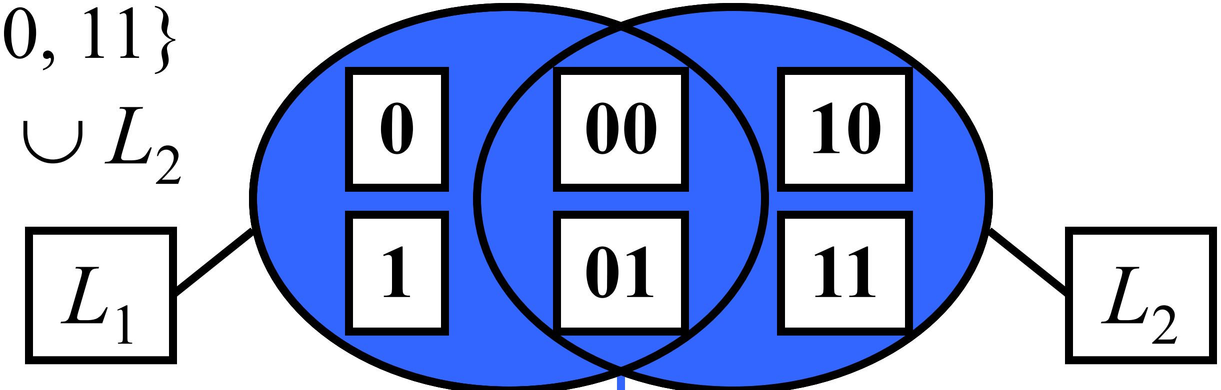
Myšlenka: Sjednocení L_1 a L_2 je $L_1 \cup L_2$

Definice: Nechť L_1 a L_2 jsou dva jazyky nad Σ.
Sjednocení jazyků L_1 a L_2, $L_1 \cup L_2$, je definováno:

$$L_1 \cup L_2 = \{x: x \in L_1 \textbf{ nebo } x \in L_2\}$$

Příklad: Uvažujme jazyky $L_1 = \{0, 1, 00, 01\}$, $L_2 = \{00, 01, 10, 11\}$

Určeme: $L_1 \cup L_2$



$$\boldsymbol{L_1 \cup L_2 = \{0, 1, 00, 01, 10, 11\}}$$

Průnik jazyků

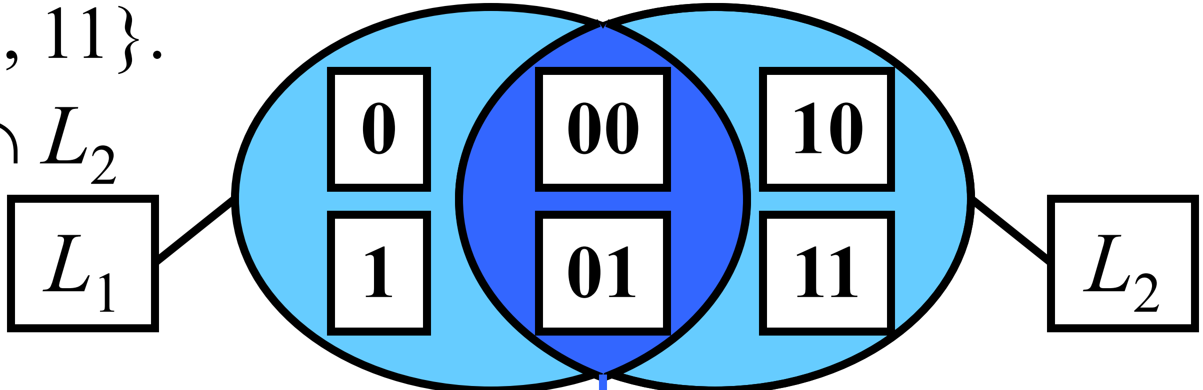
Myšlenka: Průnik L_1 a L_2 je $L_1 \cap L_2$

Definice: Nechť L_1 a L_2 jsou dva jazyky nad Σ.
Průnik jazyků L_1 a L_2, $L_1 \cap L_2$, je definován:

$$L_1 \cap L_2 = \{x: x \in L_1 \textbf{ a } x \in L_2\}$$

Příklad: Uvažujme jazyky $L_1 = \{0, 1, 00, 01\}$, $L_2 = \{00, 01, 10, 11\}$.

Určeme: $L_1 \cap L_2$



$$L_1 \cap L_2 = \{00, 01\}$$

Rozdíl jazyků

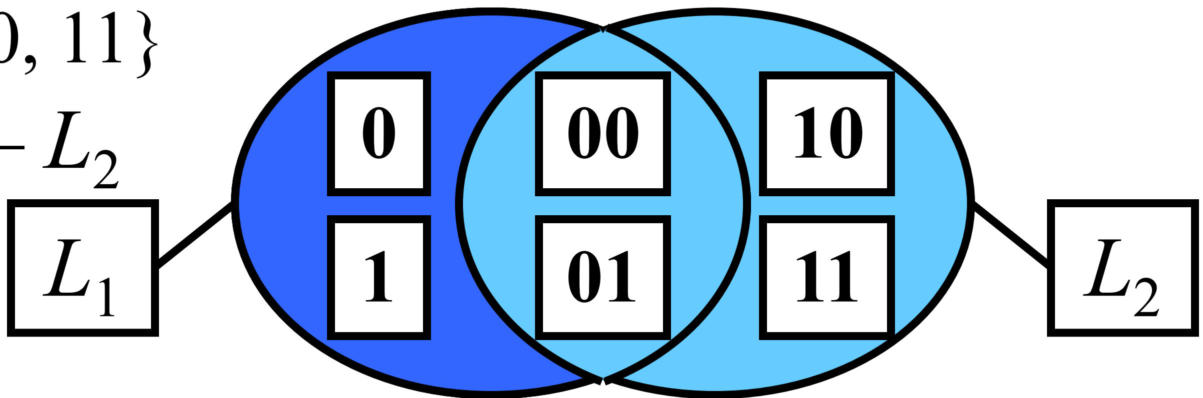
Myšlenka: Rozdíl jazyků L_1 a L_2 je $L_1 - L_2$

Definice: Nechť L_1 a L_2 jsou dva jazyky nad Σ.
Rozdíl jazyků L_1 a L_2, $L_1 - L_2$, je definován:

$$L_1 - L_2 = \{x: x \in L_1 \text{ a } x \notin L_2\}$$

Příklad: Uvažujme jazyky $L_1 = \{0, 1, 00, 01\}$, $L_2 = \{00, 01, 10, 11\}$

Určeme: $L_1 - L_2$



$L_1 - L_2 = \{0, 1\}$

Doplněk jazyka

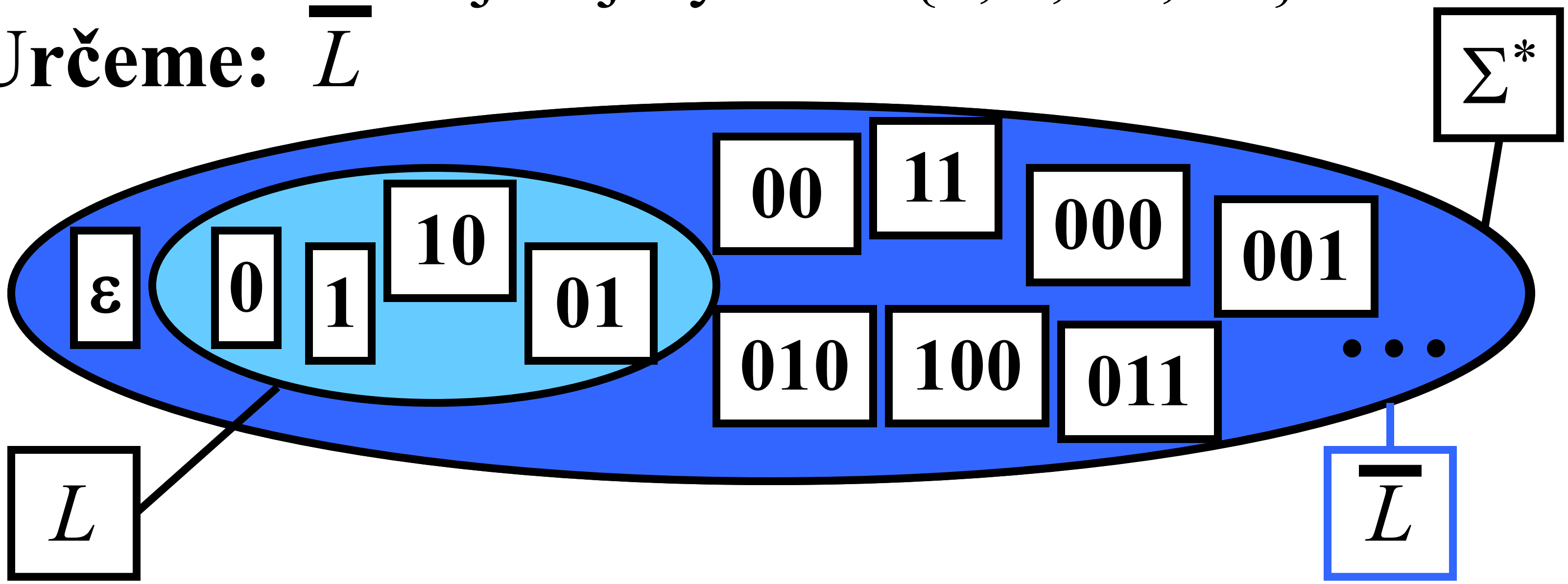
Myšlenka: $\overline{L} = \Sigma^* - L$

Definice: Nechť L je jazyk nad abecedou Σ.
Doplněk jazyka L, $\overline{L}$, je definován:

$$\overline{L} = \Sigma^* - L$$

Příklad: Uvažujme jazyk $L = \{0, 1, 01, 10\}$

Určeme: $\overline{L}$



Konkatenace jazyků

Myšlenka: $\boldsymbol{L_1L_2 = \{xy: x \in L_1 \text{ a } y \in L_2\}}$

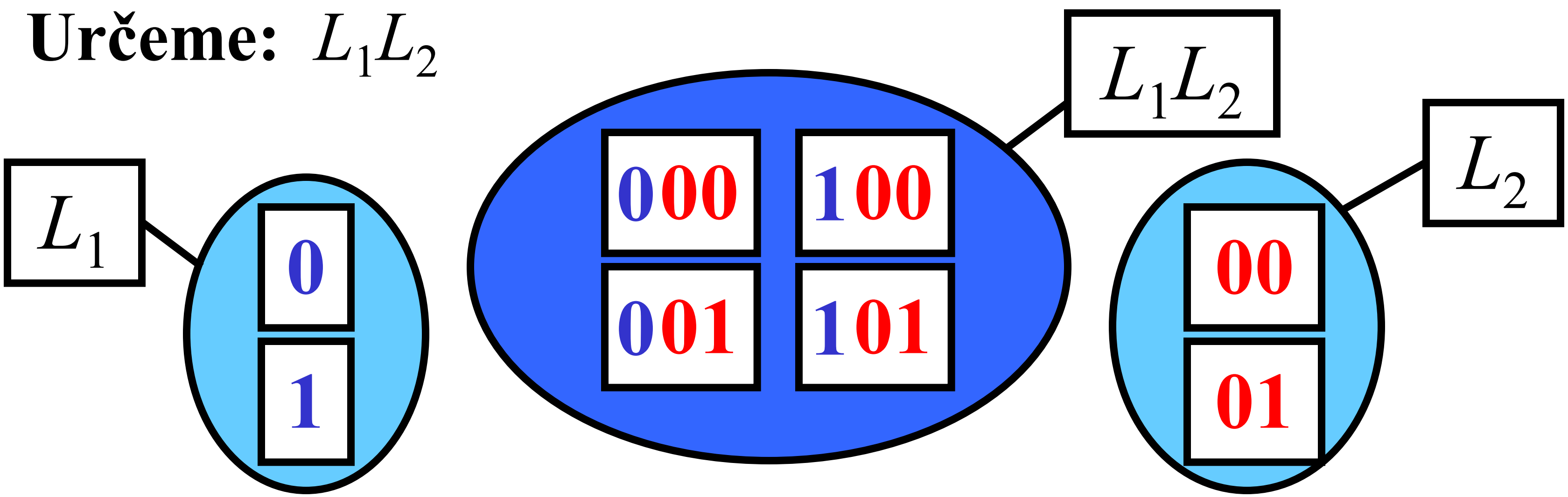
Definice: Nechť L_1 a L_2 jsou dva jazyky nad Σ. *Konkatenace jazyků* L_1 a L_2, L_1L_2 , je definována jako

$$L_1L_2 = \{xy: x \in L_1 \text{ a } y \in L_2\}$$

Pozn.: 1) $L\{\varepsilon\} = \{\varepsilon\}L = L$ **2)** $L\emptyset = \emptyset L = \emptyset$

Příklad: Uvažujme jazyky $L_1 = \{0, 1\}$, $L_2 = \{00, 01\}$

Určeme: L_1L_2



Reverzace jazyka

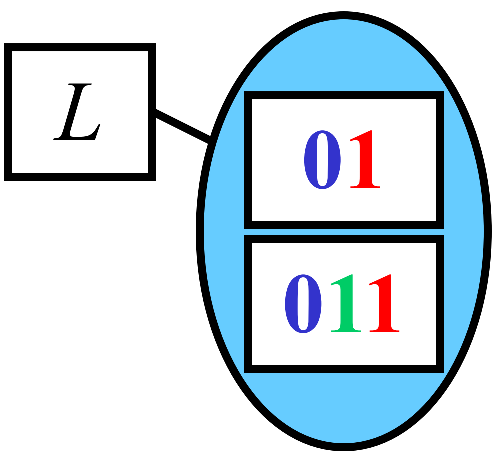
Myšlenka: $\boldsymbol{reverse(L) = \{reverse(x): x \in L\}}$

Definice: Nechť L je jazyk nad abecedou Σ.
Reverzace jazyka L, $reverse(L)$, je definována:

$$reverse(L) = \{reverse(x): x \in L\}$$

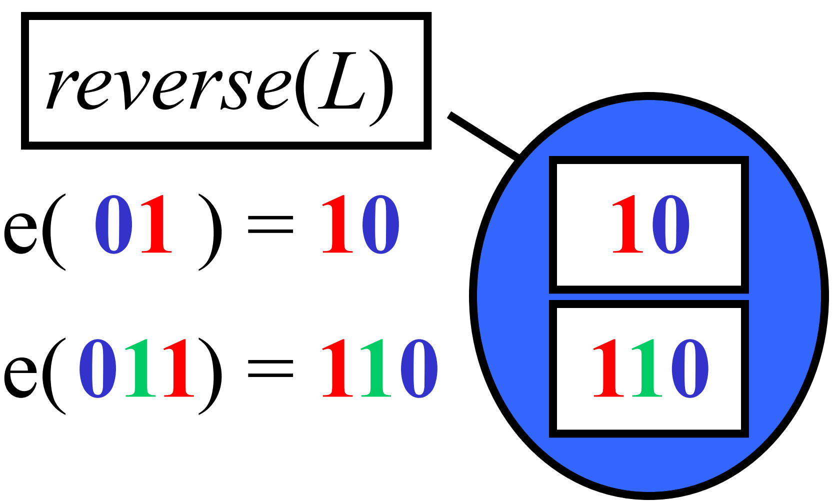
Příklad: Uvažujme $L = \{01, 011\}$

Určeme: $\boldsymbol{reversal}(L)$



reverse(01) = 10

reverse(011) = 110

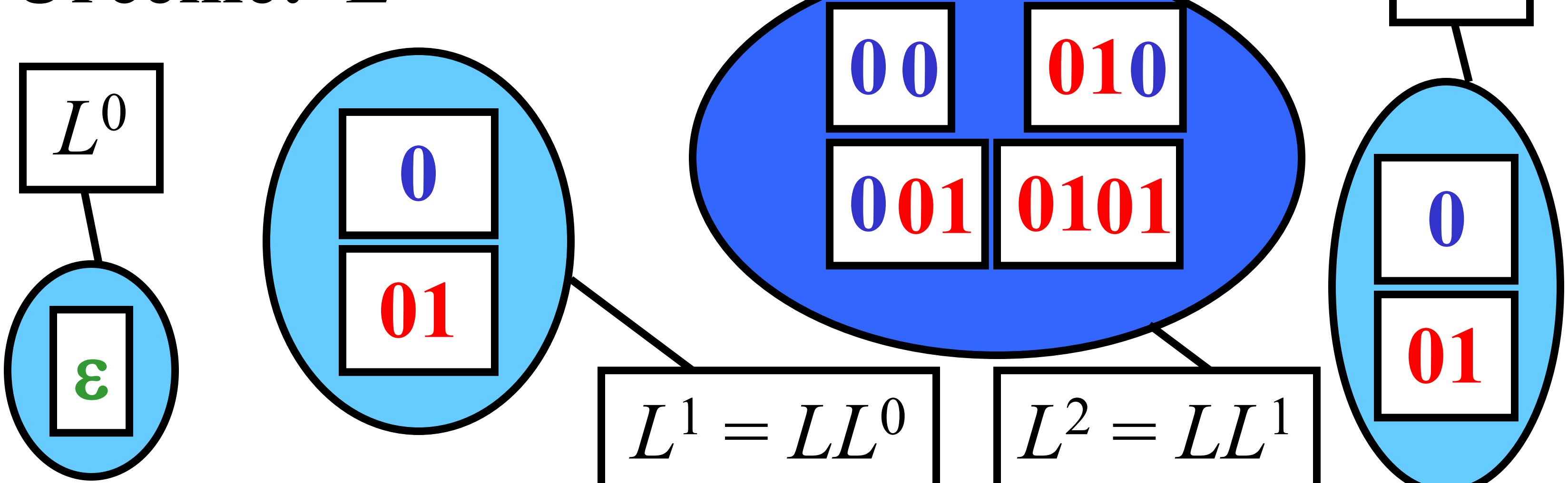


Mocnina jazyka

Myšlenka: $L^i = \underbrace{LL\ldots L}_{i\text{-krát}}$

Definice: Nechť L je jazyk nad abecedou Σ.
Pro $i \geq 0$, i-tá *mocnina jazyka* L, L^i, je definována:
1) $L^0 = \{\varepsilon\}$ **2)** pro $i \geq 1$: $L^i = LL^{i-1}$

Příklad: Uvažujme $L = \{0, 01\}$
Určeme: L^2



Iterace jazyka

Myšlenka: $L^* = L^0 \cup L^1 \cup L^2 \cup \ldots \cup L^i \cup \ldots$

$L^+ = L^1 \cup L^2 \cup \ldots \cup L^i \cup \ldots$

Definice: Nechť L je jazyk nad abecedou Σ. *Iterace jazyka* L, L^*, a *pozitivní iterace jazyka* L, L^+, jsou definovány $L^* = \bigcup_{i=0}^{\infty} L^i$, $L^+ = \bigcup_{i=1}^{\infty} L^i$

Pozn.: 1) $L^+ = LL^* = L^*L$ **2)** $L^* = L^+ \cup \{\varepsilon\}$

Příklad:

Uvažujme jazyk $L = \{0, 01\}$ nad $\Sigma = \{0, 1\}$.

Určeme: L^* **a** L^+

$L^0 = \{\varepsilon\}$, $L^1 = \{0, 01\}$, $L^2 = \{00, 001, 010, 0101\}$, ...

$L^* = L^0 \cup L^1 \cup L^2 \cup \ldots = \{\varepsilon, 0, 01, 00, 001, 010, 0101, \ldots\}$

$L^+ = L^1 \cup L^2 \cup \ldots = \{0, 01, 00, 001, 010, 0101, \ldots\}$