Formal Languages and Compilers

Alexander Meduna

- These lecture notes are based on *Automata and Languages* by Alexander Meduna, Springer, 2000

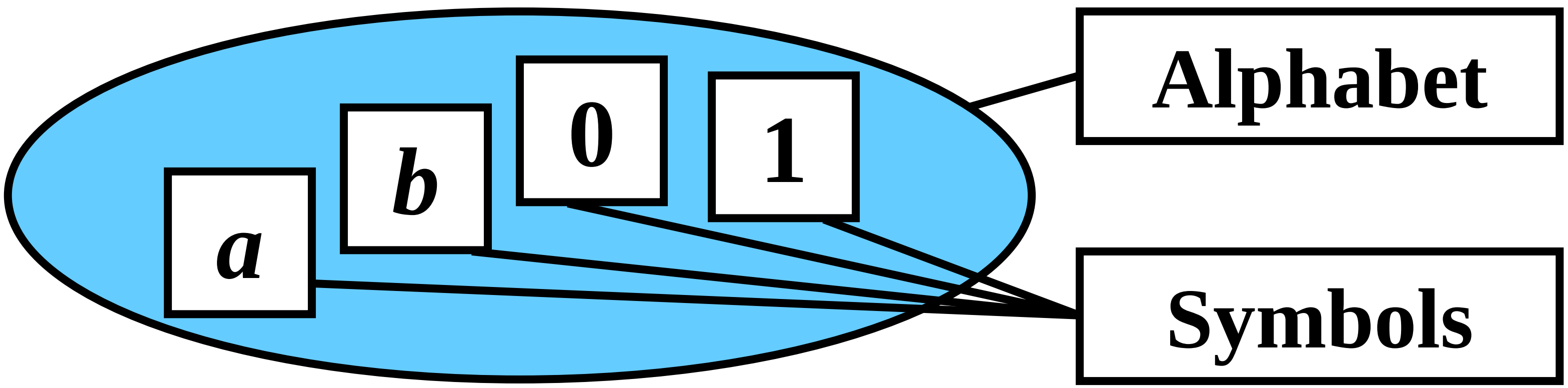
***Acknowledgement*: The author is indebted to Roman Lukáš for his great help during the preparation of these lecture notes.**

Part I.
Alphabets, Strings, and Languages

Alphabets and symbols

Definition: An *alphabet* is a finite, nonempty set of elements, which are called *symbols*.

Example:



If we denote this alphabet as Σ, then Σ **= {*a*, *b*, 0, 1}**

String

Gist: $x = a_1 a_2 \ldots a_n$

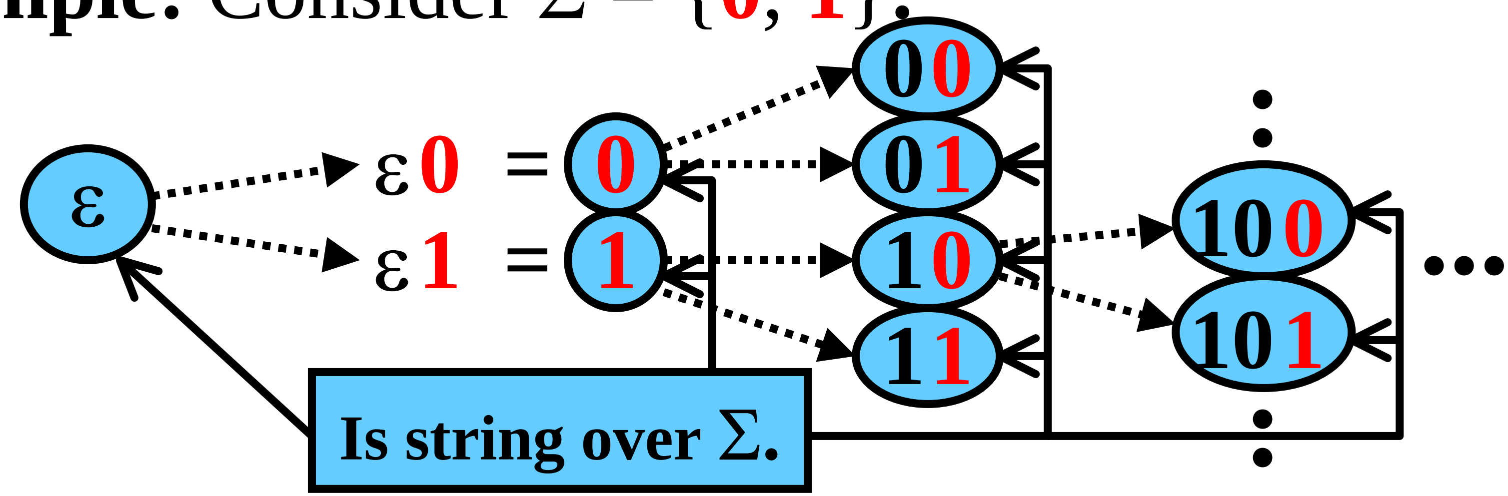
Definition: Let Σ be an alphabet.

1) ε is a string over Σ

2) if x is a string over Σ and $a \in \Sigma$ then xa is a string over Σ

Note: ε denotes ***the empty string*** that contains no symbols.

Example: Consider Σ = {**0**, **1**}:



Length of String

Gist: $|a_1a_2\ldots a_n| = n$

Definition: Let x be a string over Σ.
The *length* of x, $|x|$, is defined as follows:
1) if $x = \varepsilon$, then $|x| = 0$
2) if $x = a_1\ldots a_n$, then $|x| = n$
for some $n \geq 1$, and $a_i \in \Sigma$ for all $i = 1,\ldots,n$

Note: The length of x is the number of all symbols in x.

Example: Consider x =1010

Task: $|x|$

$$x = 1\ 0\ 1\ 0$$

$a_1a_2a_3a_4$ → $n = 4$, thus $|x| = 4$

Concatenation of Strings

Gist: ***xy***

Definition: Let x and y be two strings over Σ. The *concatenation* of x and y is xy.

Note: $x\varepsilon = \varepsilon x = x$

Examples:

Concatenation of **101** and **001** is **101001**

Concatenation of ε and **001** is ε**001** = **001**

Power of String

Gist: $x^i = \underbrace{xx\ldots x}_{i\text{-times}}$

Definition: Let x be a string over Σ.

For $i \geq 0$, the i-th *power* of x, x^i, is defined as

1) $x^0 = \varepsilon$ **2)** if $i \geq 1$ then $x^i = xx^{i-1}$

Note: $x^ix^j = x^jx^i = x^{i+j}$, where $i, j \geq 0$

Example: Consider $x = 10$

Task: x^3

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

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

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

$x^0 = \varepsilon$

$x^1 = 10\varepsilon = 10$

$x^2 = 1010$

$x^3 = 101010$

Reversal of String

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

Definition: Let x be a string over Σ.
The *reversal* of x, *reversal*(x), is defined as:

1) if $x = \varepsilon$ then reversal(ε) **=** ε

2) if $x = a_1 \ldots a_n$ then reversal($a_1 \ldots a_n$) = $a_n \ldots a_1$ for some $n \geq 1$, and $a_i \in \Sigma$ for all $i = 1,\ldots,n$

Example: Consider x =1010

Task: reversal(x)

reversal($\boldsymbol{a_1 a_2 a_3 a_4}$) = $\boldsymbol{a_4 a_3 a_2 a_1}$, so

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

Prefix of String

Gist: *x* is a prefix of *xz*

Definition: Let x and y be two strings over Σ; x is *prefix* of y if there is a string z over Σ so

$$xz = y$$

Note: if $x \notin \{\varepsilon, y\}$ then x is ***proper prefix*** of y.

Example: Consider 1010

Task: All prefixes of **1010**

Prefixes of 1010:
- ε
- **1**
- **10**
- **101**
- **1010**

Proper prefixes of 1010: **1**, **10**, **101**

Suffix of String

Gist: *x* is a suffix of *zx*

Definition: Let x and y be two strings over Σ; x is *suffix* of y if there is a string z over Σ so

$$zx = y$$

Note: if $x \notin \{\varepsilon, y\}$ then x is ***proper suffix*** of y.

Example: Consider 1010

Task: All suffixes of **1010**

Suffixes of 1010: ε, **0**, **10**, **010**, **1010**

Proper suffixes of 1010: **0**, **10**, **010**

Substring

Gist: *x* is a substring of *zxz'*

Definition: Let x and y be two strings over Σ; x is *substring* of y if there are two string z, z' over Σ so $\boldsymbol{zxz' = y}$.

Note: if $x \notin \{\varepsilon, y\}$ then x is ***proper substring*** of y.

Example: Consider 1010

Task: All substrings of **1 0 1 0**

Substrings of 1010:

- ε
- **1, 0**
- **10, 01**
- **101, 010**
- **1010**

Proper substrings of 1010:

- **1, 0**
- **10, 01**
- **101, 010**

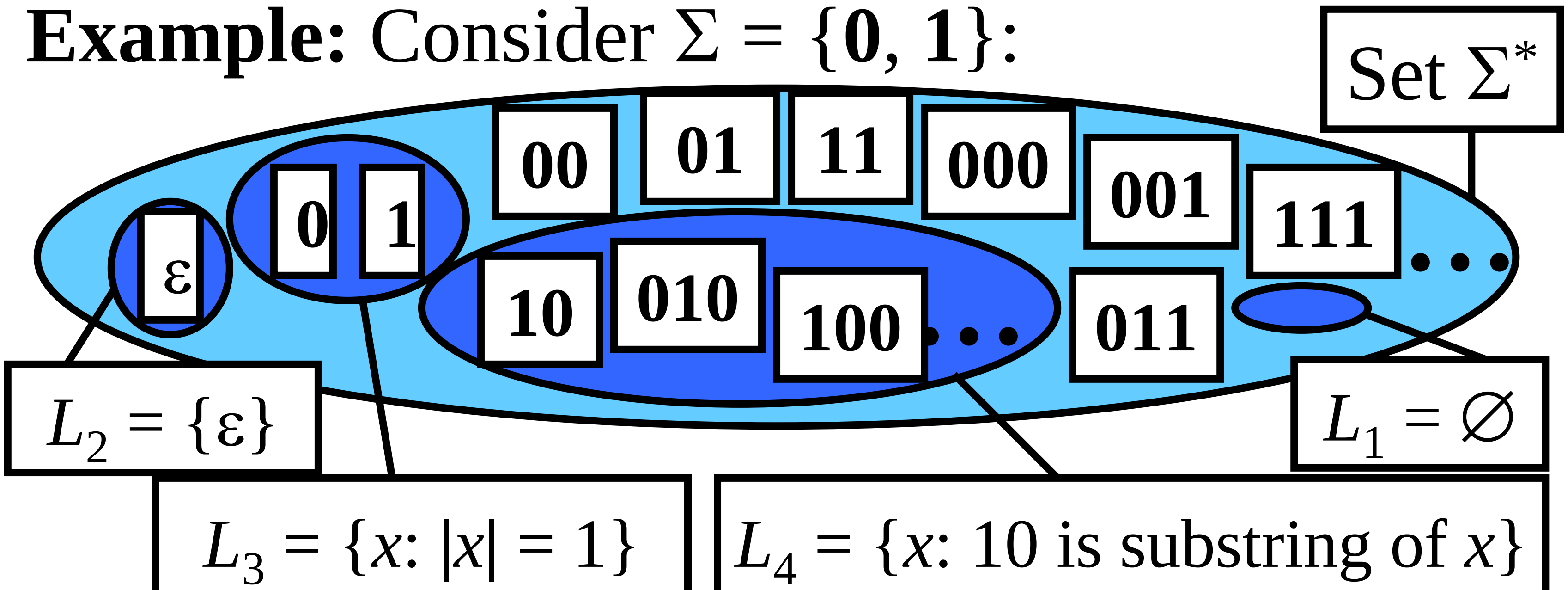
Languages

Gist: $L \subseteq \Sigma^*$

Definition: Let Σ^* denote the set of all strings over Σ. Every subset $L \subseteq \Sigma^*$ is a *language* over Σ.

Note: Σ^+ denote the set $\Sigma^* - \{\varepsilon\}$.

Example: Consider $\Sigma = \{\mathbf{0}, \mathbf{1}\}$:



L_1, L_2, L_3, L_4 are languages over Σ

Finite and Infinite Languages

Gist: finite language contains a finite number of strings

Definition: A language, L, is *finite* if L contains a finite number of strings; otherwise, L is *infinite*.

Note: Let S be a set; card(S) is the number of its members.

Examples:

- $L_1 = \emptyset$ is **finite** because card(L_1) = **0**
- $L_2 = \{\varepsilon\}$ is **finite** because card(L_2) = **1**
- $L_3 = \{x: |x| = 1\} = \{0, 1\}$ is **finite** because card(L_3) = **2**
- $L_4 = \{x$: 10 is substring of $x\} = \{10, 010, 100, \ldots\}$ is **infinite**

Union of Languages

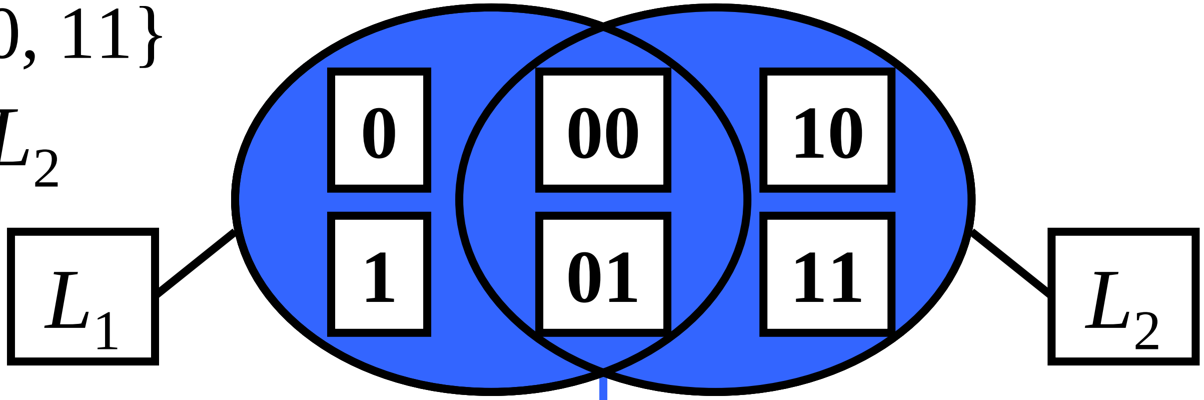
Gist: Union of L_1 and L_2 is $L_1 \cup L_2$

Definition: Let L_1 and L_2 be two languages over Σ.
The *union* of L_1 and L_2, $L_1 \cup L_2$, is defined as

$$L_1 \cup L_2 = \{x: x \in L_1 \text{ or } x \in L_2\}$$

Example: Consider languages $L_1 = \{0, 1, 00, 01\}$, $L_2 = \{00, 01, 10, 11\}$

Task: $L_1 \cup L_2$



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

Intersection of Languages

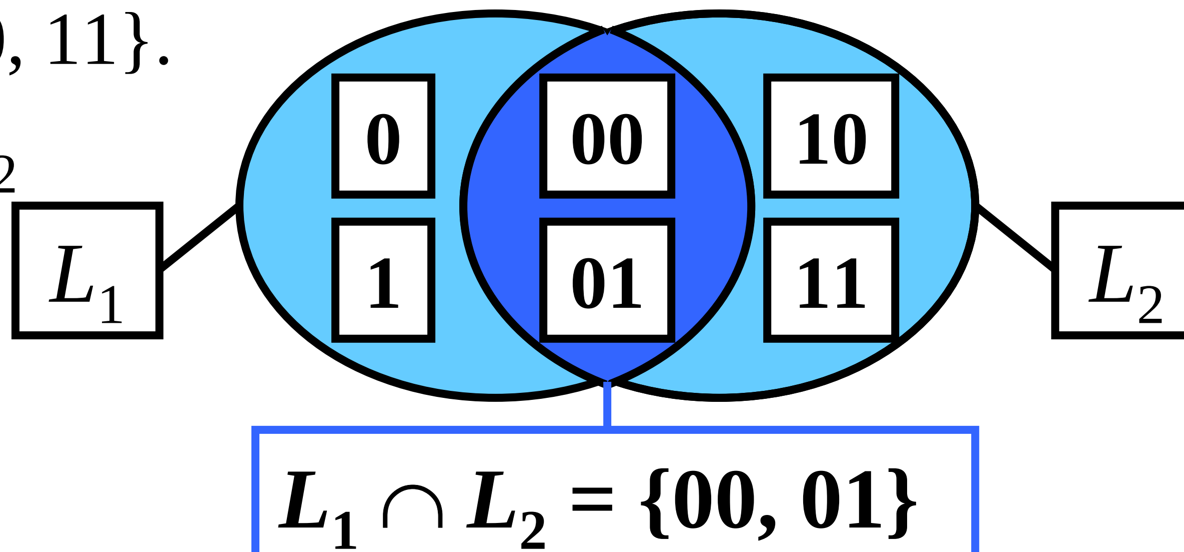
Gist: Intersection of L_1 and L_2 is $L_1 \cap L_2$

Definition: Let L_1 and L_2 be two languages over Σ. The *intersection* of L_1 and L_2, $L_1 \cap L_2$, is defined as:

$$L_1 \cap L_2 = \{x: x \in L_1 \text{ and } x \in L_2\}$$

Example: Consider languages $L_1 = \{0, 1, 00, 01\}$, $L_2 = \{00, 01, 10, 11\}$.

Task: $L_1 \cap L_2$



Difference of Languages

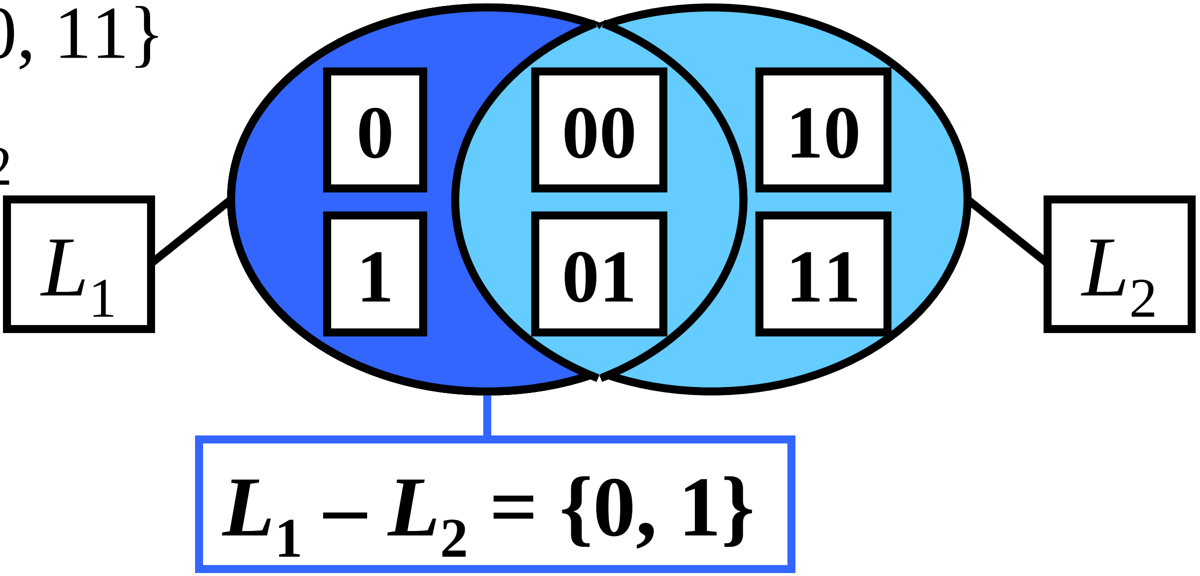
Gist: Difference of L_1 and L_2 is $L_1 - L_2$

Definition: Let L_1 and L_2 be two languages over Σ. The *difference* of L_1 and L_2, $L_1 - L_2$, is defined as

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

Example: Consider languages $L_1 = \{0, 1, 00, 01\}$, $L_2 = \{00, 01, 10, 11\}$

Task: $L_1 - L_2$



Complement of Language

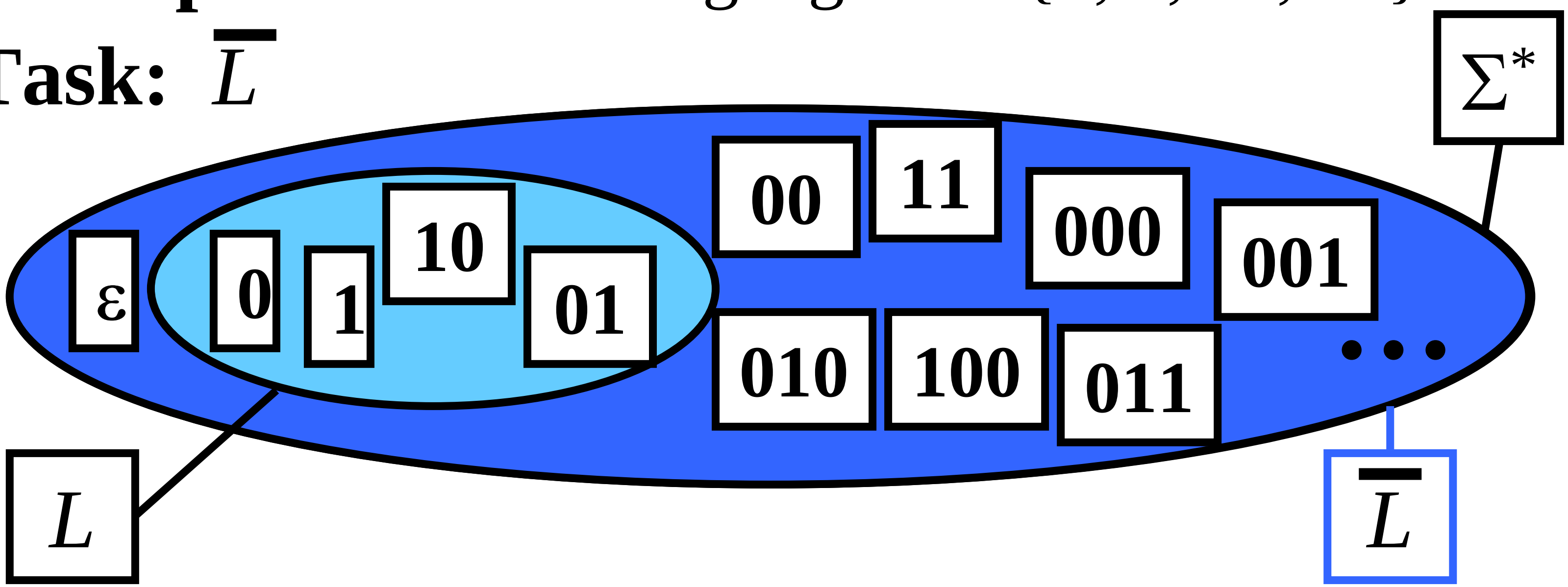
Gist: $\overline{L} = \Sigma^* - L$

Definition: Let L be a languages over Σ.
The *complement* of L, $\overline{L}$, is defined as

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

Example: Consider language L = {0, 1, 01, 10}

Task: $\overline{L}$



Concatenation of Languages

Gist: $L_1L_2 = \{xy: x \in L_1$ and $y \in L_2\}$

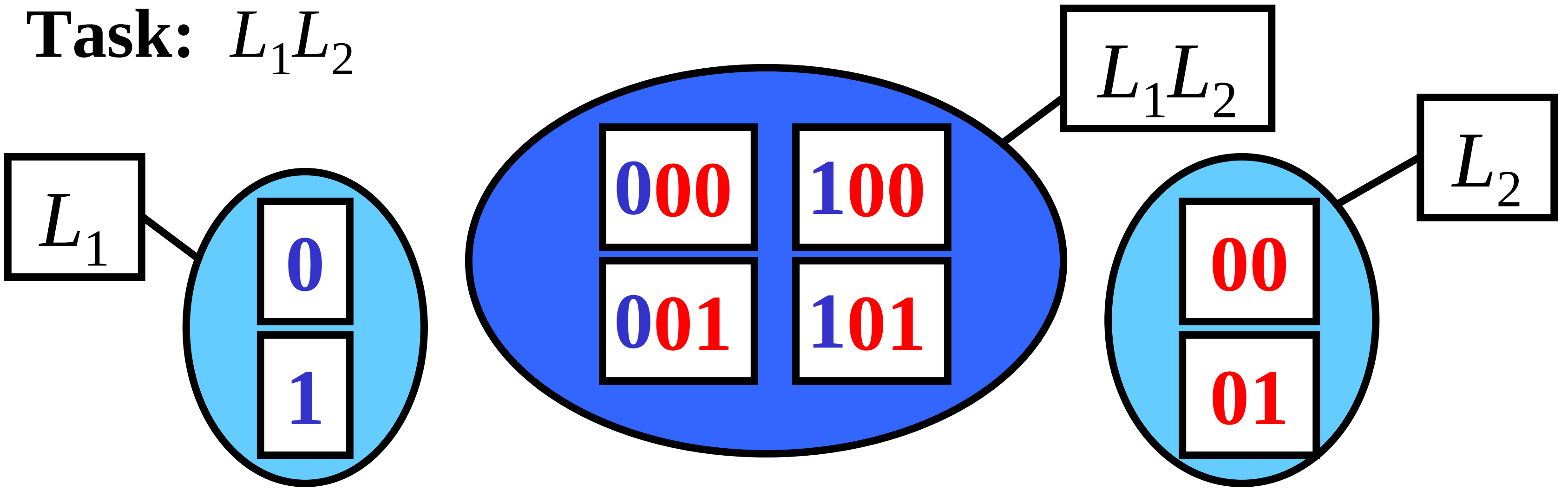
Definition: Let L_1 and L_2 be two languages over Σ. The *concatenation* of L_1 and L_2, L_1L_2, is defined as

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

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

Example: Consider languages $L_1 = \{0, 1\}$, $L_2 = \{00, 01\}$

Task: L_1L_2



Reversal of Language

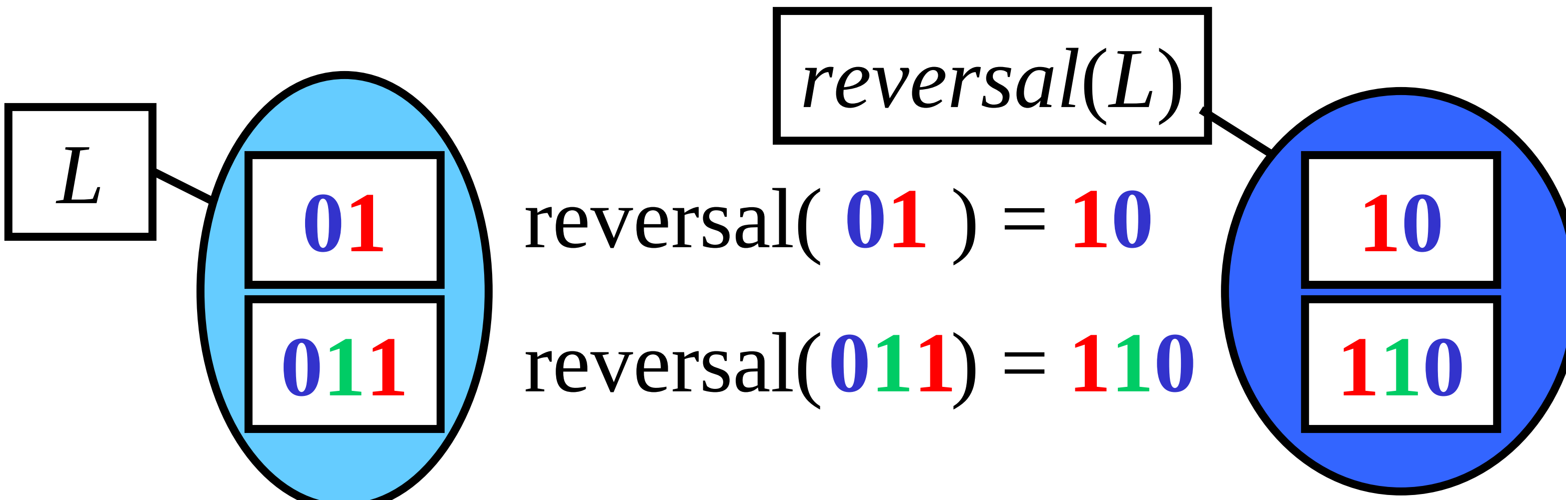
Gist: *reversal*(*L*) = {*reversal*(*x*): *x* ∈ *L*}

Definition: Let *L* be a language over Σ.
The *reversal* of *L*, *reversal*(*L*), is defined as

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

Example: Consider *L* = {01, 011}

Task: ***reversal***(*L*)



Power of Language

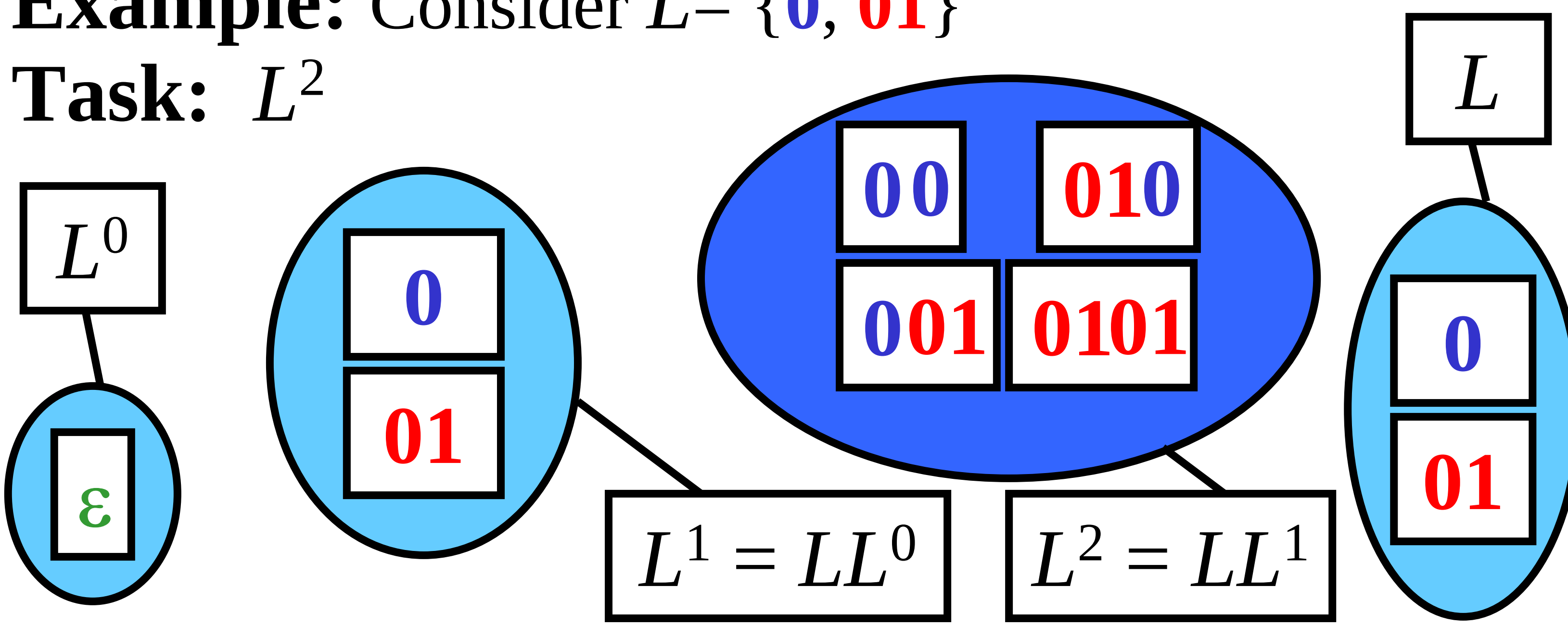
Gist: $L^i = \underbrace{LL\ldots L}_{i\text{-times}}$

Definition: Let L be a language over Σ.
For $i \geq 0$, the i-th *power* of L, L^i, is defined as:

1) $L^0 = \{\varepsilon\}$ **2)** if $i \geq 1$ then $L^i = LL^{i-1}$

Example: Consider $L = \{0, 01\}$

Task: L^2



Iteration of Language

Gist: $L^* = L^0 \cup L^1 \cup L^2 \cup \dots \cup L^i \cup \dots$

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

Definition: Let L be a language over Σ. The *iteration* of L, L^*, and the *positive iteration* of L, L^+, are defined as $L^* = \bigcup_{i=0}^{\infty} L^i$, $L^+ = \bigcup_{i=1}^{\infty} L^i$

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

Example:

Consider language $L = \{0, 01\}$ over $\Sigma = \{0, 1\}$.

Task: L^* **and** L^+

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

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

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