**Definition 26.1 (string)**  For any finite alphabet, $\Sigma$, a $\Sigma$-string is a finite sequence of characters from $\Sigma$. The **length** of a string, $s$, is the number of characters in $s$: $\text{length}(s) = |s|$. The **empty string**, $\epsilon$, is the unique string of length 0. $0 = |\epsilon|$.

**Example 26.2** For $\Sigma_{\text{bin}} = \{0, 1\}$ the following are binary strings, i.e., $\Sigma_{\text{bin}}$-strings:

<table>
<thead>
<tr>
<th>$s$</th>
<th>$\epsilon$</th>
<th>0</th>
<th>1</th>
<th>00</th>
<th>01</th>
<th>10</th>
<th>11</th>
<th>000</th>
<th>001</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>s</td>
<td>$</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

For $\Sigma_a = \{a\}$ the following are $\Sigma_a$-strings:

<table>
<thead>
<tr>
<th>$s$</th>
<th>$\epsilon$</th>
<th>$a$</th>
<th>$aa$</th>
<th>$aaa$</th>
<th>$a^4$</th>
<th>$a^5$</th>
<th>$a^6$</th>
<th>$a^7$</th>
<th>$a^8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>s</td>
<td>$</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

For $\Sigma_{abc} = \{a, b, c\}$ the following are $\Sigma_{abc}$-strings:

<table>
<thead>
<tr>
<th>$s$</th>
<th>$\epsilon$</th>
<th>$a$</th>
<th>$b$</th>
<th>$c$</th>
<th>$aa$</th>
<th>$ab$</th>
<th>$ac$</th>
<th>$ba$</th>
<th>$bb$</th>
<th>$bc$</th>
<th>$ca$</th>
<th>$cb$</th>
<th>$cc$</th>
<th>$aaa$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>s</td>
<td>$</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>
Definition 26.3 (Concatenation)  

The **concatenation** of two strings $a, b$ is written $a \cdot b$, or $ab$, and consists of $a$ immediately followed by $b$.

Example 26.4

$$a \cdot a = aa; \quad 00 \cdot 111 = 00111; \quad \epsilon \cdot s = s \cdot \epsilon = s$$

Definition 26.5 (Kleene Star)  

For $S$ a set of strings,

$$S^* \overset{\text{def}}{=} \bigcup_{i=0}^{\infty} S^i = S^0 \cup S^1 \cup S^2 \cup \ldots$$

- $S^0 \overset{\text{def}}{=} \{\epsilon\}$
- $S^1 \overset{\text{def}}{=} S$
- $S^2 \overset{\text{def}}{=} S \cdot S$
- $S^n \overset{\text{def}}{=} \underbrace{S \cdots \cdot S}_{n}$

$S^*$ is the **set of all strings** from $S$.

Example 26.6

- $\Sigma^*_{\text{bin}} = \{\epsilon, 0, 1, 00, 01, 10, 11, 000, 001, \ldots\}$
- $\Sigma^*_{ab} = \{\epsilon, a, b, aa, ab, ba, bb, aaa, aab, \ldots\}$
- $\Sigma^*_{abc} = \{\epsilon, a, b, c, aa, ab, ac, ba, bb, bc, ca, cb, cc, aaa, aab, \ldots\}$
- $\emptyset^* = \{\epsilon\}$
Definition 26.7  The set of regular expressions \( \text{regexp}(\Sigma) \) over alphabet \( \Sigma \) is the smallest set of strings such that:

1. **base cases:**
   
   (a) if \( a \in \Sigma \) then \( a \in \text{regexp}(\Sigma) \)
   
   (b) \( \emptyset \in \text{regexp}(\Sigma) \)

2. **inductive cases:** if \( e, f \in \text{regexp}(\Sigma) \) then so are the following:
   
   (a) \( (e \cup f) \)
   
   (b) \( (e \cdot f) \)
   
   (c) \( (e^*) \)

\[
\begin{align*}
e_0 &= \emptyset & \in & \text{regexp} \left( \{a, b\} \right) \\
e_1 &= a & \in & \text{regexp} \left( \{a, b\} \right) \\
e_2 &= a^* & \in & \text{regexp} \left( \{a, b\} \right) \\
e_3 &= ((a \cup b) \cdot (a \cup b))^* & \in & \text{regexp} \left( \{a, b\} \right) \\
e_4 &= a^* (ba^* ba^*)^* & \in & \text{regexp} \left( \{a, b, c\} \right) \\
e_5 &= \emptyset^* & \in & \text{regexp} \left( \{a, b\} \right)
\end{align*}
\]

**Meanings:** \( \mathcal{L}(e) \) is the language denoted by regular expression \( e \):

\[
\begin{align*}
\mathcal{L}(e_0) &= \emptyset \\
\mathcal{L}(e_1) &= \{a\} \\
\mathcal{L}(e_2) &= \{a\}^* \\
&= \{\epsilon, a, aa, a^3, a^4, \ldots\} \\
\mathcal{L}(e_3) &= \{a, b\}^{2*} = \{aa, ab, ba, bb\}^* \\
&= \{w \in \{a, b\}^* \mid |w| \equiv 0 \pmod{2}\} \\
\mathcal{L}(e_4) &= \{a\}^* \cdot (\{b\} \cdot \{a\}^* \cdot \{b\} \cdot \{a\}^*)^* \\
&= \{w \in \{a, b\}^* \mid \#a(w) \equiv 0 \pmod{2}\} \\
\mathcal{L}(e_5) &= \{\epsilon\}
\end{align*}
\]
Definition 26.8 (Meaning of Regular Expressions)

**base cases:** $a \in \Sigma$:  

$\mathcal{L}(a) \overset{\text{def}}{=} \{a\}$

$\mathcal{L}(\emptyset) \overset{\text{def}}{=} \emptyset$

**inductive cases:**  

$\mathcal{L}(e \cup f) \overset{\text{def}}{=} \mathcal{L}(e) \cup \mathcal{L}(f)$

$\mathcal{L}(e \cdot f) \overset{\text{def}}{=} \mathcal{L}(e) \cdot \mathcal{L}(f)$

$\mathcal{L}(e^*) \overset{\text{def}}{=} (\mathcal{L}(e))^*$
Precedence for Regular Operators

The order of precedence for the regular operators is the following:

1. \(*\), \(+\)
2. \(*\)
3. \(\cup\)

Example 26.9

\[ a \cup b \cdot c^* = a \cup (b \cdot (c^*)) \]

Definition 26.10 A set, \(A\), is regular iff there exists a regular expression that denotes it. In symbols, the set of regular sets over \(\Sigma\) is, \(\text{Regular}(\Sigma) = \{L(e) \mid e \in \text{regexp}(\Sigma)\}\).

Theorem 26.11 \(\text{Regular}(\Sigma)\) is the smallest set of languages that contains \(\emptyset, \{a\}, a \in \Sigma\) and is closed under \(\cup, \cdot, ^*\).
Clicker Question 26.1  What is $L(a^*ba^*)$?

A: \[ \{ w \in \{a,b\}^* \mid w \text{ contains at least one “b”} \} \]

B: \[ \{ w \in \{a,b\}^* \mid w \text{ contains exactly one “b”} \} \]
Clicker Question 26.2  What is $\mathcal{L}((0 \cup 1)^*001(0 \cup 1)^*)$?

A: $\{w \in \{0, 1\}^* \mid w \text{ contains the substring } 001\}$

B: $\{w \in \{0, 1\}^* \mid w \text{ does not contain the substring } 11\}$
Example 26.12

\[ L(0^+) = \{0, 0^2, 0^3 \ldots\} \]
\[ L((aa)^+) = \{aa, a^4, a^6, \ldots\} = \{ w \in \{a\}^* \mid |w| \equiv 0 \pmod{2} \land w \neq \epsilon\} \]
Clicker Question 26.3  What is $L(1^*(01^+)^*)$?

**A:** $\{w \in \{0, 1\}^* \mid w \text{ does not contain } 00 \text{ as a substring}\}$

**B:** $\{w \in \{0, 1\}^* \mid \text{Every } 0 \text{ in } w \text{ is immediately followed by a } 1\}$
Theorem 26.13  *Regular*$(\Sigma)$ is the *smallest set of languages* that contains $\emptyset, \{a\}, a \in \Sigma$ and is closed under $\cup, \cdot, \ast$.

**Proof:** Recall that $\text{Regular}(\Sigma) \overset{\text{def}}{=} \{L(e) \mid e \in \text{regexp}(\Sigma)\}$. Thus, $\text{Regular}(\Sigma)$ contains $\emptyset, \{a\}, a \in \Sigma$.

Let $L_0, L_1 \in \text{Regular}(\Sigma)$ be arbitrary. Therefore for some regular expressions, $e_0, e_1, L_0 = L(e_0)$ and $L_1 = L(e_1)$. Thus the following languages are also in $\text{Regular}(\Sigma)$,

\[
\begin{align*}
L_0 \cup L_1 &= L(e_0 \cup e_1) \\
L_0 \cdot L_1 &= L(e_0 \cdot e_1) \\
L_0^* &= L(e_0^*)
\end{align*}
\]

thus $\text{Regular}(\Sigma)$ is closed under the regular operations.

**Conversely,** suppose that $\mathcal{R}$ contains $\emptyset, \{a\}, a \in \Sigma$ and is closed under $\cup, \cdot, \ast$.

We must show that $\text{Regular}(\Sigma) \subseteq \mathcal{R}$.

We prove by induction on $e \in \text{regexp}(\Sigma)$ that $L(e) \in \mathcal{R}$.

**base case:** This holds because $\mathcal{R}$ contains $\emptyset, \{a\}, a \in \Sigma$.

**inductive case:** assume that $L(e_0), L(e_1) \in \mathcal{R}$.

It follows that

\[
\begin{align*}
L(e_0 \cup e_1) &= L(e_0) \cup L(e_1) \in \mathcal{R} \quad \text{because } \mathcal{R} \text{ is closed under } \cup \\
L(e_0 \cdot e_1) &= L(e_0) \cdot L(e_1) \in \mathcal{R} \quad \text{because } \mathcal{R} \text{ is closed under } \cdot \\
L(e_0^*) &= L(e_0)^* \in \mathcal{R} \quad \text{because } \mathcal{R} \text{ is closed under } \ast.
\end{align*}
\]

Thus, by induction, every language in $\text{Regular}(\Sigma)$ is in $\mathcal{R}$.  \(\square\)