Alphabets and Languages

CENG 280, 2022

Alphabets and Languages

Alphabets, languages

String operations

Language operations

Alphabets and Languages

In computational practice, data are encoded in the computer's memory as strings of bits or other symbols appropriate for manipulation by a computer. The mathematical study of the theory of computation must therefore begin by understanding the mathematics of strings of symbols..
by H.R. Lewis and C. H. Papadimitriou.

Alphabets and Languages

Alphabet: a finite set of symbols, e.g., $\{a, b, \ldots, z\}$, $\{0, 1\}$.

Alphabets and Languages

Alphabet: a finite set of symbols, e.g., $\{a, b, \ldots, z\}$, $\{0, 1\}$.

String: A string over an alphabet is a finite sequence of symbols from the alphabet, e.g.,"apple", "1011", "abc", "e" (empty string).

Alphabets and Languages

Alphabet: a finite set of symbols, e.g., $\{a, b, \ldots, z\}$, $\{0, 1\}$.

String: A string over an alphabet is a finite sequence of symbols from the alphabet, e.g.,"apple", "1011", "abc", "e" (empty string).

The set of all strings (including the empty string) over Σ is denoted by $\Sigma^\star$.

Alphabets and Languages

Alphabet: a finite set of symbols, e.g., $\{a, b, \ldots, z\}$, $\{0, 1\}$.

String: A string over an alphabet is a finite sequence of symbols from the alphabet, e.g.,“apple”, “1011”, ”abc”, ”e” (empty string).

The set of all strings (including the empty string) over Σ is denoted by $\Sigma^\star$.

What is the cardinality of $\Sigma^\star$?
Is $\Sigma^\star$ countable?

Alphabets and Languages

Alphabet: a finite set of symbols, e.g., $\{a, b, \ldots, z\}$, $\{0, 1\}$.

String: A string over an alphabet is a finite sequence of symbols from the alphabet, e.g.,“apple”, “1011”, ”abc”, ”e” (empty string).

The set of all strings (including the empty string) over Σ is denoted by $\Sigma^\star$.

What is the cardinality of $\Sigma^\star$?
Is $\Sigma^\star$ countable?

The length of a string is the length of the sequence, e.g. $|abc| = 3$.

Alphabets and Languages

Alphabet: a finite set of symbols, e.g., $\{a, b, \ldots, z\}$, $\{0, 1\}$.

String: A string over an alphabet is a finite sequence of symbols from the alphabet, e.g.,"apple", "1011", "abc", "e" (empty string).

The set of all strings (including the empty string) over Σ is denoted by $\Sigma^\star$.

What is the cardinality of $\Sigma^\star$?
Is $\Sigma^\star$ countable?

The length of a string is the length of the sequence, e.g. $|abc| = 3$.

A string $w \in \Sigma^\star$ can be considered as a function $w : \{1, \ldots, |w|\} \to \sigma$, where $w(j)$ is the symbol at the $j-$th position.

Alphabets and Languages

Alphabet: a finite set of symbols, e.g., $\{a, b, \ldots, z\}$, $\{0, 1\}$.

String: A string over an alphabet is a finite sequence of symbols from the alphabet, e.g.,“apple”, “1011”, ”abc”, ”e” (empty string).

The set of all strings (including the empty string) over Σ is denoted by $\Sigma^\star$.

What is the cardinality of $\Sigma^\star$?
Is $\Sigma^\star$ countable?

The length of a string is the length of the sequence, e.g. $|abc| = 3$.

A string $w \in \Sigma^\star$ can be considered as a function $w : \{1, \ldots, |w|\} \to \sigma$, where $w(j)$ is the symbol at the $j-$th position.

Language: Any subset L of $\Sigma^\star$ for alphabet Σ is called a language over Σ.

String operations

Concatenation: Two strings x, y over the same alphabet, e.g. $x, y \in \Sigma^\star$, can be combined. $w = x \circ y$, or simply $w = xy$. $|w| = |x| + |y|$, $w(i) = x(i)$ for $i \leq |y|$, and $w(i) = y(i - |x|)$ for $i > |x|$.

String operations

Concatenation: Two strings x, y over the same alphabet, e.g. $x, y \in \Sigma^\star$, can be combined. $w = x \circ y$, or simply $w = xy$. $|w| = |x| + |y|$, $w(i) = x(i)$ for $i \leq |y|$, and $w(i) = y(i - |x|)$ for $i > |x|$.

$w \circ e = e \circ w = w$.

String operations

Concatenation: Two strings x, y over the same alphabet, e.g. $x, y \in \Sigma^\star$, can be combined. $w = x \circ y$, or simply $w = xy$. $|w| = |x| + |y|$, $w(i) = x(i)$ for $i \leq |y|$, and $w(i) = y(i - |x|)$ for $i > |x|$.

$w \circ e = e \circ w = w$.

Concatenation is associative, $x(yz) = (xy)x$

String operations

Substring: A string v is a substring of w if w can be written as $w = xvy$.

String operations

Substring: A string v is a substring of w if w can be written as $w = xvy$.

If $w = vx$ then v is a **prefix** of w, and if $w = xv$, then v is a **suffix** of w.

String operations

Power w^i: For each string $w \in \Sigma^\star$ and natural number $i \in \mathbb{N}$, w^i is defined as (definition by induction):

$$
\begin{aligned}
w^0 &= e \\
w^{i+1} &= w^i \circ w \text{ for each } i \geq 0
\end{aligned}
$$

String operations

Reversal The reverse of a string w, denoted by w^R, is the string spelled backwards, $w^R(i) = w(|w| - i + 1)$. Inductive definition:

If $|w| = 0$, then $w^R = w = e$

If $|w| = n + 1$ for some $n \in \mathbb{N}$, then $w = ua$ for some $a \in \Sigma$, and $w^R = au^R$.

String operations

Reversal The reverse of a string w, denoted by w^R, is the string spelled backwards, $w^R(i) = w(|w| - i + 1)$. Inductive definition:

If $|w| = 0$, then $w^R = w = e$

If $|w| = n + 1$ for some $n \in \mathbb{N}$, then $w = ua$ for some $a \in \Sigma$, and $w^R = au^R$.

Theorem

For any two strings x, w, $(wx)^R = x^R w^R$.

String operations

Reversal The reverse of a string w, denoted by w^R, is the string spelled backwards, $w^R(i) = w(|w| - i + 1)$. Inductive definition:

If $|w| = 0$, then $w^R = w = e$

If $|w| = n + 1$ for some $n \in \mathbb{N}$, then $w = ua$ for some $a \in \Sigma$, and $w^R = au^R$.

Theorem

For any two strings x, w, $(wx)^R = x^R w^R$.

Proof, by induction (see Ch. 1.5 for proof techniques)

Languages

Given an alphabet Σ, any subset of $\Sigma^\star$ is called a language.

Languages

Given an alphabet Σ, any subset of $\Sigma^\star$ is called a language.

E.g. $\Sigma = \{0, 1\}$, $L_1 = \{0, 1\}$, $L_2 = \{e, 0, 1, 00, 01, 10, 11\}$,
$L_3 = \{w \in \Sigma^\star \mid |w| \leq 2\}$, $L_4 = \{w \in \Sigma^\star \mid w \text{ has an more 0's than 1's}\}$.
$L_5 = \{w \in \Sigma^\star \mid \sum_{i=1}^{|w|} w_i < 3\}$.

Languages

Given an alphabet Σ, any subset of $\Sigma^\star$ is called a language.

E.g. $\Sigma = \{0, 1\}$, $L_1 = \{0, 1\}$, $L_2 = \{e, 0, 1, 00, 01, 10, 11\}$,
$L_3 = \{w \in \Sigma^\star \mid |w| \leq 2\}$, $L_4 = \{w \in \Sigma^\star \mid w \text{ has an more 0's than 1's}\}$.
$L_5 = \{w \in \Sigma^\star \mid \sum_{i=1}^{|w|} w_i < 3\}$.

$$L = \{w \in \Sigma^\star \mid w \text{ has the property } P\}$$

Languages

Given an alphabet Σ, any subset of $\Sigma^\star$ is called a language.

E.g. $\Sigma = \{0, 1\}$, $L_1 = \{0, 1\}$, $L_2 = \{e, 0, 1, 00, 01, 10, 11\}$,
$L_3 = \{w \in \Sigma^\star \mid |w| \leq 2\}$, $L_4 = \{w \in \Sigma^\star \mid w \text{ has an more 0's than 1's}\}$.
$L_5 = \{w \in \Sigma^\star \mid \sum_{i=1}^{|w|} w_i < 3\}$.

$$L = \{w \in \Sigma^\star \mid w \text{ has the property } P\}$$

Any language is countable. Proof?

Languages

Given an alphabet Σ, any subset of $\Sigma^\star$ is called a language.

E.g. $\Sigma = \{0, 1\}$, $L_1 = \{0, 1\}$, $L_2 = \{e, 0, 1, 00, 01, 10, 11\}$,
$L_3 = \{w \in \Sigma^\star \mid |w| \leq 2\}$, $L_4 = \{w \in \Sigma^\star \mid w \text{ has an more 0's than 1's}\}$.
$L_5 = \{w \in \Sigma^\star \mid \sum_{i=1}^{|w|} w_i < 3\}$.

$$L = \{w \in \Sigma^\star \mid w \text{ has the property } P\}$$

Any language is countable. Proof?

Question: How many languages are there over a non-empty alphabet Σ?

Languages

Given an alphabet Σ, any subset of $\Sigma^\star$ is called a language.

E.g. $\Sigma = \{0, 1\}$, $L_1 = \{0, 1\}$, $L_2 = \{e, 0, 1, 00, 01, 10, 11\}$,
$L_3 = \{w \in \Sigma^\star \mid |w| \leq 2\}$, $L_4 = \{w \in \Sigma^\star \mid w \text{ has an more 0's than 1's}\}$.
$L_5 = \{w \in \Sigma^\star \mid \sum_{i=1}^{|w|} w_i < 3\}$.

$$L = \{w \in \Sigma^\star \mid w \text{ has the property } P\}$$

Any language is countable. Proof?

Question: How many languages are there over a non-empty alphabet Σ?

See Ch. 1.1, 1.4 for sets and countability.

Language operations

Languages are sets, so set operations (union, intersection, difference) can be used on languages.

Language operations

Languages are sets, so set operations (union, intersection, difference) can be used on languages.

Complement: $\overline{L} = \Sigma^{\star} \setminus L$

Language operations

Languages are sets, so set operations (union, intersection, difference) can be used on languages.
Complement: $\overline{L} = \Sigma^\star \setminus L$
Concatenation: L_1, L_2 are languages over Σ. $L = L_1 \circ L_2$ (or simply $L = L_1 L_2$) is defined as

$$L = \{w_1 w_2 \mid w_1 \in L_1 \text{ and } w_2 \in L_2\}$$

Language operations

Languages are sets, so set operations (union, intersection, difference) can be used on languages.

Complement: $\overline{L} = \Sigma^\star \setminus L$

Concatenation: L_1, L_2 are languages over Σ. $L = L_1 \circ L_2$ (or simply $L = L_1 L_2$) is defined as

$$L = \{w_1 w_2 \mid w_1 \in L_1 \text{ and } w_2 \in L_2\}$$

Kleene star: The Kleene star of a language L, denoted by $L^\star$, is the set of strings obtained by concatenating 0 or more strings from L.

Language operations

Languages are sets, so set operations (union, intersection, difference) can be used on languages.
Complement: $\overline{L} = \Sigma^\star \setminus L$
Concatenation: L_1, L_2 are languages over Σ. $L = L_1 \circ L_2$ (or simply $L = L_1 L_2$) is defined as

$$L = \{w_1 w_2 \mid w_1 \in L_1 \text{ and } w_2 \in L_2\}$$

Kleene star: The Kleene star of a language L, denoted by $L^\star$, is the set of strings obtained by concatenating 0 or more strings from L.
Note that $\Sigma^\star$ and $L^\star$ definitions are consistent.

Language operations

Languages are sets, so set operations (union, intersection, difference) can be used on languages.

Complement: $\overline{L} = \Sigma^\star \setminus L$

Concatenation: L_1, L_2 are languages over Σ. $L = L_1 \circ L_2$ (or simply $L = L_1 L_2$) is defined as

$$L = \{w_1 w_2 \mid w_1 \in L_1 \text{ and } w_2 \in L_2\}$$

Kleene star: The Kleene star of a language L, denoted by $L^\star$, is the set of strings obtained by concatenating 0 or more strings from L.

Note that $\Sigma^\star$ and $L^\star$ definitions are consistent.

$L^+ = LL^\star$

Language operations

Languages are sets, so set operations (union, intersection, difference) can be used on languages.
Complement: $\overline{L} = \Sigma^\star \setminus L$
Concatenation: L_1, L_2 are languages over Σ. $L = L_1 \circ L_2$ (or simply $L = L_1 L_2$) is defined as

$$L = \{w_1 w_2 \mid w_1 \in L_1 \text{ and } w_2 \in L_2\}$$

Kleene star: The Kleene star of a language L, denoted by $L^\star$, is the set of strings obtained by concatenating 0 or more strings from L.
Note that $\Sigma^\star$ and $L^\star$ definitions are consistent.
$L^+ = LL^\star$
Question: Is there any L such that a) $L^\star = \emptyset$? b) $L^\star = \{e\}$? c) $L = L^\star$?