Strings and Languages

Lecture 1

Tuesday, August 25, 2020

LaTeXed: May 28, 2020 11:04

1.1
Strings

Alphabet

An alphabet is a **finite** set of symbols.

Examples of alphabets:

- $\Sigma = \{0, 1\}$,
- $\Sigma = \{a, b, c, \ldots, z\}$,
- ASCII.
- UTF8.
- $\Sigma = \{\langle \mathbf{moveforward} \rangle, \langle \mathbf{moveback} \rangle\}$

Alphabet

An alphabet is a **finite** set of symbols.
Examples of alphabets:

- $\Sigma = \{0, 1\}$,
- $\Sigma = \{a, b, c, \ldots, z\}$,
- ASCII.
- UTF8.
- $\Sigma = \{\langle \textbf{moveforward} \rangle, \ \langle \textbf{moveback} \rangle\}$

String Definitions

Definition

1. A string/word over Σ is a **finite sequence** of symbols over Σ. For example, '0101001', '*string*', '$\langle\textbf{moveback}\rangle\langle\textbf{rotate90}\rangle$'
2. ϵ is the empty string.
3. The length of a string w (denoted by $|w|$) is the number of symbols in w. For example, $|101| = 3$, $|\epsilon| = 0$
4. For integer $n \geq 0$, Σ^n is set of all strings over Σ of length n. Σ^* is the set of all strings over Σ.

Inductive/recursive definition of strings

Formal definition of a **string**:

- ϵ is a string of length 0
- ax is a string if $a \in \Sigma$ and x is a string. The length of ax is $1 + |x|$

The above definition helps prove statements rigorously via induction.

- Alternative recursive definition useful in some proofs: xa is a string if $a \in \Sigma$ and x is a string. The length of xa is $1 + |x|$

Convention

- $a, b, c, \ldots$ denote elements of Σ
- $w, x, y, z, \ldots$ denote strings
- $A, B, C, \ldots$ denote sets of strings

Much ado about nothing

- ϵ is a string containing no symbols. It is not a set
- $\{\epsilon\}$ is a set containing one string: the empty string. It is a set, not a string.
- $\emptyset$ is the empty set. It contains no strings.
- $\{\emptyset\}$ is a set containing one element, which itself is a set that contains no elements.

Concatenation and properties

- If x and y are strings then xy denotes their concatenation.
- **concatenation** defined recursively :
  - $xy = y$ if $x = \epsilon$
  - $xy = a(wy)$ if $x = aw$
- xy sometimes written as $x \bullet y$.
- concatenation is **associative**: $(uv)w = u(vw)$
  hence write $uvw \equiv (uv)w = u(vw)$
- **not** commutative: uv not necessarily equal to vu
- The *identity* element is the empty string ϵ:

$$\epsilon u = u\epsilon = u.$$

Concatenation and properties

- If x and y are strings then xy denotes their concatenation.
- **concatenation** defined recursively :
  - $xy = y$ if $x = \epsilon$
  - $xy = a(wy)$ if $x = aw$
- xy sometimes written as $x \bullet y$.
- concatenation is **associative**: $(uv)w = u(vw)$
  hence write $uvw \equiv (uv)w = u(vw)$
- **not** commutative: uv not necessarily equal to vu
- The *identity* element is the empty string ϵ:

$$\epsilon u = u\epsilon = u.$$

Concatenation and properties

- If x and y are strings then xy denotes their concatenation.
- **concatenation** defined recursively :
  - $xy = y$ if $x = \epsilon$
  - $xy = a(wy)$ if $x = aw$
- xy sometimes written as $x \bullet y$.
- concatenation is **associative**: $(uv)w = u(vw)$
  hence write $uvw \equiv (uv)w = u(vw)$
- **not** commutative: uv not necessarily equal to vu
- The *identity* element is the empty string ϵ:

$$\epsilon u = u\epsilon = u.$$

Concatenation and properties

- If x and y are strings then xy denotes their concatenation.
- **concatenation** defined recursively :
  - $xy = y$ if $x = \epsilon$
  - $xy = a(wy)$ if $x = aw$
- xy sometimes written as $x \bullet y$.
- concatenation is **associative**: $(uv)w = u(vw)$
  hence write $uvw \equiv (uv)w = u(vw)$
- **not** commutative: uv not necessarily equal to vu
- The *identity* element is the empty string ϵ:

$$\epsilon u = u\epsilon = u.$$

Concatenation and properties

- If x and y are strings then xy denotes their concatenation.
- **concatenation** defined recursively :
  - $xy = y$ if $x = \epsilon$
  - $xy = a(wy)$ if $x = aw$
- xy sometimes written as $x \bullet y$.
- concatenation is **associative**: $(uv)w = u(vw)$
  hence write $uvw \equiv (uv)w = u(vw)$
- **not** commutative: uv not necessarily equal to vu
- The *identity* element is the empty string ϵ:

$$\epsilon u = u\epsilon = u.$$

Substrings, prefix, suffix

Definition

v is **substring** of w $\iff$ there exist strings x, y such that $w = xvy$.

- If $x = \epsilon$ then v is a prefix of w
- If $y = \epsilon$ then v is a suffix of w

String exponents

Definition

If w is a string then w^n is defined inductively as follows:
$w^n = \epsilon$ if $n = 0$
$w^n = ww^{n-1}$ if $n > 0$

Example: $(blah)^4 = blahblahblahblah$.

Set Concatenation

Definition

Given two sets X and Y of strings (over some common alphabet Σ) the **concatenation** of X and Y is

$$XY = \{xy \mid x \in X, y \in Y\}$$

Set Concatenation

Definition

Given two sets X and Y of strings (over some common alphabet Σ) the **concatenation** of X and Y is

$$XY = \{xy \mid x \in X, y \in Y\}$$

Example

$X = \{fido, rover, spot\}$,
$Y = \{fluffy, tabby\}$
$\implies$
$XY = \{fidofluffy, fidotabby, roverfluffy, \ldots\}$.

Σ^* and languages

Definition

1. Σ^n is the set of all strings of length n. Defined inductively:
   $\Sigma^n = \{\epsilon\}$ if $n = 0$
   $\Sigma^n = \Sigma\Sigma^{n-1}$ if $n > 0$
2. $\Sigma^* = \cup_{n \geq 0} \Sigma^n$ is the set of all finite length strings
3. $\Sigma^+ = \cup_{n \geq 1} \Sigma^n$ is the set of non-empty strings.

Definition

A language L is a set of strings over Σ. In other words $L \subseteq \Sigma^*$.

Σ^* and languages

Definition

1. Σ^n is the set of all strings of length n. Defined inductively:
   $\Sigma^n = \{\epsilon\}$ if $n = 0$
   $\Sigma^n = \Sigma\Sigma^{n-1}$ if $n > 0$
2. $\Sigma^* = \cup_{n \geq 0} \Sigma^n$ is the set of all finite length strings
3. $\Sigma^+ = \cup_{n \geq 1} \Sigma^n$ is the set of non-empty strings.

Definition

A language L is a set of strings over Σ. In other words $L \subseteq \Sigma^*$.

Exercise

Answer the following questions taking $\Sigma = \{0, 1\}$.

1. What is Σ^0?
2. How many elements are there in Σ^3?
3. How many elements are there in Σ^n?
4. What is the length of the longest string in Σ?
5. Does Σ^* have strings of infinite length?
6. If $|u| = 2$ and $|v| = 3$ then what is $|u \bullet v|$?
7. Let u be an arbitrary string in Σ^*. What is ϵu? What is $u\epsilon$?
8. Is $uv = vu$ for every $u, v \in \Sigma^*$?
9. Is $(uv)w = u(vw)$ for every $u, v, w \in \Sigma^*$?

THE END

...

(for now)