# CSCI/MATH 2113 Discrete Structures

## 6.1 Language: The Set Theory of Strings

Alyssa Motas

March 2, 2021

## Contents

# 1 Definition of an alphabet

An *alphabet* is a finite nonempty set. We write $\Sigma$ for an alphabet and we sometimes call the elements of $\Sigma$ *letters*. For example, we may have $\Sigma = \{0, 1\}$ or $\Sigma = \{a, b, c, d, e\}$.

# 2 Powers of an alphabet

If $\Sigma$ is an alphabet and $n \in \mathbb{Z}^+$, we define the *powers* of $\Sigma$ recursively as follows:

1. $\Sigma^1 = \Sigma$; and
2. $\Sigma^{n+1} = \{xy \mid x \in \Sigma, y \in \Sigma^n\}$, where $xy$ denotes the juxtaposition of $x$ and $y$.

## 2.1 Example

Let $\Sigma$ be an alphabet. With $\Sigma = \{0, 1\}$, we find that

$$\Sigma^2 = \{00, 01, 10, 11\} \text{ and } |\Sigma^2| = |\Sigma|^2 = 2^2 \text{ two-symbol strings.}$$

In general, we have $|\Sigma^n| = |\Sigma|^n$.

# 3 Empty string

For an alphabet $\Sigma$, we define $\Sigma^0 = \{\lambda\}$, where $\lambda$ denotes the *empty string*. That is, the string consisting of *no* symbols taken from $\Sigma$. Note that even though $\lambda \notin \Sigma$, we do have $\varnothing \subseteq \Sigma$. Also, $\{\lambda\} \neq \varnothing$ because $|\{\lambda\}| = 1 \neq 0 = |\varnothing|$.

# 4 Union of alphabets

If $\Sigma$ is an alphabet, then

(a) $\Sigma^+ = \bigcup_{n=1}^{\infty} \Sigma^n = \bigcup_{n \in \mathbb{Z}^+} \Sigma^n$;

(b) $\Sigma^* = \bigcup_{n=0}^{\infty} \Sigma^n$.

The difference between (a) and (b) is that $\lambda \in \Sigma^n$ only when $n = 0$. Also, $\Sigma^* = \Sigma^+ \cup \Sigma^0$.

We shall also refer to the elements of $\Sigma^+$ or $\Sigma^*$ as *words* and sometimes as *sentences*. Finally, we note that even though the sets $\Sigma^+$ and $\Sigma^*$ are *infinite*, the elements of these sets are *finite* strings of symbols.

### 4.1 Example

- For $\Sigma = \{0, 1\}$ the set $\Sigma^*$ consists of all finite strings (binary words) of 0's and 1's together with the empty string.
- If $\Sigma = \{+, \times, 0, 1, \ldots, 9, (, ), \quad \}$ we have

$$((14 + 12) \times 3) \times 1009 \in \Sigma^* \text{ or } ) + (x)1 + (\times 3) \in \Sigma^*.$$

# 5 Equality of sets

If $w_1, w_2 \in \Sigma^+$, then we may write

$$w_1 = x_1 x_2 \ldots x_m \text{ and } w_2 = y_1 y_2 \ldots y_n$$

for $m, n \in \mathbb{Z}^+$ and $x_1, x_2, \ldots, x_m, y_1, y_2, \ldots, y_n \in \Sigma$. We say that the strings $w_1$ and $w_2$ are *equal*, and we write $w_1 = w_2$, if $m = n$, and $x_i = y_i$ for all $1 \leq i \leq m$.

# 6 Length

Let $w = x_1 x_2 \ldots x_n \in \Sigma^+$, where $x_i \in \Sigma$ for each $1 \leq i \leq n$. We define the *length* of $w$, which is denoted by $||w||$, as the value $n$. For the case of $\lambda$, we have $||\lambda|| = 0$.

# 7 Concatenation

Let $x, y \in \Sigma^+$ with $x = x_1 x_2 \ldots x_m$ and $y = y_1 y_2 \ldots y_n$, so that each $x_i$, for $1 \leq i \leq m$, and each $y_j$, for $1 \leq j \leq n$, is in $\Sigma$. The *concatenation* of $x$ and $y$, which we write as $xy$, is the string

$$x_1 x_2 \ldots x_m y_1 y_2 \ldots y_n.$$

The concatenation of $x$ and $\lambda$ is $x\lambda = x_1 x_2 \ldots x_m \lambda = x_1 x_2 \ldots x_m = x$, and the concatenation of $\lambda$ and $x$ is $\lambda x = \lambda x_1 x_2 \ldots x_m = x_1 x_2 \ldots x_m = x$. Finally, the concatenation of $\lambda$ and $\lambda$ is $\lambda\lambda = \lambda$.

Here, we have defined a closed binary operation on $\Sigma^*$ (and $\Sigma^+$). This operation is associative but not commutative unless $|\Sigma| = 1$. The $\lambda$ is also the identity for the operation of concatenation. We also have

$$||xy|| = ||x|| + ||y||, \qquad \text{for all } x, y \in \Sigma^*.$$

# 8 Powers of a string

For each $x \in \Sigma^*$, we define the *powers* of $x$ by $x^0 = \lambda, x^1 = x, x^2 = xx, x^3 = xx^2, \dots, x^{n+1} = xx^n, \dots,$ where $n \in \mathbb{N}$.

# 9 Proper prefix and suffix

If $x, y \in \Sigma^*$ and $w = xy$, then the string $x$ is called a *prefix* of $w$, and if $y \neq \lambda$, then $x$ is said to be a *proper prefix.* Similarly, the string $y$ is called a *suffix* of $w$; it is a *proper suffix* when $x \neq \lambda$.

In general, for an alphabet $\Sigma$, if $n \in \mathbb{Z}^+$ and $x_i \in \Sigma$, for all $1 \leq i \leq n$, then each of $\lambda, x_1, x_1x_2, x_1x_2x_3, \dots,$ and $x_1x_2x_3 \dots x_n$ is a prefix of the string $x = x_1x_2x_3 \dots x_n$. And $\lambda, x_n, x_{n-1}x_n, x_{n-2}x_{n-1}x_n, \dots,$ and $x_1x_2x_3 \dots x_n$ are all suffixes of $x$. So, $x$ has $n+1$ prefixes, $n$ of which are proper, and the situation is the same for suffixes.

# 10 Substring

If $x, y, z \in \Sigma^*$ and $w = xyz$, then $y$ is called a *substring* of $w$. When at least one of $x$ and $z$ is different from $\lambda$ (so that $y$ is different from $w$), we call $y$ a *proper substring or subword.*

# 11 Language

For a given alphabet $\Sigma$, any subset of $\Sigma^*$ is called a *language* over $\Sigma$. This includes the subset $\varnothing$, which we call the *empty language.*

## 11.1 Example

With $\Sigma = \{0, 1\}$, the sets

$$A = \{0, 01, 001\}$$

and

$$B = \{0, 01, 001, 0001, \dots\}$$

are examples of languages over $\Sigma$.