Computer Science 351

Alphabets, Strings, and Languages

Instructor: Wayne Eberly

Department of Computer Science
University of Calgary

Lecture #1 — Continued

Goals for Today

Goals for Today:

- Introduce ***mathematical structures*** that will be used, intensively, during the beginning and middle of this course.

Note:

- A ***mathematics review*** is also available, and should be used, if you do not remember properties of ***sets*** and ***sequences*** — mathematical objects which were introduced in CPSC 251 and used here.

Alphabets

Definition: An ***alphabet*** is a finite non-empty set.

- Alphabets in this course will often be given names Σ, Γ, or Σ_i or Γ_i, for some non-negative integer i, when more than one alphabet must be considered.
- Alphabets that will initially include simple ones like $\{\mathtt{0}, \mathtt{1}\}$, $\{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$, the set of decimal ***digits*** $\{\mathtt{0}, \mathtt{1}, \mathtt{2}, \ldots, \mathtt{9}\}$, or the set of lower-case letters in the Roman alphabet $\{\mathtt{a}, \mathtt{b}, \mathtt{c}, \ldots, \mathtt{z}\}$. More complicated alphabets, that are more useful for the problems being solved, will be considered later on.

Alphabets

- Another example — that will not be used in this course (because it is too large) is the set of possible values that can be represented using a single word of machine memory — which could be modelled as the set of integers between 0 and $2^k - 1$, where k is the size (in bits) of a word of machine memory.
- The elements of an alphabet are often called ***symbols***.

Strings

Definition: A ***string***, over an alphabet Σ, is a finite sequence of elements of Σ.

- For example, if $\Sigma = \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$ then the sequence

  $$\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{a}$$

  is an example of a string over the alphabet Σ. Strings are often written by listing the symbols in the string immediately after the other, without spaces or commas separating them, so that

  $$\mathtt{abca}$$

  is another (more frequently used) representation of the same string over Σ.

Strings

Definition: The ***length*** of a string ω, over an alphabet Σ, is the same as its length when it is considered as a sequence.

- For example, the if ω is the above string, $\mathtt{abca}$ over Σ, then the length of ω is 4.
- The length of a string ω is often written as $|\omega|$ — so that $|\omega| = 4$ for the string $\omega = \mathtt{abca}$ as above.

Strings

Definition: The string of length zero (over any alphabet Σ) is called the ***empty string*** and is denoted by λ.[1]

Definition: For any alphabet Σ, $\mathbf{\Sigma}^{\star}$ is the set of *all* strings over Σ.

[1] Some other references denoted the empty string by "ε" instead.

Strings

Definition: Suppose that

$$\mu = a_1 a_2 \ldots a_n \qquad \text{and} \qquad \nu = b_1 b_2 \ldots b_m$$

are strings over an alphabet Σ (that is, suppose that $\mu, \nu \in \Sigma^\star$) with lengths n and m, respectively, so that

$$\alpha_1, \alpha_2, \ldots, \alpha_n, \beta_1, \beta_2, \ldots, \beta_m \in \Sigma.$$

Then the ***concatenation*** of μ and ν, denoted by $\mu \cdot \nu$, is the string

$$\alpha_1 \alpha_2 \ldots \alpha_n \beta_1 \beta_2 \ldots \beta_m$$

over Σ, with length $n + m$, obtained by listing the symbols in ν after the symbols in μ.

Strings

Definition: If $\mu, \omega \in \Sigma^\star$, them μ is a ***substring*** of ω if there exist strings $\nu, \varphi \in \Sigma^\star$ such that $\omega = \nu \cdot \mu \cdot \varphi$.

Strings

- Another (equivalent) way to define "substring" is as follows: If

  $$\omega = a_1 a_2 \dots a_n$$

  and

  $$\mu = b_1 b_2 \dots b_m$$

  are strings in $\Sigma^\star$ with lengths n and m respectively (so that $a_1, a_2, \dots, a_n, b_1, b_2, \dots, b_m \in \Sigma$), then μ is a ***substring*** of ω if $m \leq n$ and there exists an integer i such that $0 \leq i \leq n - m$, and such that $b_j = a_{j+i}$ for every integer j such that $1 \leq j \leq m$.

Strings

Definition: Once again, suppose that $\mu, \omega \in \Sigma^\star$. Then μ is a ***prefix*** of ω if there exists a string $\nu \in \Sigma^\star$ such that

$$\omega = \mu \cdot \nu,$$

and μ is a ***suffix*** of ω if there exists a string $\nu \in \Sigma^\star$ such that

$$\omega = \nu \cdot \mu.$$

Strings

- Another (equivalent) way to define “prefix” and “suffix” is as follows: If

$$\omega = a_1 a_2 \ldots a_n$$

and

$$\mu = b_1 b_2 \ldots b_m$$

are strings in $\Sigma^\star$ with lengths n and m respectively (so that $a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_m \in \Sigma$), then μ is a ***prefix*** of ω if $m \leq n$ and $b_j = a_j$ for every integer j such that $1 \leq j \leq n$; μ is a ***suffix*** of ω if $m \leq n$ and $b_j = a_{j+n-m}$ for every integer j such that $1 \leq j \leq n - m$, instead.

Languages

Definition: A ***language*** over an alphabet Σ is a subset of $\Sigma^\star$.

- That is, a ***language*** over an alphabet Σ is a set of (some of the) strings over Σ.

Languages

- One kind of computational problem is a ***decision problem*** — one where the required answer is either “Yes” or “No”. Strings of symbols over an alphabet Σ can be used to represent, or “encode”, instances of a decision problem — so that the set of instances of a decision problem, for which the corresponding answer is “Yes”, is encoded by a language over Σ.
- We will, therefore, be considering various (kinds of) algorithms that decide membership in languages, as well as properties of languages, in much of this course.