

1 Definitions

1.1 Alphabet

Let Σ be a finite set of symbols (alternatively called characters), called the alphabet. No assumption is made about the nature of the symbols.

1.2 Strings

A string (or word) over Σ is any finite sequence of symbols from Σ . For example if $\Sigma = \{0, 1\}$ then 01011 is a string over Σ .

The **length of a string S is the number of symbols** in s (the length of the sequence) and can be any non-negative integer. It is often denoted as $|S|$.

The **empty string** is the unique string over Σ of length 0, and is denoted as ε .

The **set of all strings over Σ of length n** is denoted Σ^n . For example, if $\Sigma = 0, 1$, then $\Sigma^2 = \{00, 01, 10, 11\}$. Note that $\Sigma^0 = \{\varepsilon\}$ for any alphabet Σ .

The **set of all strings of any length over Σ** is the *Kleene closure* of Σ and is denoted as Σ^* . Also:

$$\Sigma^* = \bigcup_{n \in \mathbb{N} \cup \{0\}} \Sigma^n$$

The **set of all non-empty strings over Σ** is denoted by Σ^+

1.3 Subsequences

A **subsequence** of S is a string that can be **obtained by deleting any number (from 0 to n) of non-consecutive characters** from S , including S and the empty string too.

1.3.1 Let us count: subsequences

A subsequence of S can be described by a binary string B of n elements, telling us for each position i if the corresponding character is kept in the subsequence (1) or deleted (0).

- $S = \text{apple}, B = 11010 \rightarrow \text{subsequence} = \text{apl}$

Thus, the **number of subsequences** corresponds to the number of binary strings of length n , hence

$$\text{subsequences} = 2^n$$

Notice that if two subsequences produced by different choices are **identical, they will be counted twice**.

1.4 Substrings

A string s is said to be a **substring** (or *factor*) of t if there exist (possibly empty) strings u and v such that $t = usv$.

Given a string t , **suffixes** and **prefixes** are special substrings of t .

1.4.1 Let us count: substrings

Given a string of length n :

- We have 1 substring of length n (starting at 1)
- We have 2 substrings of length $n-1$ (starting at 1 and 2)
- We have 3 substrings of length $n-2$ (starting at 1,2 and 3)
- ...
- We have $n - i + 1$ substrings of length i (starting at all positions from 1 to $n - i + 1$)
- We have n substrings of length 1 (starting at every position)

Plus, 1 substring of length 0.

This brings the **total number of substrings** of a string of length n to

$$substrings = \sum_{i=1}^n i = \frac{n(n+1)}{2} + 1$$

1.5 Prefixes and suffixes

A string s is said to be a **prefix** of t if there exists a string u such that $t = su$. If u is nonempty, s is said to be a *proper* prefix of t .

Simmetrically, a string s is said to be a **suffix** of t if there exists a string u such that $t = us$. If u is nonempty, s is said to be a *proper* suffix of t .

1.5.1 Let us count: prefixes and suffixes

The number of all possible prefixes/suffixes of a string of length n is

$$prefixes/suffixes = n + 1$$

n are non-empty, $+1$ because we include also the empty string as prefix/suffix.

1.6 Reverse and palindrome

The **reverse** of a string is a string with the same symbols but in reverse order. For example, if $s = abc$ (where a , b , and c are symbols of the alphabet), then the reverse of s is cba .

A string that is the reverse of itself is called a **palindrome**, which also includes the empty string and all strings of length 1.

1.7 Rotations and permutations

A string $s = uv$ is said to be a **rotation** of t if $t = vu$. For example, if $\Sigma = \{0,1\}$ the string 00111001 is a rotation of 0100110 , where $u = 00110$ and $v = 01$.

1.7.1 Let us count: rotations

Given a string S , with $|S| = n$, we denote $S=AB$, where **A is a prefix** substring, and **B a suffix** substring.

For every possible pair A, B the string **$R=BA$ is a rotation of S** , including S itself, that is, the case where $A = \varepsilon$ and $B = S$.

Since we have n different non empty suffixes B , the number of rotations is

$$rotations = n$$

It is not $n + 1$ because the two suffixes $B = S$ and $B = \varepsilon$ produce exactly the same rotation, given by S itself.

1.7.2 Let us count: permutations

Given a string S , with $|S| = n$, the number of permutations of the characters in S is

$$permutations = n!$$

Notice that the formula does not consider the case of identical permutations: that is, a string of n identical characters, will have n identical permutations.