Basic Terminology

Alphabet = set of symbols (characters).

Often denoted Σ. In this course, Σ is always finite and nonempty. The characters should at least have an equality relation, will normally have an ordering, too.

Examples: $\{0, 1\}$

$\{a, b, c\}$

$\{A, C, G, T\}$

$\{0, 1, 2, \ldots, 9\}$

String on Σ = finite sequence of characters from Σ. Examples:

01000101

abracadabra

ACCTGTAGTT

This␣is␣a␣sentence

The empty string is denoted ε.

Length $|x|$ of string x is the number of characters in x.

Notation for parts of strings:

string x = abracadabra
position i: 0 1 2 3 4 5 6 7 8 9 10

x[2] = r x[1..5] = braca

a d a ← subsequence

x = abracadabra

abr cad bra

prefix (= substring starting at start of x)

substring

suffix (= substring ending at end of x)

Equality of strings

$x = y$

$\Leftrightarrow$

$|x| = |y|$ AND $x[i] = y[i]$ for all i from 0 to $|x|-1$

Lexicographic Ordering of strings

$x \leq y$

$\Updownarrow$

x is a prefix of y

OR

there exists index $i < |x|, |y|$
such that

$x[j] = y[j]$ for $0 \leq j < i$

$x[i] < y[i]$

[Exercise : verify this is an ordering, i.e. is reflexive, transitive, antisymmetric]

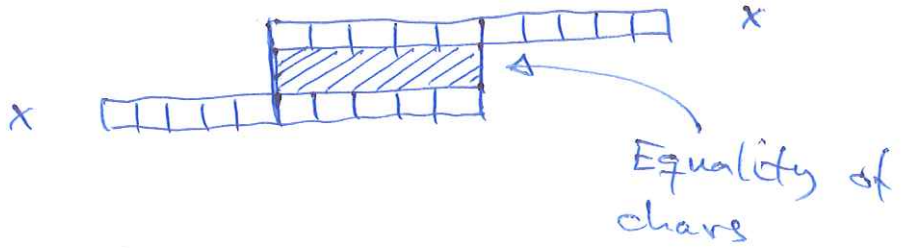
Concatenation : xy means string x concatenated by string y.

Powers of a string :

$$x^0 = \varepsilon$$
$$x^1 = x$$
$$x^2 = xx$$
$$x^3 = xxx$$
$$\vdots$$

Border of string x : a proper substring of x (i.e. of length $< |x|$) which is both a suffix and a prefix of x.

Figure :



Border(x) = Longest border of x.

$\text{border}_x(k) = |\text{Border}(x[0..k])|$

for $0 \leq k < |x|$

Example :

$x = a\ a\ b\ a\ a\ b\ a$

k	0	1	2	3	4	5	6
$\text{border}_x(k)$	0	1	0	1	2	3	4