Intro to Analysis of Algorithms
Computational Foundations
Sections 9.1-9.2
Chapter 9

Michael Soltys

CSU Channel Islands

[Ed: 4th, last updated: September 16, 2025]

Sections 9.1 & 9.2: Alphabets, strings and languages

Since long ago "markings" have been used to store & process information. The following pictures are from the *Smithsonian Museum of Natural History, Washington D.C.*

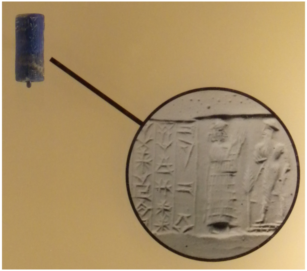
Engraved ocher plaque
Blombos Cave, South Africa
77,000–75,000 years old

Ishango bone
Congo, 25,000–20,000 years old
leg bone from a baboon; 3 rows of tally marks, to *add* or *multiply* (**?**)

Reindeer antler with tally marks
La Madeleine, France
17,000–11,500 years old

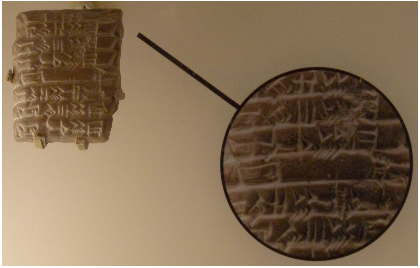
About 8,000 years ago, humans were using symbols to represent words and concepts. True forms of writing developed over the next few thousand years.

Cylinder seals were rolled accross wet clay tablets to produce raised designs

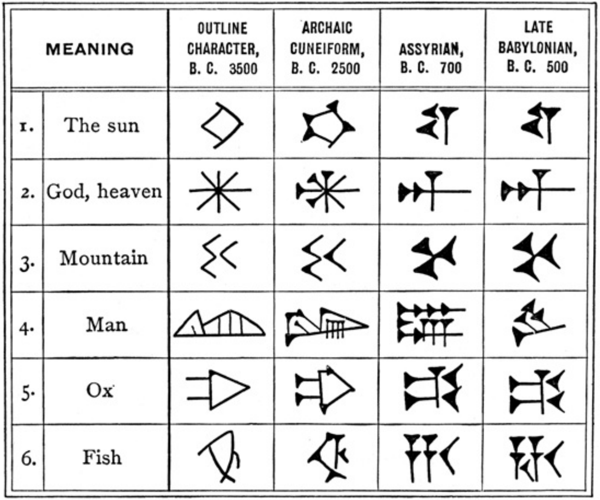


cylinder seal in lapis lazuli, Assyrian culture, Babylon, Iraq, 4,100–3,600 years ago

Cuneiform symbols stood for concepts and later for sounds and syllables



cuneiform clay tablet, Chakma, Chalush, near Babylon, Iraq, 4,000–2,600 years ago



MEANING		OUTLINE CHARACTER, B. C. 3500	ARCHAIC CUNEIFORM, B. C. 2500	ASSYRIAN, B. C. 700	LATE BABYLONIAN, B. C. 500
1.	The sun				
2.	God, heaven				
3.	Mountain				
4.	Man				
5.	Ox				
6.	Fish				

An *alphabet* is a finite, non-empty set of distinct symbols, denoted usually by Σ.

e.g., $\Sigma = \{0, 1\}$ (binary alphabet)
$\Sigma = \{a, b, c, \dots, z\}$ (lower-case letters alphabet)

A *string*, also called *word*, is a finite ordered sequence of symbols chosen from some alphabet.

e.g., 010011101011

$|w|$ denotes the *length* of the string w.

e.g., $|010011101011| = 12$

The *empty string*, ε, $|\varepsilon| = 0$, is in any Σ by default.

Σ^k is the set of strings over Σ of length *exactly* k.

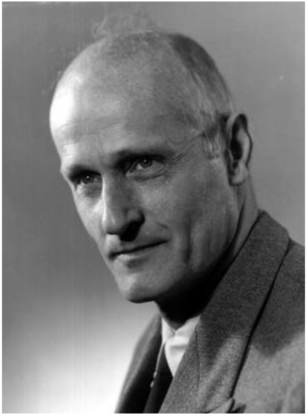
e.g., If $\Sigma = \{0, 1\}$, then

$$\Sigma^0 = \{\varepsilon\}$$
$$\Sigma^1 = \Sigma$$
$$\Sigma^2 = \{00, 01, 10, 11\}, \text{ etc. } |\Sigma^k|?$$

Kleene's star Σ^* is the set of all strings over Σ.

$$\Sigma^* = \Sigma^0 \cup \underbrace{\Sigma^1 \cup \Sigma^2 \cup \Sigma^3 \cup \ldots}_{=\Sigma^+}$$

Concatenation If x, y are strings, and $x = a_1a_2 \ldots a_m$ & $y = b_1b_2 \ldots b_n \Rightarrow x \cdot y = \underbrace{xy}_{\textit{juxtaposition}} = a_1a_2 \ldots a_mb_1b_2 \ldots b_n$



Stephen Cole Kleene

A *language* L is a collection of strings over some alphabet Σ, i.e., $L \subseteq \Sigma^*$. E.g.,

$$L = \{\varepsilon, 01, 0011, 000111, \ldots\} = \{0^n 1^n | n \geq 0\} \tag{1}$$

Note:

- $w\varepsilon = \varepsilon w = w$.
- $\{\varepsilon\} \neq \emptyset$; one is the language consisting of the single string ε, and the other is the empty language.

Two fundamental questions:

- How do we describe a language? (1) is just an *informal set-theoretic* description.
- Given a language $L \subseteq \Sigma^*$ and a string $x \in \Sigma^*$, how do we check if $x \in L$? E.g.,

$$L = \{\underbrace{10}_{2}, \underbrace{11}_{3}, \underbrace{101}_{5}, \underbrace{111}_{7}, \ldots\} \subseteq \{0,1\}^*$$

  $w \in L$ iff $w \in \{0,1\}^*$ encodes a prime number in standard binary notation.
- What is an algorithm?