Strings and regular expressions

Given a set A, we can construct the set A^* consisting of all finite sequences of elements of A. Often the set A is not a set of numbers, but some set of symbols. In this case, A is called an **alphabet**, and the finite sequences in A^* are called **words** from A, or sometimes **strings** from A. For this case in particular, the sequences in A^* are *not* written with commas. We assume that A^* contains the **empty sequence** or **empty string**, containing no symbols, and we denote this string by Λ. This string will be useful in Chapters 7 and 8.

Example 14. Let $A = \{a, b, c, \ldots, z\}$, the usual English alphabet. Then A^* consists of all ordinary words, such as ape, sequence, antidisestablishmentarianism, and so on, as well as "words" such as yxaloble, zigadongdong, cya, and pqrst. All finite sequences from A are in A^*, whether they have meaning or not. ◆

Cantenetion

If $w_1 = s_1 s_2 s_3 \cdots s_n$ and $w_2 = t_1 t_2 t_3 \cdots t_k$ are elements of A^* for some set A, we define the **catenation** of w_1 and w_2 as the sequence $s_1 s_2 s_3 \cdots s_n t_1 t_2 t_3 \cdots t_k$. The catenation of w_1 with w_2 is written as $w_1 \cdot w_2$ and is another element of A^*. Note that if w belongs to A^*, then $w \cdot \Lambda = w$ and $\Lambda \cdot w = w$. This property is convenient and is one of the main reasons for defining the empty string Λ.

Example 15. Let $A = \{\text{John, Sam, Jane, swims, runs, well, quickly, slowly}\}$. Then A^* contains real sentences such as "Jane swims quickly" and "Sam runs well," as well as nonsense sentences such as "Well swims Jane slowly John." Here we separate the elements in each sequence with spaces. ◆

Example 16. Let $A = \{0, 1\}$. Show that the following expressions are all regular expressions over A.
(a) $0^*(0 \vee 1)^*$ (b) $00^*(0 \vee 1)^*1$ (c) $(01)^*(01 \vee 1^*)$

(a) $\{00, 01, 1, \Lambda, \ldots\}$
(b) $\{01, 0011, \ldots\}$
(c) $\{01, 0101, 1, \Lambda, \ldots\}$