1.2 ΑΛΦΑΒΗΤΑ & ΓΛΩΣΣΕΣ

ΟΡΙΣΜΟΙ

Θεωρούμε **σύμβολα** συνήθως αποτελούμενα από φυσικούς αριθμούς, λατινικά ή ελληνικά γράμματα κλπ.

Ένα (πεπερασμένο) σύνολο συμβόλων ονομάζεται **αλφάβητο**. Συνήθως, ένα αλφάβητο συμβολίζεται με Σ και ένα σύμβολο με $\sigma \in \Sigma$.

Μια **συμβολοσειρά** (**string**) ή **λέξη** είναι μια πεπερασμένη ακολουθία συμβόλων.

Το μήκος μιας συμβολοσειράς $|\sigma|$ είναι το πλήθος των συμβόλων της $|\sigma_1\sigma_2\cdots\sigma_k| = k$, $|\sigma_1\sigma_2\sigma_2| = 3$, κ.ο.κ.

Η κενή συμβολοσειρά ε είναι εκείνη που αποτελείται από κανένα σύμβολο: $|\varepsilon| = 0$.

Πρόθεμα/κατάληξη μιας συμβολοσειράς είναι οποιοδήποτε σύνολο διατεταγμένων συμβόλων στην αρχή/τέλος της, αντιστοίχως.

Έστω u, v δύο συμβολοσειρές. Η **συνένωση** (**concatenation**) είναι η συμβολοσειρά uv με πρόθεμα την u και κατάληξη την v.

Μια (**τυπική**) **γλώσσα** είναι ένα (πεπερασμένο) σύνολο συμβολοσειρών (ή λέξεων) ενός αλφαβήτου.

Συμβολίζουμε με Σ^* την γλώσσα που αποτελείται από όλες τις συμβολοσειρές (ή λέξεις) του Σ και ονομάζουμε την Σ^* **κλειστότητα άστρο** ή **καθολική** (ή **παγκόσμια**) **γλώσσα** του Σ.

Παρατήρηση:

- Σ^* (αριθμησίμως) άπειρη
- Κάθε γλώσσα L στο Σ είναι $L \subset \Sigma^*$

Παραδείγματα:

- $\Sigma = \{1\} \Rightarrow \Sigma^* = \{\varepsilon, 1, 11, 111, 1111, ...\}$
- $\Sigma = \{0,1\} \Rightarrow \Sigma^* = \{\varepsilon, 0, 1, 00, 10, 01, 11, 000, 001, 010, 100, 011, 110, 101, 111, ...\} =$
  $= \{\text{όλες οι ακολουθίες των } 0 \text{ και } 1\}$

ΠΡΑΞΕΙΣ ΣΕ ΣΥΜΒΟΛΟΣΕΙΡΕΣ

Ένωση συμβολοσειρών:

Για κάθε $w, z \in \Sigma^*$, ορίζεται η *ένωσή* τους $w \cup z$ ως η *γλώσσα* $\{w, z\}$.

Προσοχή: Συμβολίζουμε την ένωση $w \cup z$ ως *άθροισμα* $w + z$.

Συνένωση συμβολοσειρών:

Αν $w, z \in \Sigma^* \Rightarrow wz \in \Sigma^*$ με $|wz| = |w| + |z|$.

Άρα, $w\varepsilon = \varepsilon w = w$ με $|w\varepsilon| = |\varepsilon w| = |w|, \forall w \in \Sigma^*$.

Δυνάμεις συμβολοσειράς σε εκθέτη $n \in \mathbb{N}$:

Για κάθε $w \in \Sigma^*$, ορίζεται η *δύναμή* του $w^n \in \Sigma^*$ ως εξής:

$$w^n = \begin{cases} \varepsilon, & \text{για } n = 0, \\ ww^{n-1}, & \text{για } n \geq 1. \end{cases}$$

Παράδειγμα:

$\Sigma^* = \{0, 1\}, w = 011$. Τότε έχουμε:
$w^0 = \varepsilon, w^1 = w = 011, w^2 = ww = 011011,$
$w^3 = ww^2 = 011011011$ κ.ο.κ.

Αντίστροφη συμβολοσειρά:

Για κάθε $w \in \Sigma^*$, ορίζεται η *αντίστροφη συμβολοσειρά* (*ή λέξη*) $w^R \in \Sigma^*$ ως εξής:

$$w^R = \begin{cases} w, & \text{όταν } w = \varepsilon, \\ y^R a, & \text{όταν } w = ay, \text{για κάποιο } \alpha \in \Sigma \text{ και } y \in \Sigma^*. \end{cases}$$

Προφανώς, αν $w = uv$, με $u, v \in \Sigma^*$, δηλαδή, $|w| = 2$, τότε $w^R = vu$.

Παράδειγμα:

$w = 01101 = 0(1101)$
$\Rightarrow w^R = (1101)^R 0 = (101)^R 10 = (01)^R 110 = 10110$

Ορισμός: Μια συμβολοσειρά (ή λέξη) w λέγεται **παλίνδρομη** αν $w^R = w$.

Παράδειγμα:

Στο {0,1}, οι $\varepsilon, 0, 1, 11, 101, 010, 000, 111, ...$ είναι παλίνδρομες.

ΠΡΑΞΕΙΣ ΣΕ ΓΛΩΣΣΕΣ

Ένωση γλωσσών:

Αν L_1, L_2 είναι γλώσσες στο ίδιο αλφάβητο Σ, τότε η **ένωσή** τους $L_1 \cup L_2$ (που συμβολίζεται $L_1 + L_2$) είναι η γλώσσα $\{x, y \in \Sigma^* \colon x \in L_1, y \in L_2\}$.

Συνένωση γλωσσών:

Αν L_1, L_2 είναι γλώσσες στο ίδιο αλφάβητο Σ, τότε η **συνένωσή** τους $L_1 L_2$ είναι η γλώσσα $\{xy \in \Sigma^* \colon x \in L_1, y \in L_2\}$.

Παρατηρήσεις:

1. Αν L_1, L_2 είναι γλώσσες σε διαφορετικά αλφάβητα Σ_1, Σ_2 (αντιστοίχως), τότε οι $L_1 + L_2$ και $L_1 L_2$ είναι γλώσσες στο $\Sigma_1 \cup \Sigma_2$.
2. $\{\varepsilon\}L = L\{\varepsilon\} = L, \forall L \subset \Sigma^*$.

Παράδειγμα:

$$L_1 = \{10, 1\}, L_2 = \{011, 11\} \text{ (στο } \Sigma = \{0,1\}\text{)}$$
$$\Rightarrow L_1 L_2 = \{10011, 1011, 111\}$$

Δυνάμεις γλώσσας σε εκθέτη $n \in \mathbb{N}$:

Για κάθε γλώσσα $L \subset \Sigma^*$, ορίζεται η *δύναμή* της $L^n \subset \Sigma^*$ ως εξής:

$$L^n = \begin{cases} \{\varepsilon\}, & \text{για } n = 0, \\ LL^{n-1}, & \text{για } n \geq 1. \end{cases}$$

Παράδειγμα:

$$L = \{01\} \text{ (στο } \Sigma = \{0,1\}\text{)}$$
$$L^0 = \{\varepsilon\}, L^1 = L = \{01\}, L^2 = LL = \{0101\},$$
$$L^3 = LL^2 = \{010101\} \text{ κ.ο.κ.}$$

Αντιστροφή γλώσσας:

$$L \subset \Sigma^* \Rightarrow L^R \subset \Sigma^*, \text{ όπου } L^R = \left\{x^R : x \in L\right\}$$

Παρατηρήσεις: Αν L, L_1, L_2 είναι γλώσσες, τότε :

(i) $(L^R)^R = L$,
(ii) $(L_1 L_2)^R = L_2^R L_1^R$.

Συμπλήρωμα γλώσσας:

$$\overline{L} = \Sigma^* \setminus L.$$

Παράδειγμα: Έστω $\Sigma = \{0, 1, 2, \ldots, 9\}$ και L η γλώσσα που αποτελείται από όλες τις συμβολοσειρές χωρίς τα $2, 3, \ldots, 9$. Τότε $\overline{L}$ είναι η γλώσσα που αποτελείται από όλες τις λέξεις που περιέχουν ένα τουλάχιστον των $2, 3, \ldots, 9$.

Κλειστότητα Kleene γλώσσας:

Έστω L γλώσσα στο αλφάβητο Σ. Τότε η **κλειστότητα Kleene** ή **άστρο Kleene** της L ορίζεται ως

$$L^* = \bigcup_{n=0}^{+\infty} L^n$$

και η **θετική κλειστότητα** της L ως

$$L^+ = \bigcup_{n=1}^{+\infty} L^n.$$

Με άλλα λόγια, οι κλειστότητες L^*, L^+ αποτελούνται από όλες τις συμβολοσειρές που προκύπτουν από οποιεσδήποτε συνενώσεις των συμβολοσειρών της L.

Προφανώς, η L^* πάντα περιέχει την άδεια λέξη ε (διότι $n = 0$) και $\varepsilon \in L^+ \Leftrightarrow \varepsilon \in L$.

Παρατήρηση:

$$L \in \Sigma^* \Rightarrow L^n \subset \Sigma^*, \quad \forall n \in \mathbb{N}.$$

Άρα, και L^*, L^+ γλώσσες (δηλαδή, $L^*, L^+ \subset \Sigma^*$)

Πρόταση: Για κάθε γλώσσα $L \subset \Sigma^*$:

$$\begin{array}{ll} \text{(i)} & L^+ = LL^* = L^*L, \\ \text{(ii)} & (L^+)^+ = L^+, \\ \text{(iii)} & (L^*)^* = L^*. \end{array}$$

Παράδειγμα:

$$\begin{array}{l} L = \{a\} \ (a \in \Sigma) \\ L^0 = \{\varepsilon\}, L^1 = L = \{a\}, L^2 = LL = \{aa\}, L^3 = LL^2 = \{aaa\} \text{ κ.ο.κ.} \\ \Rightarrow \begin{cases} L^* = \{\varepsilon, a, a^2, a^3, \ldots\} = \sum\limits_{n=0}^{\infty} a^n \\ L^+ = \{a, a^2, a^3, \ldots\} = \sum\limits_{n=1}^{\infty} a^n \end{cases} \end{array}$$

Παράδειγμα:

$$\begin{array}{l} L = a + b \ \ (\Sigma = \{a, b\}) \\ L^0 = \{\varepsilon\}, L^1 = L = \{a, b\}, L^2 = LL = \{aa, ab, ba, bb\}, \\ L^3 = LL^2 = \{aaa, aab, aba, baa, abb, bab, bba, bbb\} \end{array}$$

κ.ο.κ.

Επομένως, $L^* = (a+b)^*$ είναι η γλώσσα όλων των συμβολοσειρών με a, b συμπεριλαμβανομένης της κενής ε ($L^* = \Sigma^*$) και $L^+ = (a+b)^+$ είναι η γλώσσα αυτή χωρίς την ε.

Παράδειγμα: (Η περίπτωση $L^+ = L^* = L$)

Έστω $\Sigma = \{0, 1, 2, ..., 9\}$ και L η γλώσσα με λέξεις χωρίς τα $2, 3, ..., 9$. Προφανώς, $\varepsilon \in L$.

Επίσης, $L^n \subset L, \forall n \in \mathbb{N}$ (διότι $x \in L^n \Rightarrow x = w_1 w_2 \cdots w_n$ χωρίς τα $2, 3, ..., 9 \Rightarrow x \in L$). Ακόμη $L \subset L^n, \forall n \in \mathbb{N}$ (διότι $x \in L \Rightarrow x = \varepsilon^{n-1}x$, αφού $\varepsilon \in L \Rightarrow x \in L^n$). Άρα, $L = L^n, \forall n \in \mathbb{N}$.

Επομένως, $L^+ = \bigcup_{n=1}^{+\infty} L^n = \bigcup_{n=1}^{+\infty} L = L$ και επίσης $L^* = L^0 \cup L^+ = L^0 \cup L = \{\varepsilon\} \cup L = L$. Άρα, $L^+ = L^* = L$.