22/10/2020

Θεωρία Αυτομάτων και Τυπικών γλωσσών

Ορισμός: Αλφάβητο είναι κάθε πεπερασμένο μη κενό σύνολο. Τα μέλη του ονομάζονται σύμβολα ή γράμματα.

π.χ.: $\Sigma_1 = \{0, 1\}$, $\Sigma_2 = \{a, b, \dots, \omega\}$

Συμβολοσειρά (string), ενός αλφαβήτου Σ είναι μια πεπερασμένη ακολουθία συμβόλων.

π.χ. Το 0101101 είναι συμβολοσειρά του αλφαβήτου $\Sigma = \{0, 1, 2\}$

※ Η (μοναδική) συμβολοσειρά μήκους 0, ονομάζεται κενή και συμβολίζεται με ε.

Το μήκος των συμβολοσειρών μήκους k συμβολίζεται με Σ^k.

π.χ. $\{0,1\}^2 = \{00, 01, 10, 11\}$

Συμβολισμός: Το σύνολο όλων των συμβολοσειρών του Σ συμβολίζεται με Σ^*.

π.χ. $\Sigma^* = \{0,1\}^* = \{\varepsilon, 0, 1, 00, 01, 10, 11, 000, \dots\}$

Πράξεις συμβολοσειρών:

▶ **Παράθεση:** Η παράθεση 2 συμβολοσειρών x, y είναι η συμβολοσειρά $x \cdot y$ που συμβολίζεται με xy ή $x \circ y$.

π.χ. $x = 011$ και $y = 1001$

τότε $x \circ y = 0111001$

Επανάληψη: Αν ω είναι μια συμβολοσειρά τότε η επανάληψη της είναι η $ω^k$ που αποτελείται από τη παράθεση k αντίγραφων του ω.

π.χ: $(01)^3 = 010101$.

Αντιστροφή: Αν ω είναι μια συμβολοσειρά τότε η αντιστροφή της είναι η $ω^R$ και προκύπτει αν διαβάσουμε την ω από το τέλος προς την αρχή

π.χ $01011^R = 11010$.

Ορισμός: Έστω Σ ένα αλφάβητο. Οποιοδήποτε υποσύνολο του $Σ^*$ ονομάζεται γλώσσα του Σ. και συμβολίζεται με L (ουσιαστικά είναι σύνολα)

Πράξεις με γλώσσες.

Τομή: $L_1 \cap L_2$

Ένωση: $L_1 \cup L_2$

Συμπλήρωμα: $\bar{L} = Σ^* - L$.

Παράθεση: $L_1 \circ L_2$ ή $L_1 L_2$

π.χ: Αν $L_1 = \{0, 1, 00\}$ και $L_2 = \{ε, 00\}$

$\Rightarrow L_1 \circ L_2 = \{0, 1, 00, 000, 100, 0000\}$

Επιμερισμοί:

$0 \cdot ε = 0$

$00ε = 00$ κ.ο.κ.

Kleene Star: Η Kleene Star L^* μιας γλώσσας L είναι η γλώσσα των συμβολοσειρών που προκύπτουν από την παράθεση μηδέν ή περισσότερων συμβολοσειρών της L:

$$L^* = \{ w : w = w_1 \circ w_2 \circ \dots \circ w_n, \text{ για } n \geq 0, \; w_1, \dots, w_n \in L \}$$

π.χ.: Αν $L = \{0, 11\}$ τότε

$L^* = \{ \varepsilon, 0, 11, 00, 110, 011, 000, 0000, 1100, 0011, 01100, 1111, \dots \}$

Αν $L = \{\varepsilon\}$ τότε $L^* = \{\varepsilon\}$

Αν $L = \{ \}$ τότε $L^* = \{\varepsilon\}$

Άρα $\forall L : \varepsilon \in L^*$.

→ Ορίζεται $L^+ = LL^*$.

Σημείωση: Υπάρχουν άπειρες συμβολοσειρές και άπειρες γλώσσες του Σ.

Ορισμός: i) Δύο σύνολα A, B θα λέγονται **ισάριθμα** αν υπάρχει αμφιμονοσήμαντη

$f: A \to B$.

ii) Ένα σύνολο θα λέγεται **πεπερασμένο** αν ~~υπάρχει~~ είναι ισάριθμο με το $\{1, 2, \dots, n\}$, $n \in \mathbb{N}$.
Δηλ. αν $\exists\, f: A \to \{1, 2, \dots, n\}$ η οποία είναι αμφιμονοσήμαντη.

iii) Ένα σύνολο λέγεται μετρήσιμα άπειρο αν είναι ισάριθμο του $\mathbb{N}$ και μετρήσιμο αν είναι πεπερασμένο ή μετρήσιμα άπειρο.

↳ π.χ. Το σύνολο των ζυγών
Το $\mathbb{Z}$.
Το $\mathbb{N} \times \mathbb{N}$ ...

Θεώρημα: Το σύνολο Σ^* των συμβολοσειρών ενός αλφαβήτου Σ είναι μετρήσιμο.

Απόδ

Εξ' ορισμού γνωρίζουμε ότι το αλφάβητο Σ είναι πεπερασμένο. Μπορούμε να δημιουργήσουμε μια ακολουθία που περιέχει όλες τις συμβολοσειρές του Σ σε λεξικογραφική σειρά ως εξής: 1η συμβ. μήκους 0, μετά μήκους 1, ... κ.ο.κ.

π.χ. Αν $\Sigma = \{0, 1\}$ η ακολουθία είναι

$\varepsilon, 0, 1, 00, 01, 10, 11, 000, \dots$

Θεώρημα: Το σύνολο των γλωσσών ενός αλφαβ. Σ είναι μη μετρήσιμο. (αποδ. : διαφάνεια)

(*) Υπάρχει γλώσσα L, της οποίας οι συμβολοσ. δεν μπορούν να τυπωθούν από πρόγραμμα (C++). Αυτό συμβαίνει διότι το σύνολο των προγραμμάτων είναι μετρήσιμο, ενώ το σύνολο των γλωσσών δεν είναι.

(*) Για νδο το σύνολο των γλωσσών δεν είναι μετρήσιμο, χρησιμοποιούμε μια μέθοδο που λέγεται διαγωνιοποίηση.

Μέθοδος διαγωνιοποίησης: Έστω R μια διμελής σχέση ενός συνόλου A, δηλαδή R είναι ένα υποσύνολο του καρτεσιανού γινομένου $A \times A$: $R \subseteq \{(a_1, a_2) : a_1, a_2 \in A\}$. Τότε το διαγώνιο σύνολο $\Delta = \{a : (a,a) \in R\}$ διαφέρει από κάθε γραμμή

$R_a = \{b : (a,b) \in R\}$

__π.χ.__ $A = \{1,2,3,4,5\}$ και

$R = \{(1,3), (2,2), (2,4), (2,6), (3,3), (3,5), (4,2), (4,5), (5,1), (5,2), (5,3), (5,5)\}$

	1	2	3	4	5
1			X		X
2		X		X	X
3			X		X
4		X			X
5	X	X	X		X

Οι γραμμές της R είναι

$R_1 = \{3, 5\}$

$R_2 = \{2, 4, 5\}$

$R_3 = \{3, 5\}$

$R_4 = \{2, 5\}$

$R_5 = \{1, 2, 3, 5\}$

Η διαγώνιος είναι το σύνολο $\Delta = \{1, 4\}$ 6

και είναι διαφορετικό από κάθε γραμμή

της R.