Κεφάλαιο 2

ΚΑΝΟΝΙΚΕΣ ΓΛΩΣΣΕΣ & ΠΕΠΕΡΑΣΜΕΝΑ ΑΥΤΟΜΑΤΑ

2.1 ΚΑΝΟΝΙΚΕΣ ΕΚΦΡΑΣΕΙΣ ΚΑΙ ΚΑΝΟΝΙΚΕΣ ΓΛΩΣΣΕΣ

Ορισμός: Έστω Σ αλφάβητο. Οι **κανονικές εκφράσεις** στο Σ ορίζονται ως εξής:

(1) Το $\emptyset$ είναι κανονική έκφραση.

(2) Το ε είναι κανονική έκφραση.

(3) Για κάθε $\alpha \in \Sigma$, το α είναι κανονική έκφραση.

(4) Αν r και s είναι κανονικές εκφράσεις, τότε και τα $r + s (= r \cup s), rs, r^*$ είναι κανονικές εκφράσεις.

(5) Τίποτε άλλο δεν είναι κανονική έκφραση εκτός εάν προκύπτει από τα (1) ως (4).

Σε κάθε κανονική έκφραση α αντιστοιχεί μια κανονική γλώσσα $L(\alpha)$ έτσι ώστε:

(1′) $L(\emptyset) = \emptyset$.

(2′) L(ε)=$\{\varepsilon\}$.

(3′) Για $\alpha \in \Sigma$, $L(\alpha) = \{\alpha\}$.

(4′) $L(r + s) = L(r) + L(s)$ $(+ = \cup)$,
$L(rs) = L(r)L(s)$,
$L(r^*) = L(r)^*$.

Συμβολισμοί:

(i) Συνήθως, εκτός αν δημιουργείται σύγχυση, η κανονική γλώσσα $L(r)$, που αντιστοιχεί στην κανονική έκφραση r, συμβολίζεται και αυτή με r.

(*ii*) Δεχόμαστε ότι το * έχει μεγαλύτερη προτεραιότητα από τη συνένωση και την ένωση. Και η συνένωση έχει μεγαλύτερη προτεραιότητα από την ένωση. Έτσι, η κανονική έκφραση $(r(s^*)+t)$ γράφεται rs^*+t. Επίσης, η έκφραση rr^* γράφεται r^+.

Οι κανονικές εκφράσεις χρησιμοποιούνται για την περιγραφή (κανονικών) γλωσσών, όπως φαίνεται στα παρακάτω παραδείγματα.

Παράδειγμα 1: Έστω η γλώσσα $L = c^*(a+bc^*)^*$. Ζητούμε να χαρακτηρίσουμε μια τυπική συμβολοσειρά της.

Προφανώς, η L είναι στο αλφάβητο $\Sigma = \{a, b, c\}$. Έχουμε τον παρακάτω πίνακα, στον οποίο φαίνεται ποιας μορφής συμβολοσειρές παράγονται από τις κανονικές εκφράσεις που περιέχονται στην γλώσσα του παραδείγματος:

Κανονική έκφραση	Παραγόμενες συμβολοσειρές της μορφής
bc^*	bc^n
$a+bc^*$	a, bc^n
$(a+bc^*)^*$	ε,a^p,$(bc^n)^m$, $a^p(bc^n)^m$, $(bc^n)^m a^p$
$c^*(a+bc^*)^*$	c^q, $c^q a^p$, $c^q(bc^n)^m$,$c^q a^p(bc^n)^m$, $c^q(bc^n)^m a^p$

Ελέγχουμε για το αν περιέχονται τα a, b, c, ab, bc, ac στη γλώσσα L:

a, b, c	ναι
ab	ναι ($a^p(bc^n)$ για $p = 1$)
bc	ναι (bc^n για $n = 1$)
ac	όχι

Άρα, οι συμβολοσειρές της L **δεν** περιέχουν το ac.

Παράδειγμα 2: Έστω η γλώσσα $L = (0+1)^*00(0+1)^*$. Ζητούμε να χαρακτηρίσουμε μια τυπική συμβολοσειρά της L.

Τώρα Σ=$\{0,1\}$. Η κανονική έκφραση $(0+1)^*$ παράγει όλες τις συμβολοσειρές των 0 και 1, ενώ η 00 παράγει μόνο την 00. Επομένως, όλες οι συμβολοσειρές της L περιέχουν δυο διαδοχικά 00.

Παράδειγμα 3: $L = (1+10)^*$.

Η γλώσσα αυτή περιέχει συμβολοσειρές με 1 και 10. Επομένως, όλες οι συμβολοσειρές της L αρχίζουν με 1 και δεν περιέχουν δυο διαδοχικά 0.

Παράδειγμα 4: $L = (0+\varepsilon)(1+10)^*$.

Επειδή η $0+\varepsilon$ περιέχει το 0, η L περιέχει όλες τις συμβολοσειρές των 0 και 1, που δεν έχουν δυο διαδοχικά 0.

Παράδειγμα 5: $L = (0+1)^*011$.

Όλες οι συμβολοσειρές των 0 και 1, που τελειώνουν σε 011.

Παράδειγμα 6: Να βρεθεί η γλώσσα L στην κανονική της έκφραση, όταν όλες οι συμβολοσειρές τελειώνουν σε 1 και δεν περιέχουν το 00.

$L = (0 + \varepsilon)(1 + 10)^*1$.

Αλλιώς: (ΠΡΟΣΟΧΗ: οι κανονικές εκφράσεις <u>ΔΕΝ</u> είναι μοναδικές)

$L = (1 + 01)^+$.

Παράδειγμα 7: Να βρεθεί η γλώσσα L με συμβολοσειρές που περιέχουν άρτιο αριθμό 0.

Προφανώς, $L = L_1^*$, όπου η L_1 περιέχει δυο 0. Άρα, ένα μέρος της L_1 είναι 01^*0, ενώ το άλλο μέρος δεν μπορεί να έχει 0. Έτσι, $L_1 = 1 + 01^*0 \Rightarrow L = (1 + 01^*0)^*$.

Παράδειγμα 8: Να βρεθεί η γλώσσα L, όταν όλες οι συμβολοσειρές περιέχουν ακριβώς δυο διαδοχικά 0 (00).

Προφανώς, $L = L_1 00 L_1$, όπου η L_1 δεν περιέχει διαδοχικά 0. Άρα, για παράδειγμα, το ένα μέρος της L_1 είναι το 10 και το άλλο το 1. Δηλαδή, $L_1 = 1 + 10 \Rightarrow L = (1 + 10)^*00(1 + 10)^*$.

Παράδειγμα 9: Να βρεθεί η γλώσσα L που δεν περιέχει το 11.

Αν $L = L_1^*$, προφανώς, κανένα μέρος της L_1 δεν μπορεί να είναι το 1 ή το 11. Άρα, $L_1 = 0 + 10 \Rightarrow L = (0 + 10)^*$.

Παράδειγμα 10: Να βρεθεί η γλώσσα L που δεν περιέχει το 111.

Η L μπορεί να είναι μια γλώσσα που δεν περιέχει το 11 (όπως στο προηγούμενο παράδειγμα) και τερματίζεται με 0. Δηλαδή, $L = (0 + 10)^*0^*$.

ΑΣΚΗΣΕΙΣ

1. Βρείτε τις κανονικές εκφράσεις της L, όταν οι συμβολοσειρές της L:

- **(a)** περιέχουν ακριβώς δυο 0,
- **(b)** περιέχουν τουλάχιστον δυο 0,
- **(c)** δεν περιέχουν ούτε το 00 ούτε το 11,
- **(d)** κάθε λέξη έχει άρτιο αριθμό συμβόλων και τουλάχιστον δυο.

2. Περιγράψτε τις γλώσσες που αντιστοιχούν στις ακόλουθες εκφράσεις:

- **(a)** $(11+0)^*(00+1)^*$,
- **(b)** $0^*1(0^*10^*1)^*0^*$,
- **(c)** $0^*1(0+1)^*$,
- **(d)** $(00+01+10+11)^*$.

ΑΠΑΝΤΗΣΕΙΣ

1(a) $L = 1^*01^*01^*$,

1(b) $L = (0+1)^*01^*0(0+1)^*$,

1(c) $L = (\varepsilon + 10)^*$,

1(d) $(00+01+10+11)^*$.

2(a) δεν περιέχει τα 1010 και 0101,

2(b) δεν περιέχει δυο διαδοχικά 0 (00) εκτός αν τελειώνει σε 0,

2(c) αρχίζει με 0^n1,

2(d) όμοια με $1(d)$.