Φροντιστήριο 5 – Λύσεις

Άσκηση 1

Να αποφασίσετε ποιες από τις πιο κάτω γλώσσες είναι κανονικές αιτιολογώντας πλήρως τις απαντήσεις σας.

(α) $L_1 = \{a^n b^{2n} \mid n \geq 0\}$

(β) $L_2 = \{w \mid w \in \{0, 1\}^*$ και τα δύο πρώτα σύμβολα της w είναι ανόμοια$\}$

(γ) $L_3 = \{w^{rev} \mid w = a^n b^m, n, m \geq 0\}$

(δ) $L_4 = \{xwx^{rev}y \mid x, w, y \in \{a, b\}^+\}$

(ε) $L_5 = \{a^n b^{n+m} c^m \mid m \leq 2, n \geq 0\}$

(ζ) $L_6 = \{0^n \mid$ ο n είναι πρώτος αριθμός$\}$

(η) $\{w \mid$ η w περιέχει τις υπολέξεις 01 και 10 τις ίδιες φορές$\}$

Λύση

(α) $L_1 = \{a^n b^{2n} \mid n \geq 0\}$

Υποθέτουμε για να φτάσουμε σε αντίφαση ότι η L_1 είναι κανονική. Τότε, σύμφωνα με το Λήμμα της Άντλησης, υπάρχει p, το μήκος άντλησης της γλώσσας, τέτοιο ώστε κάθε λέξη της γλώσσας με μήκος μεγαλύτερο από p να ικανοποιεί την ιδιότητα που περιγράφεται στο Λήμμα.

Ας επιλέξουμε τη λέξη $w = a^p b^{2p}$.

Τότε, σύμφωνα με το λήμμα, $w = xyz$ έτσι ώστε η υπολέξη xy να βρίσκεται μέσα στις p πρώτες θέσεις της w ($|xy| \leq p$) η y να είναι μη κενή ($|y| \neq 0$) και οποιαδήποτε επανάληψη της υπολέξης y να διατηρεί την προκύπτουσα λέξη εντός της γλώσσας ($xy^i z \in L_1$, $i \geq 0$).

Αφού $|xy| \leq p$, τότε πρέπει να ισχύει ότι τόσο το x όσο και το y αποτελούνται μόνο από a. Επομένως, $x = a^\lambda$, $y = a^\mu$, $w = a^\nu b^{2p}$ όπου $\lambda+\mu+\nu = p$.

Επίσης, από το λήμμα, πρέπει να ισχύει ότι $xy^2z \in L_1$.

Αλλά, $xy^2z = a^\lambda a^{2\mu} a^\nu b^{2p} = a^{\lambda+2\mu+\nu} b^{2p} = a^{p+\mu} b^{2p}$ το οποίο, από τον ορισμό της γλώσσας δεν ανήκει στην L_1.

Αυτό μας οδηγεί σε αντίφαση και επομένως η υπόθεσή μας ότι η γλώσσα L_1 είναι κανονική ήταν εσφαλμένη.

Συμπέρασμα: Η L_1 είναι μη κανονική.

(β) $L_2 = \{w \mid w \in \{0, 1\}^*$ και τα δύο πρώτα σύμβολα της w είναι ανόμοια$\}$

Η γλώσσα είναι κανονική. Για να το δείξουμε αρκεί να παρουσιάσουμε μια κανονική έκφραση η οποία να περιγράφει την γλώσσα. Η έκφραση αυτή είναι η

$$01\{0, 1\}^* \cup 10\{0, 1\}^*$$

(γ) $L_3 = \{w^{rev} \mid w = a^n b^m, n, m \geq 0\}$

Η γλώσσα είναι κανονική. Για να το δείξουμε αρκεί να παρουσιάσουμε μια κανονική έκφραση η οποία να περιγράφει την γλώσσα. Η έκφραση αυτή είναι η

$$b^* a^*$$

(δ) $L_4 = \{xwx^{rev}y \mid x, w, y \in \{a, b\}^+\}$

Η γλώσσα είναι κανονική. Για να το δείξουμε αρκεί να παρουσιάσουμε μια κανονική έκφραση η οποία να περιγράφει την γλώσσα. Η έκφραση αυτή είναι η

$$1\{0, 1\}^+1\{0, 1\}^+ \cup 0\{0, 1\}^+0\{0, 1\}^+$$

(ε) $L_5 = \{a^n b^{n+m} c^m \mid m \leq 2, n \geq 0\}$

Υποθέτουμε για να φτάσουμε σε αντίφαση ότι η L_5 είναι κανονική. Τότε, σύμφωνα με το Λήμμα της Άντλησης, υπάρχει p, το μήκος άντλησης της γλώσσας, τέτοιο ώστε κάθε λέξη της γλώσσας με μήκος μεγαλύτερο από p να ικανοποιεί την ιδιότητα που περιγράφεται στο Λήμμα.

Ας επιλέξουμε τη λέξη $w = a^p b^p$.

Τότε, σύμφωνα με το λήμμα, $w = xyz$ έτσι ώστε η υπολέξη xy να βρίσκεται μέσα στις p πρώτες θέσεις της w ($|xy| \leq p$) η y να είναι μη κενή ($|y| \neq 0$) και οποιαδήποτε επανάληψη της υπολέξης y να διατηρεί την προκύπτουσα λέξη εντός της γλώσσας ($xy^iz \in L_5$).

Αφού $|xy| \leq p$, τότε πρέπει να ισχύει ότι τόσο το x όσο και το y αποτελούνται μόνο από a. Επομένως, $x = a^\lambda$, $y = a^\mu$, $w = a^\nu b^p$ όπου $\lambda+\mu+\nu = p$.

Επίσης, από το λήμμα, πρέπει να ισχύει ότι $xy^2z \in L_1$.

Αλλά, $xy^2z = a^\lambda a^{2\mu} a^\nu b^p = a^{\lambda+2\mu+\nu} b^p = a^{p+\mu} b^p$ και, από τον ορισμό της γλώσσας, $xy^2z \notin L_5$.

Αυτό μας οδηγεί σε αντίφαση και επομένως η υπόθεσή μας ότι η γλώσσα L_5 είναι κανονική ήταν εσφαλμένη.

Συμπέρασμα: Η L_5 είναι μη κανονική.

(ζ) $L_6 = \{0^n \mid$ ο n είναι πρώτος αριθμός$\}$

Υποθέτουμε για να φτάσουμε σε αντίφαση ότι η L_1 είναι κανονική. Τότε, σύμφωνα με το Λήμμα της Άντλησης, υπάρχει p, το μήκος άντλησης της γλώσσας, τέτοιο ώστε κάθε λέξη της γλώσσας με μήκος μεγαλύτερο από p να ικανοποιεί την ιδιότητα που περιγράφεται στο Λήμμα.

Ας επιλέξουμε τη λέξη $w = 0^q$ όπου q είναι ένας πρώτος αριθμός μεγαλύτερος από τον p.

Τότε, σύμφωνα με το λήμμα, $w = xyz$ έτσι ώστε η υπολέξη xy να βρίσκεται μέσα στις p πρώτες θέσεις της w ($|xy| \leq p$) η y να είναι μη κενή ($|y| \neq 0$) και οποιαδήποτε επανάληψη της υπολέξης y να διατηρεί την προκύπτουσα λέξη εντός της γλώσσας ($xy^iz \in L_{16}$).

Έστω, $x = 0^\lambda$, $y = 0^\mu$, $w = 0^\nu$ όπου $\lambda+\mu+\nu = q$.

Από το λήμμα, πρέπει να ισχύει ότι $xy^{(\lambda+\nu)}z \in L_6$.

Αλλά, $xy^{(\lambda+\nu)}z = 0^\lambda 0^{\mu(\lambda+\nu)} 0^\nu = 0^{\lambda+\mu(\lambda+\nu)+\nu} = 0^{(\mu+1)(\lambda+\nu)}$ και, προφανώς ο αριθμός $(\mu+1)(\lambda+\nu)$ δεν είναι πρώτος αριθμός. Επομένως η λέξη $xy^{(\lambda+\nu)}z$ δεν ανήκει στη γλώσσα L_6.

Αυτό μας οδηγεί σε αντίφαση και επομένως η υπόθεσή μας ότι η γλώσσα L_6 είναι κανονική ήταν εσφαλμένη.

Συμπέρασμα: Η L_6 είναι μη κανονική.

(η) L_7 = {*w* | η *w* περιέχει τις υπολέξεις 01 και 10 τις ίδιες φορές}

Λέξεις της L_7 περιλαμβάνουν τις 101, 1001, 1001001, 010, 010110, 1, 0, ε.

Πιο γενικά, παρατηρούμε ότι αν μια λέξη ξεκινά με 1, ανήκει στη γλώσσα μόνο αν τελειώνει και με 1. Μόνο τότε εναλλάσσεται από 1 σε 0 και από 0 σε ένα τον ίδιο αριθμό φορών. Παρόμοια, αν μια λέξη ξεκινά με 0, ανήκει στη γλώσσα μόνο αν τελειώνει και με 0. Αυτό οφείλεται στο γεγονός ότι η υπολέξη 01 εμφανίζεται σε μία συμβολοσειρά στα σημεία όπου μια ομάδα από συνεχόμενα 0 ακολουθείται από μια ομάδα από συνεχόμενα 1 ενώ η υπολέξη 10 εμφανίζεται σε μία συμβολοσειρά στα σημεία όπου μια ομάδα από συνεχόμενα 1 ακολουθείται από μια ομάδα από συνεχόμενα 0. Επομένως για να είναι το πλήθος των εμφανίσεων 10 και 01 ίσος σε μια λέξη πρέπει η λέξη να έχει μια από τις πιο κάτω μορφές:

- $1^+0^+1^+\ldots0^+1^+$, ή
- $0^+1^+0^+\ldots1^+0^+$

Αυτή η περιγραφή μπορεί να διατυπωθεί μέσω της κανονικής έκφρασης:

$\varepsilon \cup 1 \cup 0 \cup 1\{0,1\}^*1 \cup 0\{1,0\}^*0$.

Κατά συνέπεια, η γλώσσα L_7 είναι κανονική.

<u>Άσκηση 2</u>

Να αποδείξετε ότι αν η γλώσσα L είναι κανονική τότε και η γλώσσα L^{rev} είναι κανονική.

<u>Λύση</u>

Για να δείξουμε το ζητούμενο υποθέτουμε ότι M είναι ένα NFA αυτόματο που αναγνωρίζει τη γλώσσα L και θα το μετατρέψουμε σε NFA που αναγνωρίζει τη γλώσσα L^{rev}. Αφού υπάρχει αυτόματο που αναγνωρίζει την L^{rev} η γλώσσα αυτή είναι κανονική.

Ας υποθέσουμε ότι το αυτόματο M = (Q, Σ, δ, q, F) αναγνωρίζει τη γλώσσα A. Κατασκευάζουμε το N = (Q', Σ, δ', q', F') ως εξής:

- $Q' = Q \cup$ {start}
- Σ: το αλφάβητο είναι το ίδιο με αυτό του M
- Για κάθε $r_1, r_2 \in Q$ και $a \in \Sigma$, θέτουμε
  $r_2 \in \delta'(r_1, a)$ αν και μόνο αν $r_1 \in \delta(r_2, a)$
  και επιπλέον
  $\delta'(\text{start}, \varepsilon) = F$
- q' = start
- $F' = \{q\}$

Επομένως, η κατασκευή μας εισάγει μία νέα κατάσταση, την start η οποία είναι η αρχική κατάσταση του αυτομάτου και η οποία οδηγείται μέσω ε-μεταβάσεων στις παλιές τελικές ($\delta'(\text{start}, \varepsilon) = F$). Από αυτές, τις παλιές τελικές καταστάσεις, ξεκινούμε να διαβάζουμε λέξεις κινούμενοι αντίστροφα πάνω στο αρχικό αυτόματο, M. Γι αυτό και η συνάρτηση δ' αντιστρέφει τις μεταβάσεις του παλιού αυτομάτου. Αναγνωρίζουμε μια λέξη μόνο εάν μας οδηγήσει στην αρχική κατάσταση του M, γεγονός που μας οδηγεί στο συμπέρασμα ότι έχουμε διαβάσει μια αντιστραμμένη λέξη από τη γλώσσα του M.

Αυτή τη διαισθητική αιτιολόγηση της κατασκευής του N μπορούμε να την τεκμηριώσουμε και με μία αυστηρή απόδειξη, δείχνοντας ότι

$w \in L(M)$ αν και μόνο αν $w^{rev} \in L(N)$.

Ας υποθέσουμε λοιπόν ότι $w = w_1w_2\ldots w_n \in L(M)$. Τότε, υπάρχει ακολουθία καταστάσεων $r_0, r_1,\ldots, r_n$ του Q που ικανοποιεί τις συνθήκες:

1. $r_0 = q$
2. $r_{i+1} \in \delta(r_i, w_{i+1})$ για $i = 0,\ldots,n-1$, και
3. $r_n \in F$

Αν θεωρήσουμε το μονοπάτι

start, $r_n,\ldots, r_1, r_0$

στο αυτόματο N, παρατηρούμε ότι

1. ξεκινά από την αρχική κατάσταση start
2. $r_n \in \delta(start, \varepsilon)$, $r_{i-1} \in \delta(r_i, w_i)$, για $i = 1,\ldots,n$, και
3. Αφού $r_0 = q$ τότε $r_0 \in F'$.

Αυτό συνεπάγεται ότι $w^{rev} \in L(N)$.

Αντιστρέφοντας τα πιο πάνω επιχειρήματα, λαμβάνουμε ότι αν $w^{rev} \in L(N)$ τότε $w \in L(M)$, και το ζητούμενο έπεται.

Άσκηση 3

Έστω L_1 και L_2 δύο μη κανονικές γλώσσες. Ισχύει απαραίτητα ότι οι πιο κάτω γλώσσες είναι μη κανονικές;

(α) $L_1 \cup L_2$

(β) $L_1 \cap L_2$

(γ) $L_1 L_2$

Λύση

(α) Έστω L_1 και L_2 δύο μη κανονικές γλώσσες. Τότε δεν ισχύει απαραίτητα ότι η γλώσσα $L_1 \cup L_2$ είναι μη κανονική. Για παράδειγμα, ενώ οι γλώσσες

$$L_1 = \{0^\alpha 1^\beta \mid \alpha = \beta\} \text{ και } L_2 = \{0^\alpha 1^\beta \mid \alpha \neq \beta\},$$

είναι μη κανονικές, η γλώσσα $L_1 \cup L_2 = 0^*1^*$ είναι κανονική (αφού μπορούμε να την περιγράψουμε ως κανονική έκφραση).

(β) Έστω L_1 και L_2 δύο μη κανονικές γλώσσες. Τότε δεν ισχύει απαραίτητα ότι η γλώσσα $L_1 \cap L_2$ είναι μη κανονική. Για παράδειγμα, ενώ οι γλώσσες

$$L_1 = \{0^\alpha 1^\beta \mid \alpha = \beta\} \text{ και } L_2 = \{0^\alpha 1^\beta \mid \alpha \neq \beta\},$$

είναι μη κανονικές, η γλώσσα $L_1 \cap L_2 = \emptyset$ είναι κανονική (αφού μπορούμε να την περιγράψουμε ως κανονική έκφραση).

(γ) Έστω L_1 και L_2 δύο μη κανονικές γλώσσες. Τότε δεν ισχύει απαραίτητα ότι η γλώσσα L_1L_2 είναι μη κανονική. Για παράδειγμα, ενώ οι γλώσσες

$$L_1 = \varepsilon \cup \{0^p \mid \text{ο p πρώτος αριθμός}\} \text{ και } L_2 = \varepsilon \cup \{0^q \mid \text{ο q δεν είναι πρώτος αριθμός}\},$$

είναι μη κανονικές, η γλώσσα $L_1L_2 = 0^*$ είναι κανονική αφού μπορούμε να την περιγράψουμε ως κανονική έκφραση.