

Ερώτημα 1

(A) Θεωρήστε το αλφάβητο $\Sigma=\{a,b,c\}$. Δημιουργήστε αντίστοιχες κανονικές εκφράσεις για τις παρακάτω τρεις γλώσσες:

(1) Όλες οι συμβολοσειρές που ΔΕΝ περιέχουν ΚΑΙ τα 3 σύμβολα ΚΑΙ αρχίζουν και τελειώνουν με το ίδιο σύμβολο. Λέξεις που ανήκουν στην γλώσσα είναι: aba, cbcbcb, bccb, ενώ λέξεις που δεν ανήκουν στην γλώσσα είναι: abca, cbc, bccba.

(2) Όλες οι συμβολοσειρές που ΔΕΝ αρχίζουν με δύο διαδοχικά ίδια σύμβολα. Λέξεις που ανήκουν στην γλώσσα είναι: abbc, a, ε, ενώ λέξεις που δεν ανήκουν στην γλώσσα είναι: aa, bbac, ccca.

(3) Όλες οι συμβολοσειρές που περιέχουν το πολύ δύο **συνεχόμενα** a. Λέξεις που ανήκουν στην γλώσσα είναι: baccaab, aacccacabbaa, ενώ λέξεις που δεν ανήκουν στην γλώσσα είναι: baccaaab, aaccabaaa.

(B) Περιγράψτε με φυσική γλώσσα τις συμβολοσειρές του αλφάβητου $\Sigma=\{0,1\}$ που περιγράφονται απ' κάθε μία από τις ακόλουθες τρεις εκφράσεις: (i) $(1+01+001)^*$, (ii) $(1^*0 + 1^*00)^*$, (iii) $(1+01+001)^* + (0+00)(1^*0 + 1^*00)^*$.

Σημείωση: Να αποφύγετε απλή περιγραφή των πράξεων μεταξύ των τελεστών με λόγια, όπως λχ στο (i): «όλες οι λέξεις με παράθεση των 1, 01, 001». Αντίθετα, η φυσική γλώσσα πρέπει να περιγράφει την ουσιαστική δομή της εκάστοτε γλώσσας.

Ενδεικτική Επίλυση

(A.1) Οι λέξεις: a, b, c αρχίζουν και τελειώνουν με το ίδιο σύμβολο και είναι φανερό ότι η έκφραση $a+b+c$ παράγει λέξεις μήκους 1 και ικανοποιεί τους περιορισμούς. Για τις λέξεις μήκους τουλάχιστον 2 εργαζόμαστε ως εξής. Οι λέξεις που δεν περιέχουν το c εκφράζονται από την $(a+b)^*$, συνεπώς μπορούν να έχουν είτε το a με την έκφραση $a(a+b)^*a$, είτε το b με την έκφραση $b(a+b)^*b$, ώστε να εμφανίζετε αντίστοιχα ως το ίδιο σύμβολο αρχής και τέλους της λέξης. Όμοια τις λέξεις που δεν περιέχουν το b της εκφράζει η $(a+c)^*$ και πρέπει να έχουν είτε το a με την έκφραση $a(a+c)^*a$, είτε το c με την έκφραση $c(a+c)^*c$, ώστε να εμφανίζετε αντίστοιχα ως το ίδιο σύμβολο αρχής και τέλους της λέξης. Συνολικά έχουμε την έκφραση: $a + b + c + a(a+b)^*a + b(a+b)^*b + a(a+c)^*a + c(a+c)^*c + c(c+b)^*c + b(c+b)^*b$.

(A.2) Οι λέξεις μήκους 0 είναι αποδεκτές και είναι η κενή συμβολοσειρά ε. Οι λέξεις μήκους 1 είναι όλες αποδεκτές και παράγονται από την έκφραση $a+b+c$. Για τις λέξεις μήκους τουλάχιστον 2 πρέπει να αποκλείσουμε όσες έχουν στην αρχή 2 διαδοχικά ίδια σύμβολα. Έτσι οι επιτρεπτές λέξεις θα δίνονται από την έκφραση $(ab + ba + ac + ca + bc +$

cb) που περιγράφει την αρχή κάθε λέξης, ενώ μετά παραθέτουμε οποιαδήποτε λέξη παραγόμενη από την $(a+b+c)^*$. Τελικά έχουμε την:

$$\varepsilon + a + b + c + (ab + ba + ac + ca + bc + cb) (a+b+c)^*.$$

(A.3) Η έκφραση $b + c$ δεν επιφέρει περιορισμό στις εμφανίσεις των b, c μέσα στην λέξη. Όμως κάθε εμφάνιση του a πρέπει να ικανοποιεί τον περιορισμό της εκφώνησης. Η έκφραση $ab + ac$ εξασφαλίζει ότι θα έχουμε ακριβώς 1 εμφάνιση του a κάθε φορά που θα το επιλέγουμε ως επόμενο σύμβολο της λέξης. Αντίστοιχα η έκφραση $aab + aac$ ότι θα έχουμε 2 διαδοχικές εμφανίσεις του a . Συνεπώς η έκφραση $(b + c + ab + ac + aab + aac)^*$ παράγει λέξεις με το πολύ 2 διαδοχικές εμφανίσεις του a . Όμως όλες αυτές τελειώνουν με το σύμβολο b ή c . Θα πρέπει να συμπεριλάβουμε και τις λέξεις με τέλος a ή aa που επίσης είναι επιτρεπτές. Συνεπώς η έκφραση $(b + c + ab + ac + aab + aac)^*(a + aa + \varepsilon)$ είναι η ζητούμενη.

(B.i) Η γλώσσα αποτελείται από την κενή λέξη ε , επιπλέον από τις λέξεις όπου κάθε τμήμα τους αποτελούμενο από συνεχόμενες εμφανίσεις του 0 έχει μήκος το πολύ 2, και επιπρόσθετα, στο τέλος αυτών των λέξεων (μήκους ≥ 1) έχουμε 1.

(B.ii) Η γλώσσα αποτελείται από την κενή λέξη ε , επιπλέον από τις λέξεις με αρχή 1 όπου κάθε τμήμα τους αποτελούμενο από συνεχόμενες εμφανίσεις του 0 έχει μήκος το πολύ 2, και στο τέλος αυτών των λέξεων (μήκους ≥ 1) έχουμε 0.

(B.iii) Η γλώσσα αποτελείται από την κενή λέξη ε , επιπλέον από τις λέξεις όπου κάθε τμήμα τους αποτελούμενο από συνεχόμενες εμφανίσεις του 0 έχει μήκος το πολύ 2, τέλος, οι μη κενές λέξεις της γλώσσας δεν μπορεί να αρχίζουν με 1 και να τελειώνουν με 0.

Ερώτημα 2

(A) Συμβολίζουμε με $(w)_c$ το πλήθος των εμφανίσεων του συμβόλου c στη συμβολοσειρά w . Για παράδειγμα $(abbbabab)_a = 3$. Έστω η γλώσσα

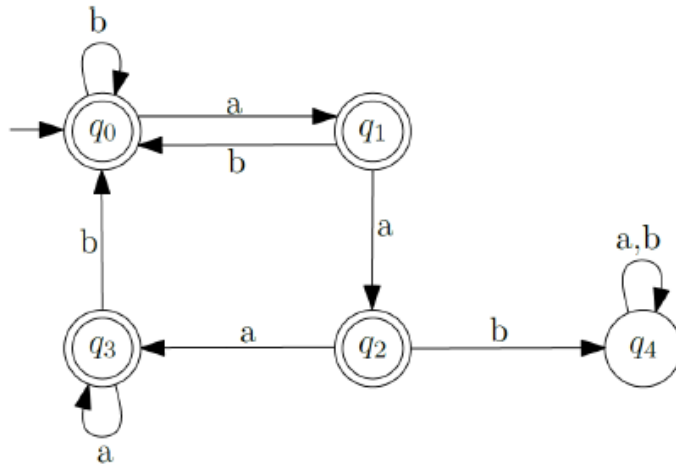
$$L = \{w \in \{a,b\}^* \mid -2 \leq (u)_a - (u)_b \leq 2 \text{ για κάθε πρόθεμα } u \text{ της } w \text{ και } |(w)_a - (w)_b| = 1\}$$

Δηλαδή, μια συμβολοσειρά w ανήκει στην L αν και μόνο αν σε κάθε αρχικό τμήμα της το πλήθος των a και το πλήθος των b διαφέρουν κατά απόλυτη τιμή το πολύ 2 και επιπλέον στη w το πλήθος των a είναι κατά ένα μεγαλύτερο ή κατά ένα μικρότερο από το πλήθος των b .

Για παράδειγμα, οι συμβολοσειρές $aabbbba$, $baabaab$, a και aab ανήκουν στην L , ενώ οι συμβολοσειρές $aabaabb$ και $abababab$ δεν ανήκουν στην L .

Σχεδιάστε ένα αιτιοκρατικό πεπερασμένο αυτόματο που να αναγνωρίζει την L .

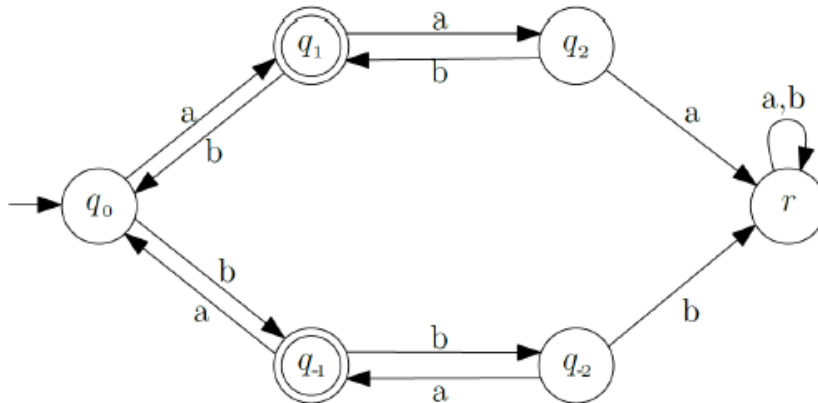
(B) Έστω το παρακάτω αιτιοκρατικό πεπερασμένο αυτόματο M :



Για κάθε κατάσταση q_i του M , $i \in \{0,1,2,3\}$, δώστε μια κανονική έκφραση R_i που να περιγράφει το σύνολο των συμβολοσειρών $\{w \mid \delta(q_0, w) = q_i\}$, δηλαδή το σύνολο όλων των συμβολοσειρών οι οποίες οδηγούν το αυτόματο M από την αρχική του κατάσταση q_0 στην κατάσταση q_i .

Ενδεικτική επίλυση

(A) Παρακάτω δίνεται το διάγραμμα καταστάσεων ενός αιτιοκρατικού πεπερασμένου αυτόματου που αναγνωρίζει τη γλώσσα L :



Τα αυτόματο χρησιμοποιεί τις καταστάσεις $q_0, q_1, q_2, q_{-1}, q_{-2}$ για να θυμάται τη διαφορά $(u)_a - (u)_b$ για το πρόθεμα u της εισόδου που έχει διαβάσει μέχρι την τρέχουσα χρονική στιγμή, δηλαδή για το πρόθεμα της εισόδου που έχει ως τελευταίο σύμβολο το πιο πρόσφατο σύμβολο που έχει διαβάσει. Αν σε κάποια από τις καταστάσεις q_{-2}, q_{-1}, q_0 ή q_1 διαβάσει το σύμβολο a , τότε μεταβαίνει αντίστοιχα στην κατάσταση q_{-1}, q_0, q_1 ή q_2 (η διαφορά αυξάνεται κατά 1). Αν σε κάποια από τις καταστάσεις q_{-1}, q_0, q_1 ή q_2 διαβάσει το σύμβολο b , τότε μεταβαίνει αντίστοιχα στην κατάσταση q_{-2}, q_{-1}, q_0 ή q_1 (η διαφορά

μειώνεται κατά 1). Αν ωστόσο στην κατάσταση q_2 διαβάσει a ή στην κατάσταση q_{-2} διαβάσει b , τότε μεταβαίνει στην κατάσταση καταβόθρα r (η οποία δεν είναι κατάσταση αποδοχής και στην οποία θα διαβάσει όλα τα υπόλοιπα σύμβολα της εισόδου), καθώς έχει εντοπίσει ένα πρόθεμα u της εισόδου για το οποίο ισχύει $|(u)_a - (u)_b| = 3$. Το αυτόματο αποδέχεται την είσοδο w αν τερματίσει σε κάποια από τις καταστάσεις q_{-1} ή q_1 , περίπτωση στην οποία ισχύει $|(w)_a - (w)_b| = 1$.

(B) Οι ζητούμενες κανονικές εκφράσεις είναι οι παρακάτω:

$$R0 = (b + ab + aaaa*b)^*$$

$$R1 = (b + ab + aaaa*b)^*a$$

$$R2 = (b + ab + aaaa*b)^*aa$$

$$R3 = (b + ab + aaaa*b)^*aaaa*$$

Παρατίθεται εδώ (αν και δεν απαιτείται) και η κανονική έκφραση των συμβολοσειρών που οδηγούν το αυτόματο M στην κατάσταση q^4 , για καθαρά εκπαιδευτικούς λόγους.

$$R4 = (b + ab + aaaa*b)^*aab(a+b)^*$$

Εξήγηση:

R_0 : Το αυτόματο ξεκινάει τη λειτουργία του στην κατάσταση q_0 . Από την q_0 μπορεί να μεταβεί ξανά στον εαυτό της, περνώντας ενδεχομένως και από κάποιες ενδιάμεσες καταστάσεις διαφορετικές της q_0 , με έναν από τους παρακάτω τρόπους:

- Ακολουθώντας το βρόχο που υπάρχει στην q_0 , αν διαβάσει το σύμβολο b .
- Μεταβαίνοντας στην q_1 και επιστρέφοντας, αν διαβάσει την ακολουθία συμβόλων ab .
- Μέσω των καταστάσεων q_1 , q_2 , q_3 αν διαβάσει μια ακολουθία συμβόλων που ξεκινάει με τρία ή περισσότερα σύμβολα a ακολουθούμενα από ένα σύμβολο b , δηλαδή μια ακολουθία συμβόλων της μορφής $aaaa*b$.

Οι παραπάνω μεταβάσεις μπορεί να επαναληφθούν καμία, μία ή περισσότερες φορές, με την ανάγνωση μιας συμβολοσειράς που αποτελεί την συναρμογή ενός οποιουδήποτε πλήθους συμβολοσειρών της μορφής b , ab ή $aaaa*b$.

$R1$: Το αυτόματο μπορεί να φτάσει στην $q1$ μόνο μέσω της $q0$, αν διαβάσει σε αυτή το σύμβολο a . Συνεπώς οι συμβολοσειρές που μπορούν να οδηγήσουν το αυτόματο από την αρχική κατάσταση $q0$ στην $q1$ είναι της μορφής ua , όπου η συμβολοσειρά u οδηγεί το αυτόματο από την $q0$ στον εαυτό της. Άρα η $R1$ προκύπτει από τη συναρμογή της $R0$ με το a .

R_2 : Με το ίδιο σκεπτικό, η R_2 προκύπτει από τη συναρμογή της R_1 με το a .

R_3 : Το αυτόματο μπορεί να εισέλθει στην q_3 μόνο μέσω της q_2 , αν διαβάσει σε αυτή το σύμβολο a . Στην q_3 παραμένει όσο διαβάζει σύμβολα a . Συνεπώς οι συμβολοσειρές που μπορούν να οδηγήσουν το αυτόματο από την αρχική κατάσταση q_0 στην q_3 είναι της μορφής $va a^*$, όπου η συμβολοσειρά v οδηγεί το αυτόματο από την q_0 στην q_2 . Άρα η R_3 προκύπτει από τη συναρμογή της R_2 με το aa^* .

R_4 : Το αυτόματο μπορεί να εισέλθει στην q_4 μόνο μέσω της q_2 , αν διαβάσει σε αυτή το σύμβολο b . Στην q_4 παραμένει μέχρι το τέλος της λειτουργίας του διαβάζοντας το υπόλοιπο της εισόδου, το οποίο μπορεί να είναι οποιαδήποτε συμβολοσειρά από το $\{a,b\}^*$. Συνεπώς οι συμβολοσειρές που μπορούν να οδηγήσουν το αυτόματο από την αρχική κατάσταση q_0 στην q_4 είναι της μορφής $wb(a+b)^*$, όπου η συμβολοσειρά w οδηγεί το αυτόματο από την q_0 στην q_2 . Άρα η R_4 προκύπτει από τη συναρμογή της R_2 με την $b(a+b)^*$.