ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

Θεωρία Υπολογισμού

Σημειώσεις - Αυτόματα

Αλέξιος Καπόρης

Τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο

ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

ΕΣΠΑ
2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Αιγαίου**» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Εστω $\Sigma = \{0,1\}$.

(1) Δείξτε ότι οποιοδήποτε $x \in \Sigma^*$ ανήκει στη γλώσσα L_1 που αντιστοιχεί στην κανονική έκφραση $0^*(10^*)^*$.

(2) Δείξτε ότι η γλώσσα $L \subseteq \Sigma^*$ που αποτελείται από τις συμβολοσειρές με αριθμό μονάδων που είναι πολλαπλάσιο του $k=3$ είναι κανονική. Μπορεί να γενικευτεί η πρόταση για οποιοδήποτε φυσικό αριθμό k?

(3) Περιγράψτε τις γλώσσες που αντιστοιχούν στις παρακάτω κανονικές εκφράσεις:

(A) $(00+11+01+10)^*$

(B) $\Sigma^*100\Sigma^*$

(Γ) $(1+0^*)\varnothing$

(Δ) $((\varnothing + 1^*)^*)^*$

(E) $\varnothing^*$

(4) Δώστε τις κανονικές εκφράσεις που αντιστοιχούν στις παρακάτω γλώσσες:

(A) $\{x \in \Sigma^* \mid$ το μήκος της x είναι πολλαπλάσιο του 5$\}$

(B) $\{x \in \Sigma^* \mid$ η x τελειώνει σε 01 ή σε 11$\}$

(Γ) $\{ x \in \Sigma^* \mid$ η x περιέχει τουλάχιστον ένα 1$\}$

(Δ) $\{ x \in \Sigma^* \mid$ η x δεν περιέχει ούτε το 111, ούτε το 000$\}$

(E) $\{ x \in \Sigma^* \mid$ η x περιέχει τη συμβολοσειρά 111000$\}$

Απάντηση

(1) Η απόδειξη είναι με επαγωγή στο μήκος συμβολοσειρών στη γλώσσα Σ^*. Αν $|x|=0$, τότε $x = \in$, η κενή συμβολοσειρά, και ανήκει στη γλώσσα L_1. Αν $|x| > 0$, τότε $x = a$y και υποθέτουμε επαγωγικά ότι η y ανήκει στη γλώσσα L_1. Αρκεί να δείξουμε ότι $y \in L_1 \Rightarrow ay \in L_1$, για a = 0,1. Η περίπτωση 0y είναι προφανής, αφού οι συμβολοσειρές της L_1 αρχίζουν με πεπερασμένο (κανένα ή περισσότερα) πλήθος μηδενικών. Η περίπτωση 1y είναι επίσης φανερή γιατί αντιστοιχεί σε συμβολοσειρές της L_1 με κανένα μηδενικό μπροστά.

(2) Η γλώσσα L αντιστοιχεί στην κανονική έκφραση (0*10*10*10*)* και επομένως είναι κανονική. Αν θέσουμε ω = 0*1, τότε η παραπάνω κανονική έκφραση αναδιατυπώνεται ως $(\omega^3 0^*)^*$. Για κάθε φυσικό αριθμό k, η κανονική έκφραση $(\omega^k 0^*)^*$ αντιστοιχεί ακριβώς στη γλώσσα με αριθμό μονάδων πολλαπλάσιο του k, και επομένως είναι κανονική.

(3)

<u>(3-A)</u> Η κανονική έκφραση 00+01+10+11 αντιστοιχεί προφανώς στη γλώσσα K = {00,01,10,11}. Η γλώσσα L που αντιστοιχεί στη δοθείσα κανονική έκφραση είναι η γλώσσα $L=K^*$. Επειδή οι συμβολοσειρές της K έχουν μήκος 2, οι συνενώσεις που υπεισέρχονται στο αστέρι Kleene παράγουν πάντοτε συμβολοσειρές με άρτιο μήκος. Η L, δηλαδή, περιέχει μόνο συμβολοσειρές με άρτιο μήκος. Διατυπώνουμε επομένως την υπόθεση ότι η γλώσσα L απαρτίζεται ακριβώς από τις συμβολοσειρές x στο αλφάβητο $\Sigma=\{0,1\}$ που έχουν άρτιο μήκος, $|x| = 2n$. Για να το αποδείξουμε αυστηρά, προχωρούμε με επαγωγή. Για n = 0, $|x| = 0$, ώστε $x = \in$ είναι η κενή συμβολοσειρά, η οποία προφανώς ανήκει στην L. Εστω ότι ο ισχυρισμός αληθεύει για n. Αν τώρα $|x| = 2(n+1)=2n+2$, δηλαδή $x = aby$, με $|y| = 2n$, τότε λόγω της επαγωγικής υπόθεσης $y \in L$. Επειδή τα a,b είναι 0 ή 1, το ab είναι ένα από τα 00,01,10,11. Ωστε προκύπτει ότι και $x = aby \in L$. Συνεπώς η $L \subseteq \Sigma^*$ είναι ακριβώς η γλώσσα των συμβολοσειρών με άρτιο μήκος.

(3-B) Η έκφραση που δίνεται είναι η (0+1)*100(0+1)* και αντιστοιχεί προφανώς στη γλώσσα που περιέχει τις συμβολοσειρές στη Σ^* οι οποίες περιέχουν τη συμβολοσειρά 100 τουλάχιστον μια φορά.

(3-Γ) Εστω L η κανονική γλώσσα που αντιστοιχεί στην κανονική έκφραση 1+0*. Τότε, από τον ορισμό, $L\varnothing = \{xy \mid x\in L,\ y\in\varnothing\}=\varnothing$.

(3-Δ) Είναι $\varnothing+1^*=1^*$ και επειδή γενικά $L^{**}=L^*$, η συγκεκριμένη γλώσσα απαρτίζεται από συμβολοσειρές με μοναδικό στοιχείο το 1, συν την κενή συμβολοσειρά.

(3-Ε) Είναι $\varnothing^* = \bigcup_{n=0}^{\infty}\varnothing^n = \varnothing^0 \cup \left(\bigcup_{n=1}^{\infty}\varnothing^n\right) = \varnothing^0 \cup \varnothing = \varnothing^0 = \{\in\}$, διότι εξ ορισμού $L^0=\{\in\}$, για οποιαδήποτε γλώσσα L.

(4)

(4-Α) Αν θέσουμε (0+1)=ω, τότε η ζητούμενη κανονική έκφραση είναι η $(\omega^5)^*$.

(4-Β) Η γλώσσα αυτή είναι ένωση δύο γλωσσών, με αντίστοιχες κανονικές εκφράσεις (0+1)*01 και (0+1)*11. Συνεπώς η ζητούμενη κανονική έκφραση είναι η (0+1)*01+(0+1)*11.

(4-Γ) Η ζητούμενη κανονική έκφραση είναι η (0+1)*1(0+1)*.

(4-Δ) Η ζητούμενη κανονική έκφραση είναι η

$$\omega = (\in+0+00)(10+110+100)^*(\in+1+11)$$

Έστω $\Sigma = \{0,1\}$.

(α) Κατασκευάστε πεπερασμένα αυτόματα που δέχονται συμβολοσειρές

(1) είτε με άρτιο αριθμό μονάδων, είτε με άρτιο αριθμό μηδενικών
(2) που περιέχουν το 01110 τουλάχιστον μια φορά
(3) με πλήθος του 1 που είναι πολλαπλάσιο του 5

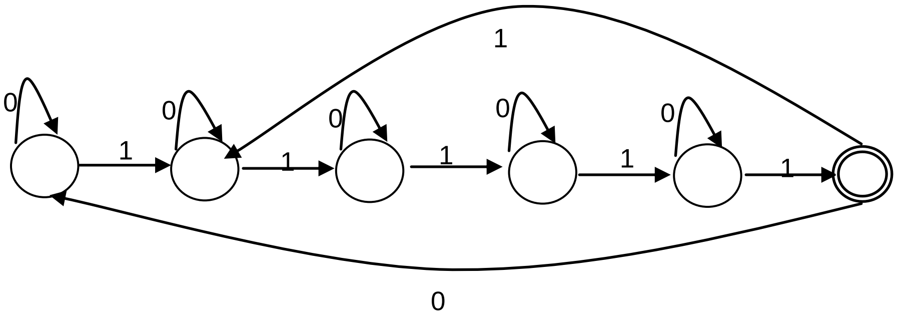
0

(β) Κατασκευάστε πεπερασμένα αυτόματα που αναγνωρίζουν τις γλώσσες

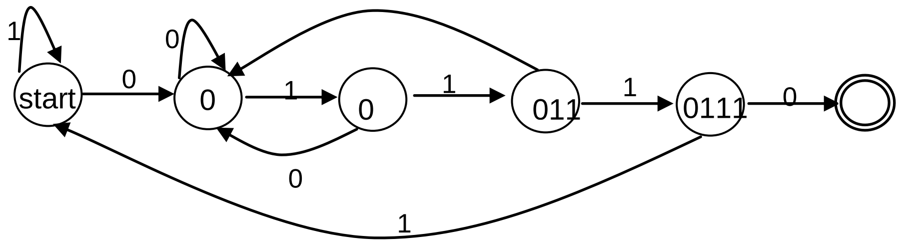
(1) $L_1 = (11+110)^*0$
(2) $L_2 = (01)^*10$
(3) $\Sigma^* - L_1$
(4) $L_1 \cap L_2$
(5) $\Sigma^* - (L_1 - L_2)$

Λύση

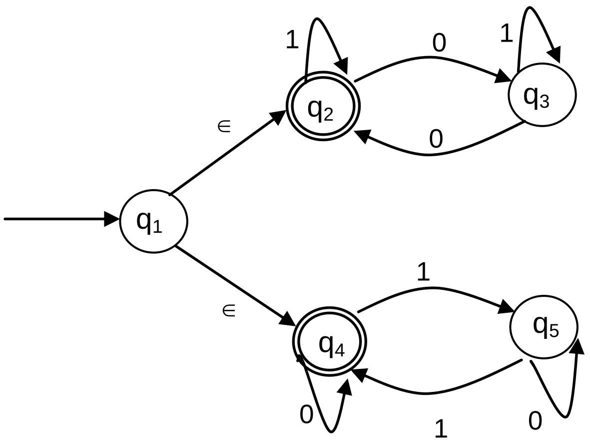
4α1



4α2



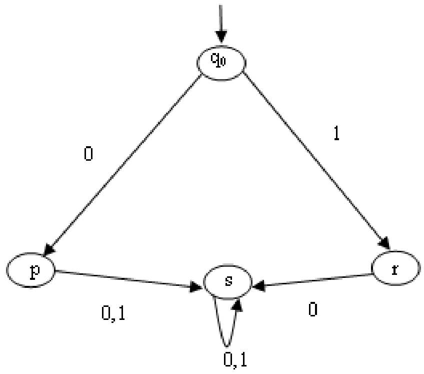
4α3



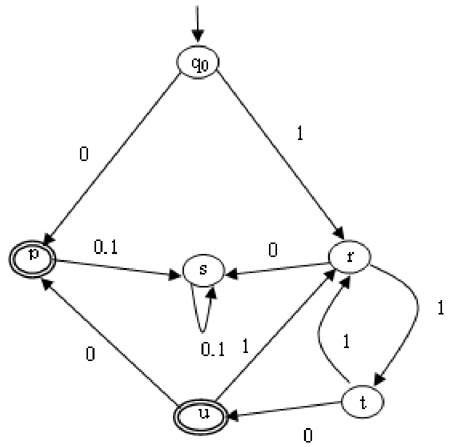
4(β)

1. Η κενή συμβολοσειρά δεν ανήκει στην γλώσσα L, που σημαίνει ότι η αρχική κατάσταση του αυτομάτου q_0 ανήκει σε αποδεκτή κατάσταση. Το 0 ανήκει στην L, άρα από την q_0 μας οδηγεί σε αποδεκτή κατάσταση. Το 1 δεν ανήκει στην L. Εάν έχει διαβαστεί η συμβολοσειρά 110 τελειώνουμε σε αποδεκτή κατάσταση, ενώ στην περίπτωση της συμβολοσειράς 1110 όχι.

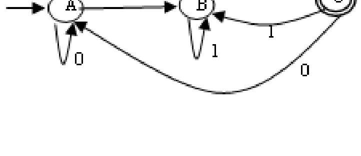
Η γλώσσα L περιέχει το 0 αλλά δεν περιέχει άλλες συμβολοσειρές που να ξεκινούν με 0. Επίσης δεν περιέχει συμβολοσειρές που να ξεκινούν από 10. Πρέπει λοιπόν να εισάγουμε μια κατάσταση s που αναπαριστά όλες τις συμβολοσειρές που αποτυγχάνουν για κάποιο λόγο να είναι πρόθεμα κάποιου στοιχείου της L. (βλέπε σχήμα.)



Στην περίπτωση που το αυτόματο είναι στην κατάσταση r και δέχεται σαν είσοδο 1 δεν πρέπει να παραμείνει στη r γιατί πρέπει να διαχωριστεί το 1 από το 11. Πρέπει λοιπόν να δημιουργηθεί μια νέα κατάσταση t. Από την t με την είσοδο του 0 οδηγούμαστε σε αποδεκτή κατάσταση ($110 \in L$) Αυτή η αποδεκτή κατάσταση δεν είναι η p γιατί το 110 είναι πρόθεμα της L. Η νέα κατάσταση είναι η u . Αν γίνει είσοδος του 0 οδηγούμαστε στην ίδια κατάσταση όπως στην αρχή άρα δ(u,0)=p. Επίσης δ(u,1)= δ(t,1)= r. Το αυτόματο που κατασκευάστηκε είναι:

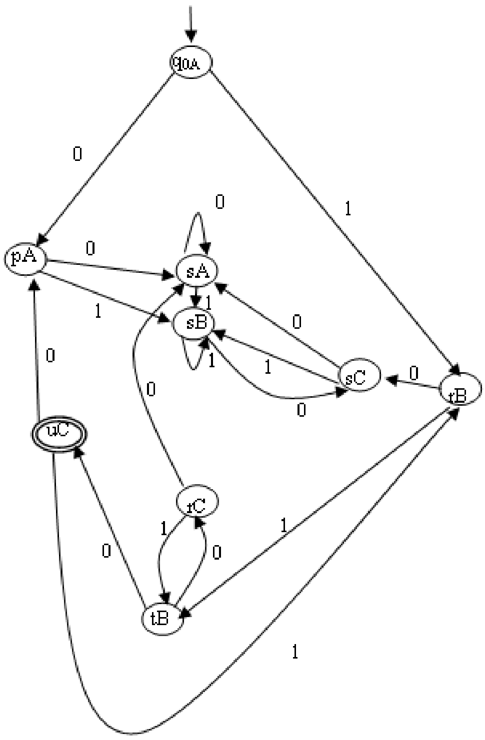


2) Το αυτόματο που κατασκευάστηκε για τη L2=(01)*10 είναι:



3) Το συμπλήρωμα δημιουργείται με την μετατροπή των μη τελικών καταστάσεων σε τελικές.

4) Έστω ότι η L_1 αναγνωρίζεται από το αυτόματο $M1=(P,\Sigma,p_0,\delta_1,F_1)$ και η L_2 από το $M_2=(P,\Sigma,q_0,\delta_2,F_2)$ Το αυτόματο που κατασκευάζεται από την $L_1 \cap L_2$ είναι $M=(P,\Sigma,r_0,\delta,F)$ και έχει καταστάσεις που δημιουργούνται από τα ζεύγη $[p,q] \in P \times Q$. Στο αυτόματο του σχήματος παραλήφθηκαν οι καταστάσεις που δεν είναι προσεγγίσιμες από την αρχική κατάσταση.



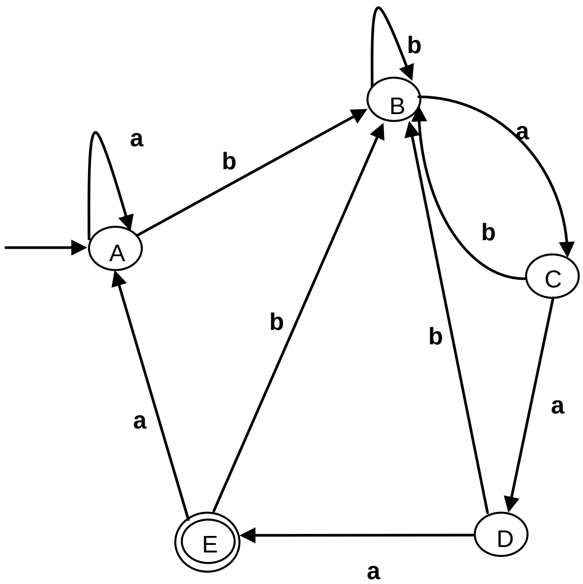
5) Το αυτόματο που κατασκευάζεται από την διαφορά L_1-L_2 είναι το ίδιο με το M, όπως κατασκευάστηκε από την τομή, με τη μόνη διαφορά ότι τελική κατάσταση είναι η [p,A]

Το αυτόματο που κατασκευάζεται από την Σ^* - (L_1-L_2) δημιουργείται με την μετατροπή των μη τελικών καταστάσεων σε τελικές.

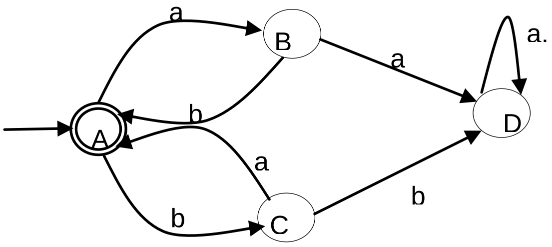
(α) Κατασκευάστε ένα NFA για την κανονική έκφραση (01+11*)*. Ποιο είναι το ισοδύναμο DFA για το NFA που δημιουργήσατε παραπάνω;

(β) Βρείτε κανονικές εκφράσεις που να αντιστοιχούν στα αυτόματα των παρακάτω σχημάτων:

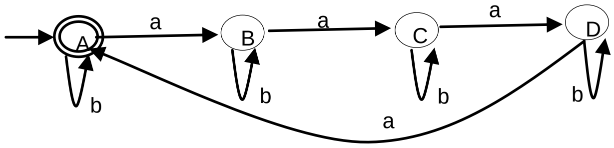
i)



ii)

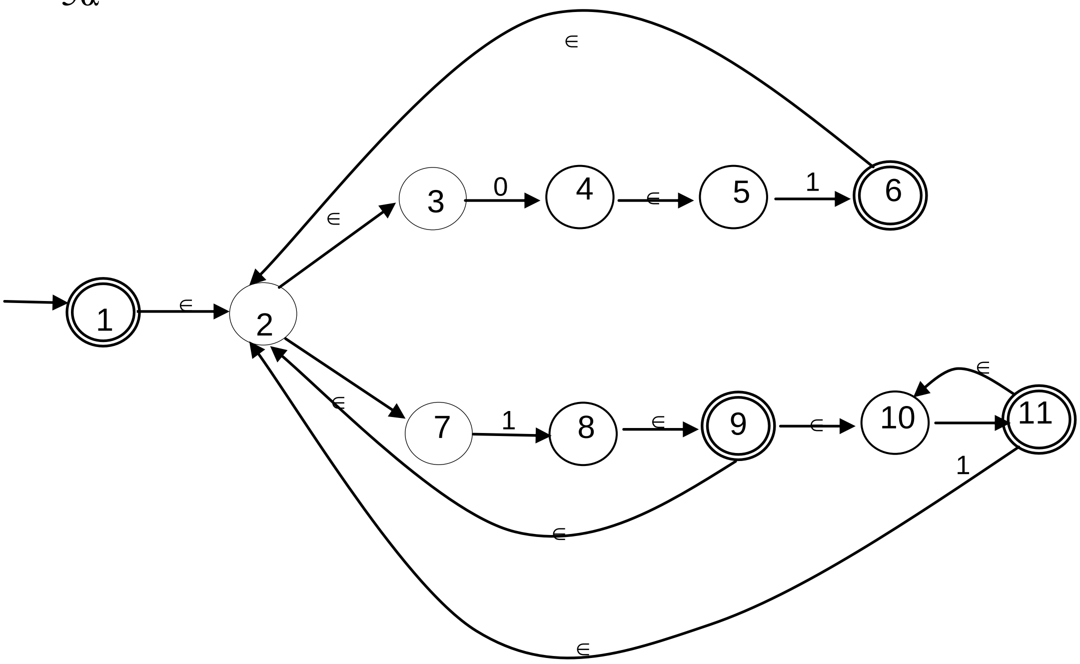


iii)

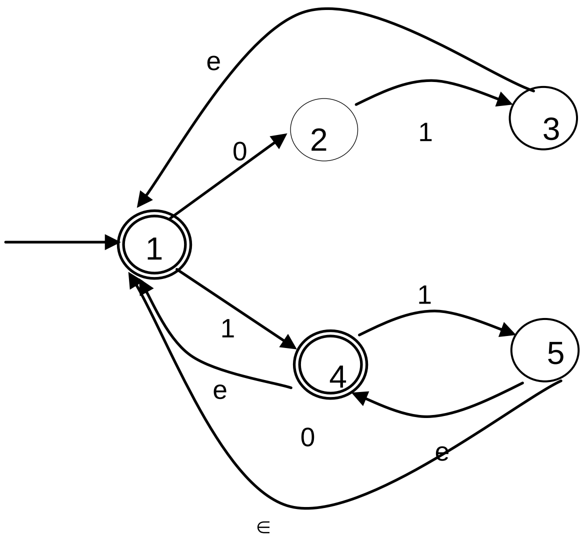


Λύση

5α



Απλοποιημένο ΜΠΑ-ε της 5α



$\delta(5,1)=\varepsilon(\hat{\delta}(\varepsilon(5),1)=\varepsilon(\hat{\delta}(\{5,4\},1))=\varepsilon(\hat{\delta}(5,1)\cup\hat{\delta}(4,1)=\varepsilon(\varnothing\cup\{5\})=\{5,4\}$

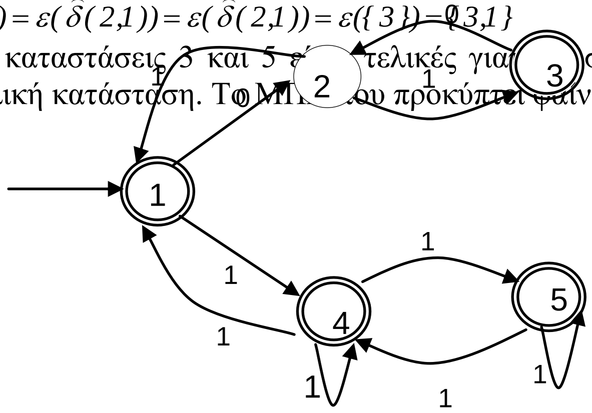
$\delta(3,0)=\varepsilon(\hat{\delta}(\varepsilon(3),0)=\varepsilon(\hat{\delta}(\{3,1\},0))=\varepsilon(\hat{\delta}(3,0)\cup\hat{\delta}(1,0))=\varepsilon(\varnothing\cup\{2\})=\{2\}$

$\delta(4,1)=\varepsilon(\hat{\delta}(\varepsilon(4),1)=\varepsilon(\hat{\delta}(\{4,1\},1))=\varepsilon(\hat{\delta}(4,1)\cup\hat{\delta}(1,1))=\varepsilon(\varnothing\cup\{4\})=\{4\}$

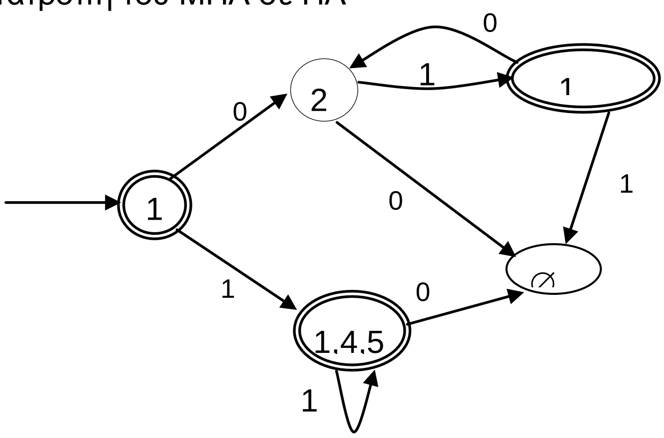
$\delta(5,1)=\varepsilon(\hat{\delta}(\varepsilon(5),1)=\varepsilon(\hat{\delta}(\{5,1\},1))=\varepsilon(\hat{\delta}(5,1)\cup\hat{\delta}(1,1))=\varepsilon(\varnothing\cup\{4\})=\{4\}$

$\delta(2,1)=\varepsilon(\hat{\delta}(2,1))=\varepsilon(\hat{\delta}(2,1))=\varepsilon(\{3\})=\{3,1\}$

Οι καταστάσεις 3 και 5 είναι τελικές γιατί τα σύνολα ε(3) και ε(5) περιέχουν μια τελική κατάσταση. Το ΜΠΑ που προκύπτει φαίνεται στο παρακάτω σχήμα:



Μετατροπή του ΜΠΑ σε ΠΑ



5β)

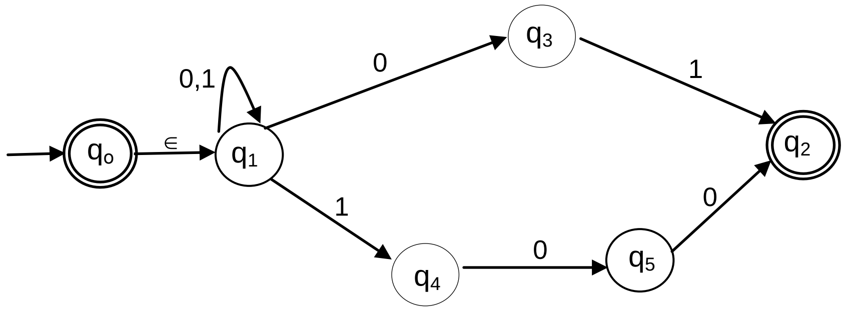
i) Η μικρότερη συμβολοσειρά που γίνεται δεκτή από το αυτόματο είναι η baaa. Ξεκινώντας από το A με την είσοδο του b το αυτόματο μεταβαίνει στην κατάσταση B, με άλλα λόγια οποιοδήποτε συμβολοσειρά έχει εισαχθεί προηγουμένως από τη στιγμή που εισάγετε το σύμβολο b, δεν είναι δυνατόν να γίνει διαχωρισμός της συμβολοσειράς από το σύμβολο b. Κάθε συμβολοσειρά που τελειώνει σε baaa είναι αποδεκτή από το

αυτόματο. Η γλώσσα που δέχεται το αυτόματο είναι το σύνολο όλων των συμβολοσειρών που τελειώνουν σε baaa και μια κανονική έκφραση που αντιστοιχεί στο αυτόματο είναι:
(a +b)*baaa

ii) Σύμφωνα με την άσκησης 4.6 του βιβλίου σελίδα 84 μια κανονική έκφραση είναι: (ab+ba)*

iii) Ξεκινάμε από την κατάσταση A. Εφόσον δ(A,b)=A ξεκινάμε την κανονική έκφραση με b*. δ(A,a)=B και δ(B,b)=B έχουμε b*ab*στη συνέχεια. Το ίδιο ισχύει και στην επόμενη κατάσταση κατά συνέπεια η κανονική έκφραση γίνεται b*ab*ab*. Από την κατάσταση C,δ(C,a)=D και δ(D,b)=D, οπότε η κανονική έκφραση γίνεται b*ab*ab*ab*. δ(D,a)=A που αποτελεί την τελική κατάσταση οπότε μια κανονική έκφραση είναι η :(b*(ab*ab*ab*a)*)* ή (b+(ab*ab*ab*a))*. Η γλώσσα που δέχεται το αυτόματο είναι L={x∈{a,b}*| x αναπαριστά μια συμβολοσειρά όπου το πλήθος των a είναι πολλαπλάσιο του 4}

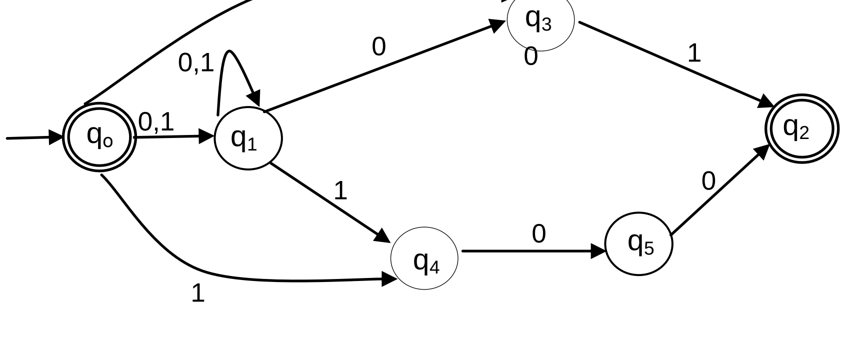
α) Βρείτε τη γλώσσα που αναγνωρίζει το αυτόματο.



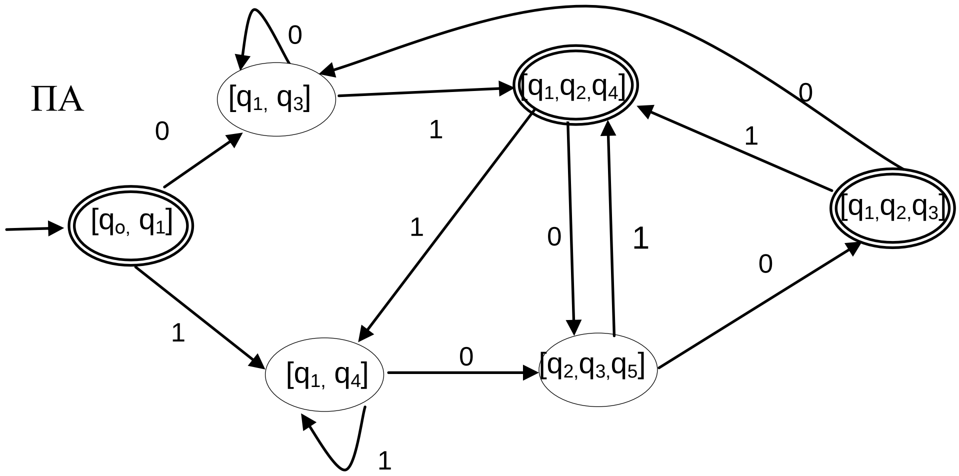
Ποιο είναι το ισοδύναμο ΜΠΑ και ΠΑ για το ΜΠΑ-ε;
Λύση:

α) L={w∈{0,1}*|w είναι κενό ή τελειώνει σε 01 ή 100}

ΜΠΑ



ΠΑ



(α) Ποια από τις παρακάτω γλώσσες στο αλφάβητο {0,1} είναι κανονική και ποια όχι; Δώστε πλήρεις αποδείξεις.

(1) $L_k=\{(01)^{kn} \mid n \geq 0\}$, για κάθε $k \geq 0$.

(2) $L=\{0^{2n^2} \mid n \geq 0\}$

(β) Δείξτε ότι η γλώσσα $\{0^m1010^n \mid m,n \geq 0\}$ είναι ανεξάρτητη συμφραζομένων. Είναι η γλώσσα αυτή κανονική; Εξετάστε τώρα την υπογλώσσα $\{0^n1010^n \mid n \geq 0\}$. Είναι η γλώσσα αυτή ανεξάρτητη συμφραζομένων; Είναι κανονική; (Δώστε πλήρεις αποδείξεις). Είναι δυνατό να κατασκευαστεί ένα ντετερμινιστικό αυτόματο στοίβας που να αναγνωρίζει την υπο-γλώσσα $\{0^n1010^n \mid n \geq 0\}$; Αν ναι, να περιγράψετε ένα τέτοιο αυτόματο.

Απάντηση

(α-1) Ας εξετάσουμε κατ' αρχήν ένα παράδειγμα, για k έστω 3. Η συμβολοσειρά $(01)^{3n}$ είναι η ίδια με την $(010101)^n$, για οποιοδήποτε $n \geq 0$. Συνεπώς η γλώσσα L_3 περιγράφεται από την κανονική έκφραση $(010101)^*$ και άρα είναι κανονική. Διαισθητικά βέβαια παρόμοιο είναι το επιχείρημα για οποιοδήποτε k. Αυστηρά, η απόδειξη είναι με επαγωγή στο k.

Για k=0, $L_0 = \{(01)^0 \mid n \geq 0\}=\{\in\}$ και είναι κανονική.

Εστω ότι η L_k είναι κανονική. Τότε

$$
\begin{aligned}
&L_{k+1} = \{(01)^{(k+1)n} \mid n \geq 0\} = \\
&= \{(01)^{kn+n} \mid n \geq 0\} = \\
&= \{(01)^{kn}(01)^n \mid n \geq 0\} = \\
&= \{(01)^{kn} \mid n \geq 0\}^\frown\{(01)^n \mid n \geq 0\} = \\
&= L_k L_1
\end{aligned}
$$

Παρατηρούμε ότι L_1 είναι η γλώσσα που αντιστοιχεί στην κανονική έκφραση $(01)^*$. Επίσης, λόγω της επαγωγικής υπόθεσης η γλώσσα L_k είναι κανονική και συνεπώς και η L_{k+1} είναι κανονική, ως συνένωση κανονικών γλωσσών.

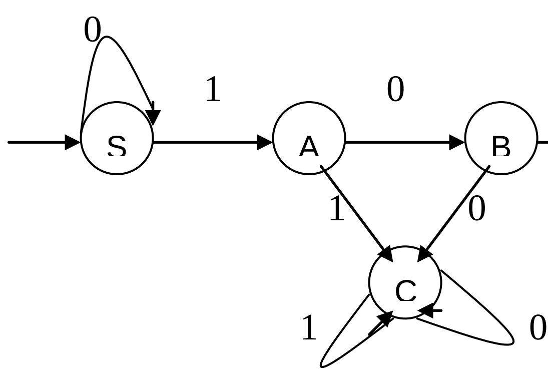
Συμπεραίνουμε από την αρχή της επαγωγής ότι κάθε γλώσσα L_k είναι κανονική.

(α-2) Έστω ότι η L είναι κανονική και έστω p το μήκος άντλησης που προβλέπεται από το Λήμμα Άντλησης. Θεωρούμε τη συμβολοσειρά $s = 0^{2p^2}$. Τότε $|s| = 2p^2 \geq p$ και η s μπορεί να γραφεί ως συνένωση $s = xyz$, όπου $|y| \neq 0$, $|xy| \leq p$ και για κάθε i, $xy^iz \in L$. Παρατηρούμε ότι $s = 0^{2p^2} = 0^{p^2}0^{p^2} = xyz$. Επειδή $|xy| \leq p$, η συμβολοσειρά xy είναι αρχικό τμήμα της συμβολοσειράς 0^{p^2}, δηλαδή $0^{p^2} = xyu$, για κάποιο u με $|u| > 0$ και τέτοιο ώστε $z = u0^{p^2}$.

Εξετάζουμε τώρα τη συμβολοσειρά $xy^2z = xyyz = xyyu0^{p^2}$. Εφόσον πρέπει να ανήκει στην L, το μήκος της είναι της μορφής $2(p+k)^2$, για κάποιο $k \geq 1$. Συνεπώς, $|xyyu| = 2(p+k)^2 - p^2 = |xyu| + |y| = p^2 + |y|$, από το οποίο προκύπτει ότι $|y| = 4pk + 2k^2 > p$, για οποιοδήποτε $k \geq 1$. Αυτό όμως αντίκειται στην υπόθεση ότι $|xy| \leq p$.

Από τα παραπάνω συμπεραίνουμε ότι η L δεν είναι κανονική γλώσσα.

(β) Για τη γλώσσα $\{0^m1010^n \mid m,n \geq 0\}$, είναι φανερό ότι είναι κανονική, με αντίστοιχη κανονική έκφραση την 0^*1010^*. Ενας τρόπος για να σχεδιάσουμε μια γραμματική ανεξάρτηση συμφραζομένων για τη γλώσσα αυτή, είναι να σχεδιάσουμε πρώτα ένα αυτόματο για την κανονική έκφραση 0^*1010^* .



Από την απόδειξη του Θεωρήματος 7.1 (τόμος Γ, σελ 154) προκύπτει ότι η αντίστοιχη Ανεξάρτητη Συμφραζομένων Γραμματική είναι η

$S \rightarrow 0S \mid 1A$

$A \rightarrow 0B \mid 1C$

$B \rightarrow 1F \mid 0C$

$C \rightarrow 0C \mid 1C$

$F \rightarrow \varepsilon \mid 0F$

Για την υπογλώσσα $\{0^n1010^n \mid n \geq 0\}$, ας δεχτούμε ότι είναι κανονική, έστω p το μήκος άντλησης και έστω $s = 0^p1010^p$. Επειδή $|s| \geq p$, η s μπορεί να γραφεί ως η συνένωση $s = xyz$, όπου $|y| > 0$, $|xy| \leq p$ και για κάθε ι, xy^iz ανήκει στη γλώσσα. Τότε θα πρέπει η συμβολοσειρά $xyyz$ να είναι της μορφής $0^{p+k}1010^{p+k}$, για κάποιο $k \geq 1$. Αλλά, επειδή $|xy| \leq p$, είναι φανερό ότι η συμβολοσειρά xy είναι αρχικό τμήμα της 0^p, δηλαδή $s=xyu1010^p$ (όπου $z=u1010^p$), οπότε $xyyz = xyyu1010^p$ και είναι αδύνατο η συμβολοσειρά αυτή να είναι της μορφής $0^{p+k}1010^{p+k}$, για οποιοδήποτε $k \geq 1$. Επομένως δεν ανήκει στη γλώσσα $\{0^n1010^n \mid n \geq 0\}$. Συνεπώς η γλώσσα αυτή δεν είναι κανονική.

Η γλώσσα $\{0^n1010^n \mid n \geq 0\}$ είναι όμως Ανεξάρτητη Συμφραζομένων, με αντίστοιχη γραμματική την παρακάτω:

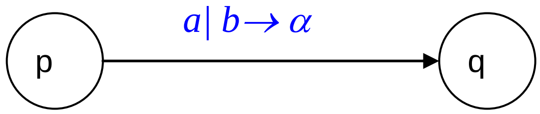
$$A \rightarrow 0A0 \mid 1B1$$

$$B \rightarrow 0$$

Ενα ντετερμινιστικό αυτόματο στοίβας που να αναγνωρίζει την υπο-γλώσσα $\{0^n1010^n \mid n \geq 0\}$ περιγράφεται παρακάτω

Αριθμός κίνησης	Κατάσταση	Σύμβολο Εισόδου	Σύμβολο Σωρού	Κινήσεις
1	q_0	0	Z_0	$(q_0,0Z_0)$
2	q_0	0	0	$(q_0,0Z)$
3	q_0	1	0	$(q_0,1Z)$
4	q_0	0	1	$(q_1,1)$
5	q_1	0	0	(q_1,ε)
6	q_1	1	1	(q_1,ε)
7	q_1	ε	Z_0	(q_2,Z_0)
8	q_0	1	Z_0	$(q_0, 1Z_0)$

Σε ένα αυτόματο στοίβας (push-down automaton) μια *μετάβαση* εξαρτάται από το τρέχον στάδιο, από το υπό ανάγνωση σύμβολο της συμβολοσειράς εισόδου και, τελικά, από το σύμβολο στην κορυφή του σωρού. Η μετάβαση οδηγεί σε κάποιο νέο στάδιο μηχανής και σε έναν τροποποιημένο σωρό. Διαγραμματικά, μια μετάβαση μπορεί να απεικονιστεί ως εξής



όπου (1) p, q είναι στάδια της μηχανής

(2) $\alpha \in \Sigma \cup \{\varepsilon\}$ (Σ είναι το αλφάβητο της συμβολοσειράς εισόδου και η περίπτωση $\alpha = \varepsilon$ αντιστοιχεί σε μετάβαση χωρίς ανάγνωση συμβόλου εισόδου)

(3) $b \in \Gamma \cup \{\varepsilon\}$ και $\alpha \in \Gamma^*$ (Γ είναι το αλφάβητο του σωρού)

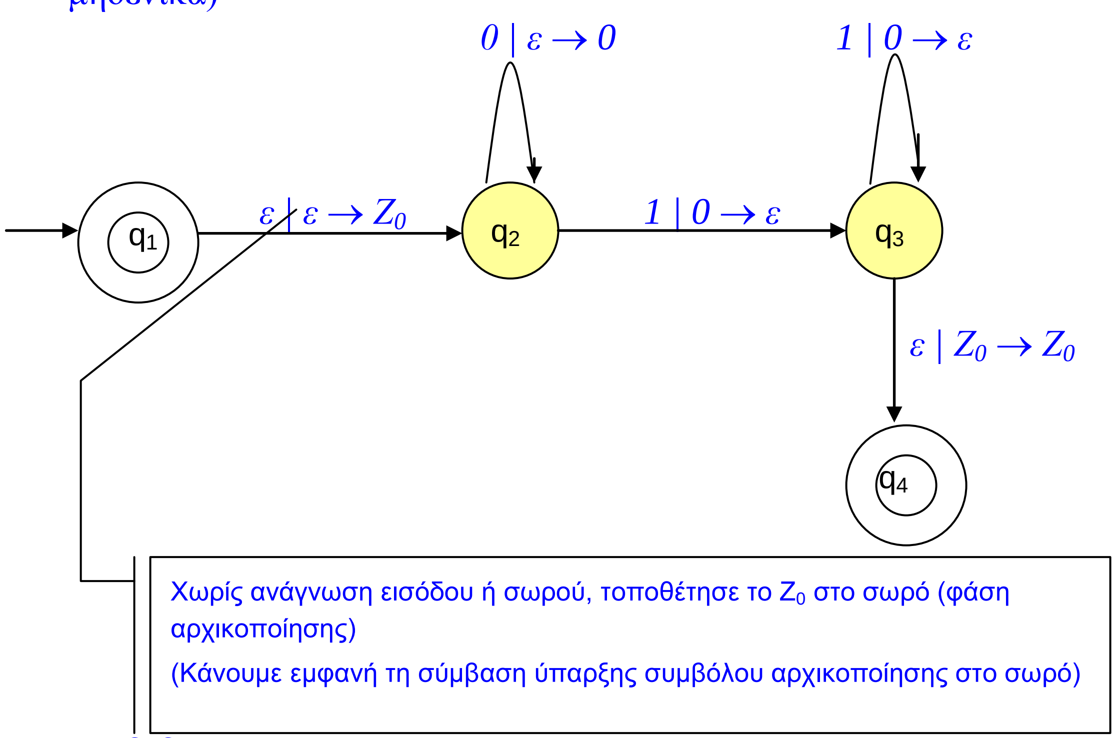
και $b \rightarrow \alpha$ υποδεικνύει την ενέργεια αντικατάστασης του b από τη συμβολοσειρά α στο σωρό, αν το b βρίσκεται στην κορυφή του σωρού. Ιδιαίτερα, οι παρακάτω συμβολισμοί έχουν το υποδεικνυόμενο νόημα:

(i) $c \rightarrow c$: καμμία μεταβολή (c στην κορυφή του σωρού)

(ii) $\varepsilon \rightarrow c$: σπρώξιμο του c στο σωρό (**push** c in the stack)

(iii) $c \rightarrow \varepsilon$: εξαγωγή του c από το σωρό (**pop** c from the stack)

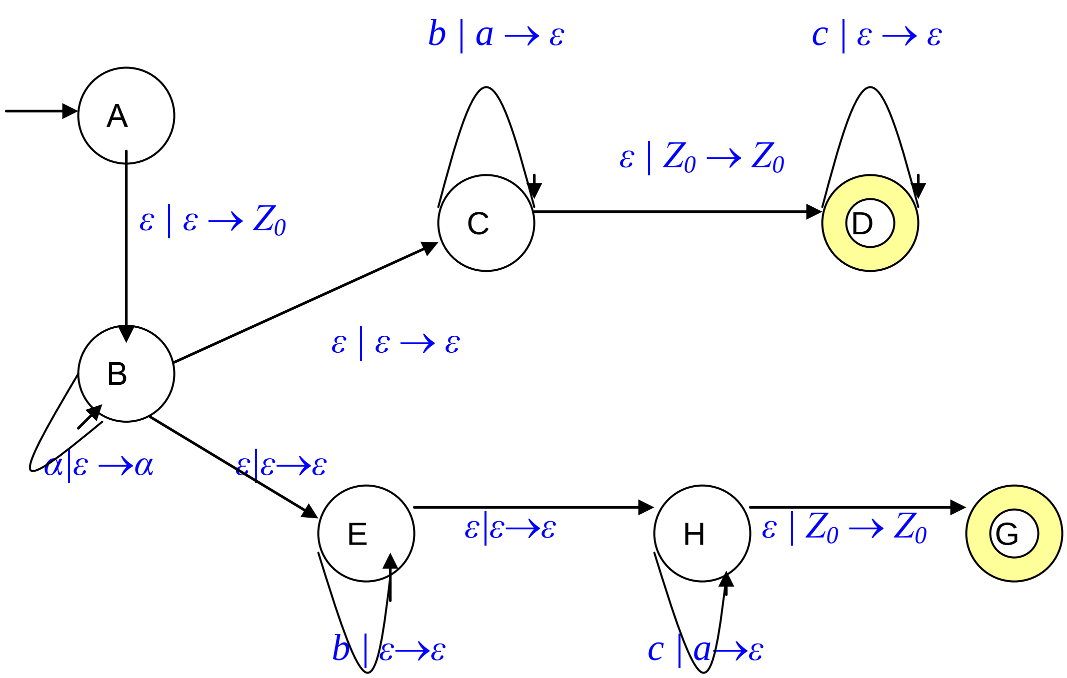
Με τις συμβάσεις αυτές, για παράδειγμα, ένα αυτόματο στοίβας που αναγνωρίζει τη γλώσσα $\{0^n 1^n \mid n \geq 0\}$ δίνεται από το παρακάτω διάγραμμα, όπου το αλφάβητο του σωρού είναι το $\Gamma = \{0, Z_0\}$ (το Z_0 χρησιμοποιείται απλά για να μπορεί να ανιχνευτεί πότε ο σωρρός έχει αδειάσει από μηδενικά)



Με δεδομένα τα παραπάνω

(4) Αναγνωρίστε την Ανεξάρτητη Συμφραζομένων Γραμματική που αντιστοιχεί στο παρακάτω αυτόματο στοίβας.

	Κατάσταση	Σύμβ. εισόδου	Σωρός	Μετάβαση
1	B	a	Z	(B, aZ)
2	B	ε	Z	(C, Z)
3	B	ε	Z	(E, Z)
4	C	b	a	(C, ε)
5	C	ε	Z_0	(D, Z_0)
6	D	c	Z_0	(D, Z_0)
7	E	b	Z	(E, Z)
8	E	ε	Z	(H, Z)
9	H	c	a	(H, ε)
10	H	ε	Z_0	(G, Z_0)



Ποια είναι η γλώσσα αυτής της γραμματικής;

(2) Περιγράψτε (διαγραμματικά ή με τον πίνακα μεταβάσεων) ένα αυτόματο στοίβας που αναγνωρίζει τη γλώσσα που παράγεται από τη γραμματική με κανόνες

$$S \rightarrow aaS \mid aB$$

$$B \rightarrow bB \mid \varepsilon$$

Είναι το αυτόματο που κατασκευάσατε ντετερμινιστικό; Αν όχι, είναι δυνατό να μετατραπεί σε ντετερμινιστικό; Αιτιολογήστε την απάντησή σας.

Απάντηση

(1) Στην πρώτη του ε-μετάβαση το αυτόματο απλά θέτει το σύμβολο Z_0 στο σωρό. Στο στάδιο B όπου μεταβαίνει με την πρώτη ε-μετάβαση, τοποθετεί το σύμβολο α στο σωρό, για κάθε σύμβολο α που διαβάζει στη συμβολοσειρά. Κατόπιν η συμπεριφορά του αυτόματου διαχωρίζεται μη-ντετερμινιστικά. Στη μία περίπτωση αφαιρείται από το σωρό ένα σύμβολο α για κάθε σύμβολο b της συμβολοσειράς που διαβάζεται. Οταν τελειώσουν τα b, αφαιρείται από το σωρό και το σύμβολο Z_0 αρχικοποίησης του σωρού, αν το ζ ήταν στην κορυφή του σωρού (αλλιώς το αυτόματο παγώνει). Στη συνέχεια, ανεξάρτητα από το πόσα σύμβολα c διαβάζονται, το αυτόματο βρίσκεται σε κατάσταση τερματισμού.

Στον άλλο κλάδο ο οποίος προέκυψε από το μη-ντετερμινιστικό διαχωρισμό, το αυτόματο «προσπερνά» όσα b τυχόν υπάρχουν στη συμβολοσειρά που διαβάζει και, κατόπιν, αρχίζει να αφαιρεί ένα α από το σωρό για κάθε c της συμβολοσειράς που διαβάζει. Οταν τελειώσουν τα c, αφαιρείται από το σωρό και το σύμβολο Z_0 αρχικοποίησης του σωρού, αν το Z_0 ήταν στην κορυφή του σωρού (αλλιώς το αυτόματο παγώνει).

Συνεπώς, το αυτόματο αναγνωρίζει με άδειο σωρό τη συμβολοσειρά ακριβώς και μόνον όταν το πλήθος των α είναι το ίδιο είτε με το πλήθος των b, είτε με το πλήθος των c.

Επομένως, η γλώσσα που αναγνωρίζεται από το αυτόματο είναι η

$$\{a^m b^n c^k \mid m,n,k \geq 0 \text{ και } m = n, \text{ ή } m = k\}$$

δηλαδή η ένωση

$$\{a^m b^m c^k \mid m,k \geq 0\} \cup \{a^m b^k c^m \mid m,k \geq 0\}$$

Η πρώτη γλώσσα είναι η συνένωση των γλωσσών

$$L_1=\{a^m b^m \mid m \geq 0\},\ L_2=\{c^k \mid k \geq 0\}$$

Για τη συνένωση έχουμε τη γραμματική

$S_1 \rightarrow TP$

$T \rightarrow \varepsilon \mid aTb$ (γραμματική της γλώσσας L_1)

$P \rightarrow \varepsilon \mid cP$ (γραμματική της γλώσσας L_2)

Η γραμματική της γλώσσας $\{a^m b^k c^m \mid m,k \geq 0\}$ είναι η

$S_2 \rightarrow aS_2b \mid B$

$B \rightarrow \varepsilon \mid bB$

Συνεπώς η ΑΣΓ για τη γλώσσα που αναγνωρίζεται από το αυτόματο στοίβας είναι η

$S \rightarrow S_1 \mid S_2$

$S_1 \rightarrow TP$

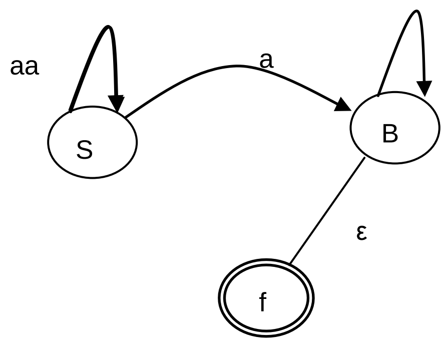
$T \rightarrow \varepsilon \mid aTb$

$P \rightarrow \varepsilon \mid cP$

$S_2 \rightarrow aS_2b \mid B$

$B \rightarrow \varepsilon \mid bB$

(2) Οι κανόνες της Γραμματικής ανήκουν στις μορφές του ορισμού 7.3 και κατά συνέπεια η γραμματική L είναι κανονική. Σύμφωνα με το θεώρημα 7.1 μια γλώσσα είναι κανονική αν και μόνο αν μπορεί να παραχθεί από μια κανονική γραμματική. Μπορούμε λοιπόν να κατασκευάσουμε ένα ΜΠΑ για την κανονική γλώσσα το οποίο παρουσιάζεται στο παρακάτω σχήμα.



Το αυτόματο είναι μη ντετερμινιστικό, μπορεί όμως να μετατραπεί σε ντετερμινιστικό (θεώρημα 5.2).

Σύμφωνα με την άσκηση αυτοαξιολόγησης 8.2 για κάθε αυτόματο Μ υπάρχει ένα αυτόματο στοίβας που αναγνωρίζει την ίδια γλώσσα, και εξομοιώνει την λειτουργία του Μ αφήνοντας τον σωρό με το σύμβολο Z_0.

(α) Βρείτε μια Κανονική ΑΣΓ για τη γλώσσα *L* που καθορίζεται με τον εξής αναδρομικό ορισμό:

(1) $0 \in L$

(2) Αν $x,y \in L$, τότε και $x1y \in L$

(3) Μόνον οι συμβολοσειρές που παράγονται από τους παραπάνω κανόνες 1 και 2 ανήκουν στην *L*,

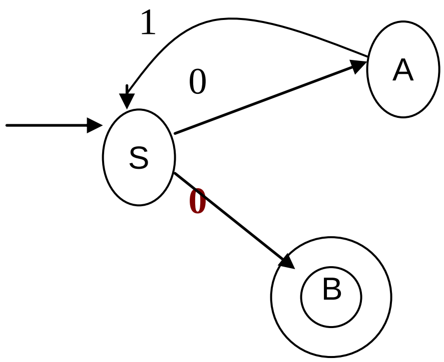
δηλαδή για τη γλώσσα της ΑΣΓ

$$T \rightarrow 0 \mid T1T$$

(β) Δείξτε ότι η γλώσσα $L = \{0^n 1 0^{2n} 1 0^{3n} \mid n \geq 0\}$, δεν είναι Ανεξάρτητη Συμφραζομένων.

Απάντηση

(α) Η γλώσσα *L* είναι κανονκή με αντίστοιχη κανονική έκφραση $(01)^*0$. Ενα αυτόματο που αναγνωρίζει αυτή τη γλώσσα είναι το εξής:



Συνεπώς η <u>Κανονική</u> ΑΣΓ που είναι ισοδύναμη της $T \to 0 \mid T1T$ είναι η

$S \to 0A \mid 0B$

$A \to 1S$

$B \to \varepsilon$

(β) Ας υποθέσουμε ότι η L είναι ανεξάρτητη συμφραζομένων και έστω p το μήκος άντλησης. Θεωρούμε τη συμβολοσειρά $s=0^p10^{2p}10^{3p}$. Επειδή $|s| \ge p$, έστω $s = uxyzv$, με $|xz|>0$, $|xyz| \le p$. Τότε θα πρέπει, π.χ., η ux^pyz^pv να είναι επίσης συμβολοσειρά της L. Διακρίνουμε δύο περιπτώσεις:

<u>Περίπτωση $xyz = 0^q$, $q \le p$, (η xyz δεν περιέχει το 1):</u>

Υπάρχουν τρεις υπο-περιπτώσεις για την s

- $s = 0^s\, 0^q \underline{0^r10^{2p}10^{3p}}$ όπου το υπογραμμισμένο είναι το v, $u = 0^s$ και $q+r+s = p$
- $s = \underline{0^p10^s}\, 0^q\, \underline{0^r10^{3p}}$ όπου τα υπογραμμισμένα είναι τα u και v, αντίστοιχα, και $s+q+r = 2p$
- $s = \underline{0^p10^{2p}10^s}\, 0^q\, 0^r$ όπου το υπογραμμισμένο είναι το u, $v = 0^r$ και $s+q+r = 3p$

Επειδή $x^pyz^p=0^{2p+k}$, για κάποιο $k \ge 0$, παίρνουμε αντίστοιχα τις συμβολοσειρές

- $0^{s+2p+k+r}10^{2p}10^{3p}$
- $0^p\,10^{s+2p+k+r}10^{3p}$
- $0^p10^{2p}10^{s+2p+k+r}$

καμμιά από τις οποίες δεν είναι της μορφής $0^n 1 0^{2n} 1 0^{3n}$

<u>Περίπτωση $xyz = 0^r 1 0^s$, $r+s+1 \le p$ (η xyz περιέχει το 1)</u>

Επομένως, $s = u\, \underline{0^r 1 0^s}\, v$ και υπάρχουν δύο υπο-περιπτώσεις:

- $s = 0^q\, \underline{0^r 1 0^s}\, 0^{2p-s} 1 0^{3p}$
- $s = 0^p 1 0^{2p-r}\, \underline{0^r 1 0^s}\, 0^{3p-s}$

Και στην περίπτωση αυτή, από την αριθμητική των εκθετών προκύπτει ότι καμμία από τις $ux^p yz^p v$ δεν ανήκει στην L.

Εστω $\Sigma = \{0,1\}$.

(5) Περιγράψτε τις γλώσσες που αντιστοιχούν στις παρακάτω κανονικές εκφράσεις:

(A) 0*1(0*10*1)*0*

(B) (1+01)*(0+01)*

(Γ) (00+11)(0+1)*+(0+1)*(00+11)

(2) Δώστε τις κανονικές εκφράσεις που αντιστοιχούν στις παρακάτω γλώσσες:

(A) $\{x \in \Sigma^* \mid$ η x δεν τελειώνει σε 01$\}$

(B) $\{x \in \Sigma^* \mid$ σε κάθε συμβολοσειρά της x κάθε 0 ακολουθείται από 11$\}$

(Γ) $\{x \in \Sigma^* \mid$ κάθε συμβολοσειρά της x περιέχει μόνο μία εμφάνιση της συμβολοσειράς 00$\}$

(Δ) $\{x \in \Sigma^* \mid$ η x δεν περιέχει τη συμβολοσειρά 110$\}$

Απάντηση

(2) (A) Λ+1+(0+1)*0+(0+1)*11

(B) $1^*(011^+)^*$

(Γ) (1+01)*(Λ+0+00)(1+10)*

(Δ) (0+10)*1*

(1) (A) $\{L=\{x \in \{0,1\}^* \mid$ η x περιέχει περιττό αριθμό από 1 $\}$

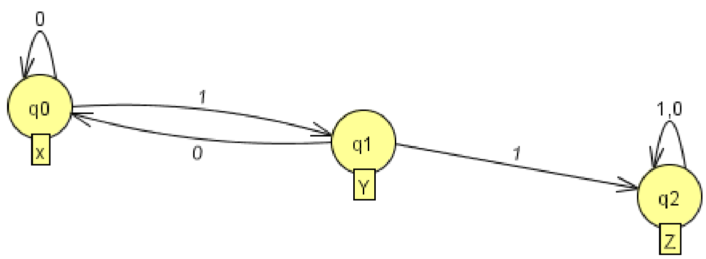
(Β){L={x∈{0,1}*| η x δεν περιέχει την 00x11}

(Γ) {L={x∈{0,1}*| η x αρχίζει ή τελειώνει με 00 ή 11}

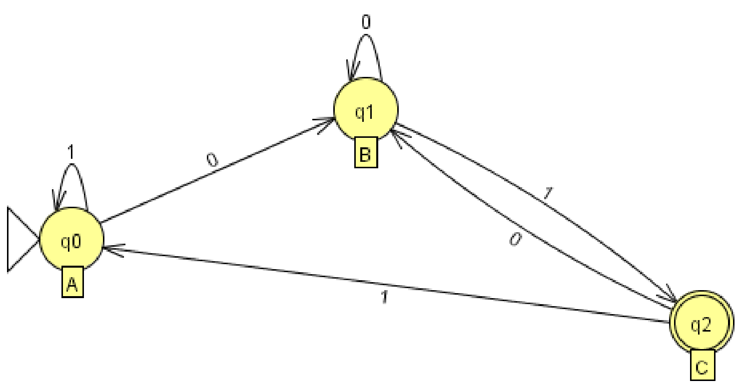
Κατασκευάστε πεπερασμένα αυτόματα που αναγνωρίζουν τις γλώσσες

(6) $L_2 = (0+10)^*11(0+1)^*$
(7) $L_1 = ((1+0)^*01)^*$
(8) $\Sigma^* - L_1$
(9) $L_1 \cup L_2$
(10) $L_1 \cap L_2$
(11) $L_1 - L_2$

Απάντηση

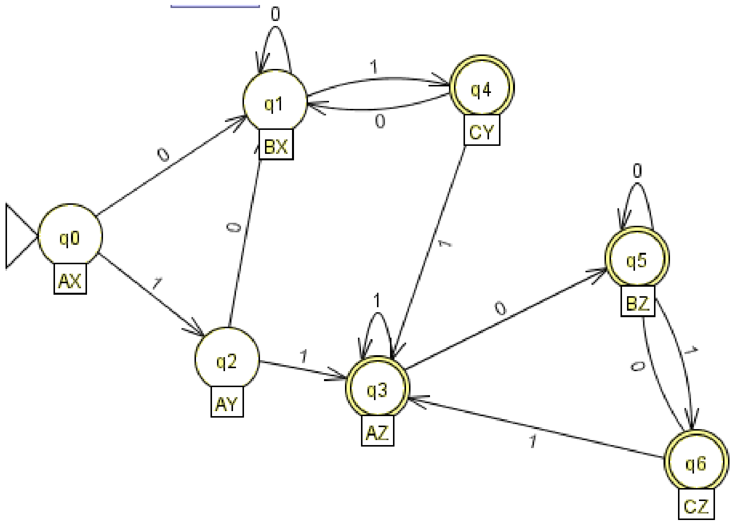


1)



2)

3) Το αυτόματο του συμπληρώματος προκύπτει με την αντιστροφή των καταστάσεων, δηλαδή οι τελικές γίνονται μη τελικές και οι μη τελικές γίνονται τελικές.



4)
5) Το αυτόματο της τομής είναι το ίδιο με της ένωσης, με τη διαφορά ότι η κατάσταση q6 είναι η μόνη τελική κατάσταση της τομής
6) Το αυτόματο της διαφοράς είναι το ίδιο με της ένωσης με τη διαφορά ότι η κατάσταση q4 είναι η μόνη τελική κατάσταση της διαφοράς.

Για κάθε μία από τις ακόλουθες κανονικές εκφράσεις δώστε ένα ντετερμινιστικό πεπερασμένο αυτόματο που αναγνωρίζει τις αντίστοιχες γλώσσες.

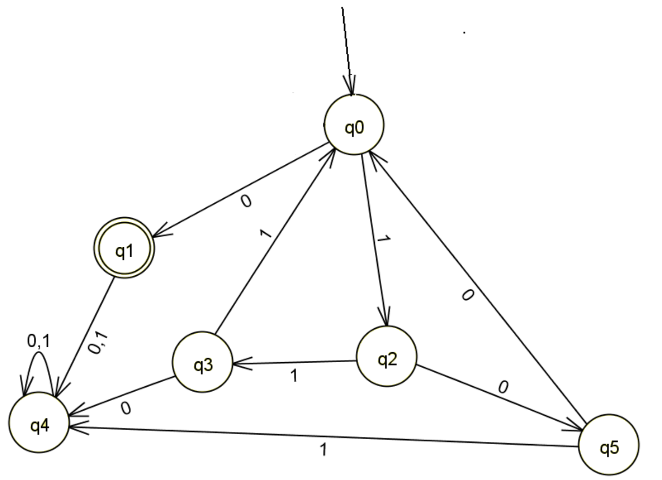
(A) (111+100)*0

(B) (11+10)*

(Γ) (0+01)*

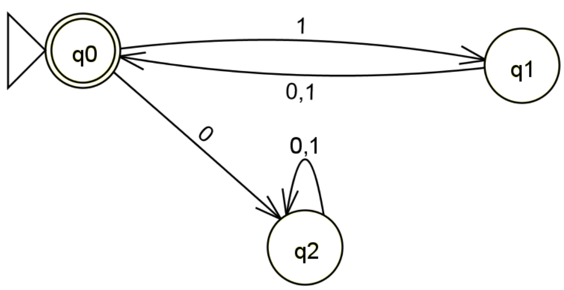
Απάντηση

(Α) Ξεκινώντας από την κατάσταση q0 με την είσοδο του συμβόλου 0 οδηγούμαστε στην τελική κατάσταση q1. Με την είσοδο της συμβολοσειράς 111 οδηγούμαστε στις καταστάσεις q2,q3,q0 και με την είσοδο της συμβολοσειράς 100 στις καταστάσεις

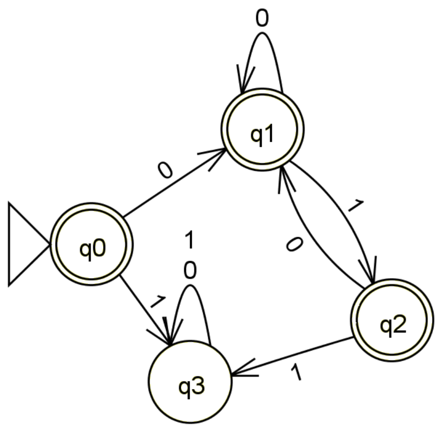


q2,q5,q0. Με την είσοδο άλλης συμβολοσειράς οδηγούμαστε στην κατάσταση q4.

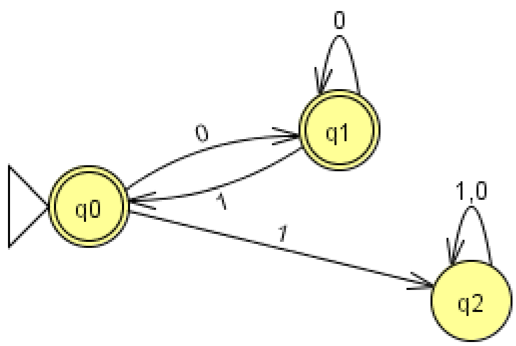
(Β)
Το αυτόματο έχει αρχική και τελική κατάσταση την q0. Με την είσοδο είτε 11 είτε 10 οδηγείτε στην τελική κατάσταση όπως και με την κενή συμβολοσειρά η οποία ανήκει στην γλώσσα.



(Γ) Το αυτόματο έχει τρεις τελικές καταστάσεις την q0 (κενή συμβολοσειρά) την q1 (με την είσοδο του 0) και την q2 (με την είσοδο της συμβολοσειράς 01).



Το αυτόματο μπορεί να απλοποιηθεί και άλλο συγχωνεύοντας τις καταστάσεις q0 και q2 οπότε προκύπτει:



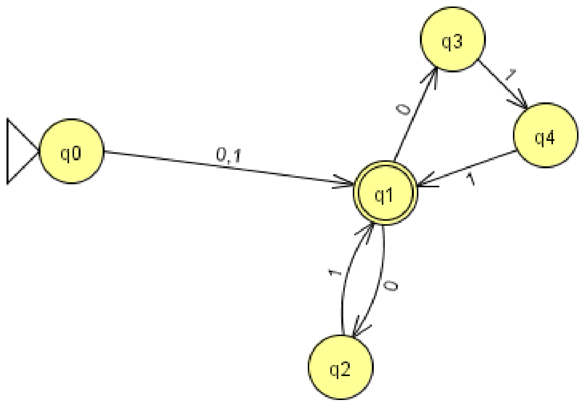
Κατασκευάστε ένα ΜΠΑ για την κανονική έκφραση (0+1)((01)*(011)*)*. Ποιο είναι το ισοδύναμο ΝΠΑ για το ΜΠΑ που δημιουργήσατε παραπάνω;

Απάντηση

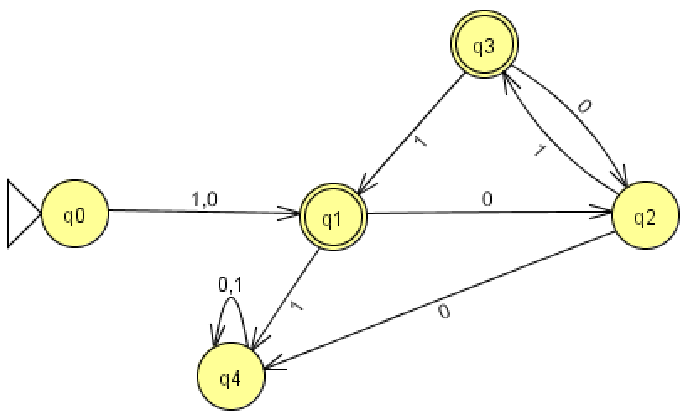
Η κανονική έκφραση μπορεί να γραφεί:

(0+1)((01)*(011)*)*=(0+1)(01+011)*.

Το αντίστοιχο Μη Ντετερμινιστικό Αυτόματο (ΜΠΑ) είναι:



Το αντίστοιχο Ντετερμινιστικό Αυτόματο (ΝΠΑ) είναι:



Ποια από τις παρακάτω γλώσσες στο αλφάβητο {a,b,c} είναι κανονική και ποια όχι? Δώστε πλήρεις αποδείξεις

1. $A = \{a^m b^2 | \ m \geq 0\}$
2. $B = \{a^m b^{m-2} c \mid m \geq 2 \}$
3. $C = \{(ab)^m c^{2m} \mid m \geq 0 \}$
4. $D_k = \{(abcc)^m (ccba)^k \mid m \geq 0 \}, \ k \geq 0$
5. $D = \{(abcc)^m (ccba)^k \mid m \geq 0 , \ k \geq 0 \}$

Απάντηση

1. Η γλώσσα $\{a^m \mid m \geq 0\}$ είναι κανονική, με κανονική έκφραση a^*, όπως κανονική είναι και η γλώσσα $\{b^2\}$, η δε A προκύπτει ως συνένωση αυτών των γλωσσών, άρα είναι κανονική

2. Εστω ότι η B είναι κανονική, p το μήκος άντλησης και $w = a^p b^{p-2} c$. Από το Λήμμα Άντλησης γνωρίζουμε ότι w είναι της μορφής w = xyz με $|xy| = q \le p$, $|y| > 0$ και έτσι ώστε για κάθε n, $xy^n z \in B$. Επομένως, $w = a^q a^{p-q} b^{p-2} c$, όπου $z = a^{p-q} b^{p-2} c$. Επειδή $|xy| = q \le p$, $|y| > 0$, η λέξη $xy^2 z = a^r a^{p-q} b^{p-2} c$, με $r > q$, δεν μπορεί να ανήκει στη γλώσσα B. Επομένως η B δεν είναι κανονική.
3. Εστω ότι C κανονική, p το μήκος άντλησης και $w = (ab)^p c^{2p} = xyz$. Επειδή $|xy| = q \le p$, $|y| > 0$, η λέξη $xy^2 z$ δεν μπορεί να έχει τη μορφή $\{(ab)^m c^{2m}$, δηλαδή δεν ανήκει στη C. Αρα η C δεν είναι κανονική.
4. Για κάθε k η γλώσσα $\{(ccba)^k\}$ είναι κανονική. Επίσης, η γλώσσα $\{(abcc)^m | m \ge 0\}$ είναι κανονική, η δε D_k προκύπτει ως συνένωση των δύο αυτών γλωσσών, άρα είναι κανονική.
5. Οι γλώσσες $\{(abcc)^m | m \ge 0 \}$, $\{(ccba)^k | k \ge 0 \}$ είναι κανονικές και επομένως η συνένωσή τους D είναι κανονική.

Να δείξετε, με χρήση του Λήμματος Αντλησης, ότι η γλώσσα

$$L = \{ a^{3n} b^{2n} c^n | n \ge 0\}$$

δεν είναι ανεξάρτητη συμφραζομένων.

Απάντηση

Εστω ότι είναι και εστω p το μήκος άντλησης και $s = a^{3p} b^{2p} c^p$. Επειδή $|s| \ge p$, η s γράφεται ως s = uxyzv, όπου $|xz| > 0$, $|xyz| \le p$ και, για κάθε k, $ux^k yz^k v \in L$.

Διακρίνουμε περιπτώσεις:

A) $xyz = a^q$, $0 < q \le p$. Τότε $s= ua^q a^r b^{2p} c^p$ και η συμβολοσειρά $ux^p yz^p a^r$ έχει μήκος 3p+2(p-1)=5p-2, οπότε $ux^p yz^p a^r b^{2p} c^p = a^{5p-2} b^{2p} c^p$ δεν μπορεί να ανήκει στην L.

B) $xyz = b^q$, $0 < q \le p$. Παρόμοιο.

Γ) $xyz = c^q$, $0 < q \le p$. Παρόμοιο.

Δ) $xyz = a^q b^t$, $u = a^{3p-q}$, $v = b^{2p-t} c^p$. Τότε στην $s = a^{3p-q} xyzb^{2p-t} c^p$, έχουμε $| a^{3p-q} xyzb^{2p-t} | = 5p$. Σύμφωνα με το Λήμμα Αντλησης, η $ux^{3p} yz^{3p} v$ θα πρέπει να είναι συμβολοσειρά της L. Αλλά

$$ux^{3p} yz^{3p} v = a^{3p-q} x^{3p-1} (xyz) z^{3p-1} b^{2p-t} c^p = a^{3p-q} x^{3p-1} (a^q b^t) z^{3p-1} b^{2p-t} c^p$$

και το μήκος $| a^{3p-q} x^{3p-1} (a^q b^t) z^{3p-1} b^{2p-t} | > 5p$, επειδή $|xz| > 0$

E) $xyz = b^q c^t$. Παρόμοιο.

Εστω η γλώσσα L στο αλφάβητο {a,b} που καθορίζεται με τον αναδρομικό ορισμό

1. $a \in L$
2. Αν $x \in L$, τότε και $xa^2 bx \in L$
3. $x \in L$ ανν αυτό προκύπτει από τα 1, 2 παραπάνω.

Είναι η L κανονική? Αν ναι, βρείτε μια Κανονική ΓΑΣ που την παράγει. Αν όχι, αποδείξτε το.

Απάντηση

Λέξεις της γλώσσας είναι

$a=(a^3b)^0a$, $a(a^2b)a= (a^3b)^1a$, $(a^3b)a(a^2b)(a^3b)a= (a^3b)^3a$, και, με επαγωγή,

$$L=\{(a^3b)^{s_n}a \mid \ s_0=0, \quad s_{n+1}=2s_n+1\}$$

Εστω ότι η L είναι κανονική και p το μήκος άντλησης. Θεωρούμε τη λέξη $w=(a^3b)^{s_p}a$, με μήκος $|w|=4s_p+1 >4p+1>p$. Τότε $w = xyz$, με $0 < |y| \leq |xy| \leq p$. Επομένως,

(1) $\quad xy = (a^3b)^m a^k$

για κάποιο φυσικό m και για $k\in\{0,1,2,3\}$ έτσι ώστε $4m+k \leq p$. Επειδή $p < s_p$, θα είναι επίσης

(2) $\quad z = a^{3-k}b(a^3b)^r a$

όπου $r = s_p-m-1$ και

(3) $\quad w = xyz = (a^3b)^m a^k a^{3-k}b(a^3b)^r a.$

Από την (1), συμπεραίνουμε επίσης ότι

(4) $\quad x = (a^3b)^t a^j, \quad y = a^{3-j}b(a^3b)^{m-t-1}a^k,$

όπου $0 \leq t \leq m$ και $0 \leq j \leq k$.

Εξετάζουμε τώρα τη λέξη

$w_2 = xy^2z = (a^3b)^t a^j a^{3-j}b(a^3b)^{m-t-1}a^k a^{3-j}b(a^3b)^{m-t-1}a^k a^{3-k}b(a^3b)^r a$

$\quad = (a^3b)^m a^{k+3-j}b(a^3b)^{m-t+r}a$

Για να είναι η w_2 στην L, θα πρέπει να είναι $k = j$, οπότε

$w_2 = xy^2z = (a^3b)^{2m+1-t+r}a$

όπου ο εκθέτης $2m+1-t+r = 2m+1-t+(s_p-m-1) = s_p+(m-t)$ θα πρέπει να είναι της μορφής $s_p+(m-t) = s_n = 2s_{p+i}+1$ για κάποιο $i \geq 0$. Ο ισχυρισμός αυτός αποδεικνύεται από την παρατήρηση ότι δεν μπορεί να είναι $m-t = 0$, διότι τότε θα είχαμε $|xyz| = |xy^2z|$, πράγμα που μπορεί να συμβεί μόνον αν έχουμε $|y| = 0$, άτοπο.

Αλλά όμως, για κάθε i, $2s_{p+i}+1 \geq 2s_p+1 = s_{p+1}$ δηλαδή $s_{p+1} \leq s_p+(m-t)$. Επειδή $4m+k \leq p$ και $k\in\{0,1,2,3\}$, έχουμε $m \leq (p-k)/4 \leq p/4$ και επειδή $0 \leq t \leq m$ και $p < s_p$ έχουμε

$$s_{p+1} \leq s_p+(m-t) \leq s_p+m \leq s_p+ p/4 < 2 s_p < 2s_p+1 = s_{p+1}$$

δηλαδή $s_{p+1} < s_{p+1}$, άτοπο.

Συμπεραίνουμε ότι η γλώσσα *L* δεν είναι κανονική.

Έστω η γλώσσα L που περιέχει όλες τις συμβολοσειρές του $(a,b)^*$ που αρχίζουν με *ab* και συνεχίζουν είτε με ένα είτε με δύο είτε με τρία *α*. Για παράδειγμα οι συμβολοσειρές *ε, a, aa b, abb, ab,* δεν ανήκουν στην L, ενώ οι *aba, abaabaa, abaaaaba, abaaabaa* ανήκουν. Η *ε* είναι η κενή συμβολοσειρά.

i. Να γράψετε κανονική παράσταση που να περιγράφει τη γλώσσα L.

ii. Να κατασκευάσετε ένα απλό μη ντετερμινιστικό πεπερασμένο αυτόματο, το οποίο δέχεται την ίδια γλώσσα με την κανονική παράσταση (δεν χρειάζεται να χρησιμοποιήσετε τον αλγόριθμο).

iii. Να μετατρέψετε το αυτόματο του ερωτήματος *ii.* σε ντετερμινιστικό πεπερασμένο αυτόματο.

Απάντηση

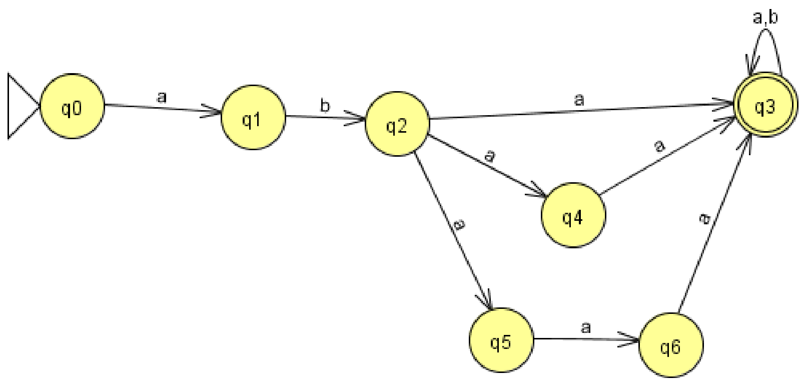
i) ab(a+aa+aaa)(a+b)*

ή απλούστερα

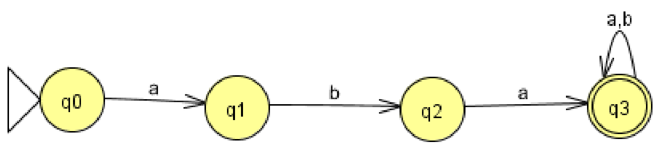
aba(a+b)*

και οι δύο κανονικές εκφράσεις περιγράφουν την ίδια γλώσσα

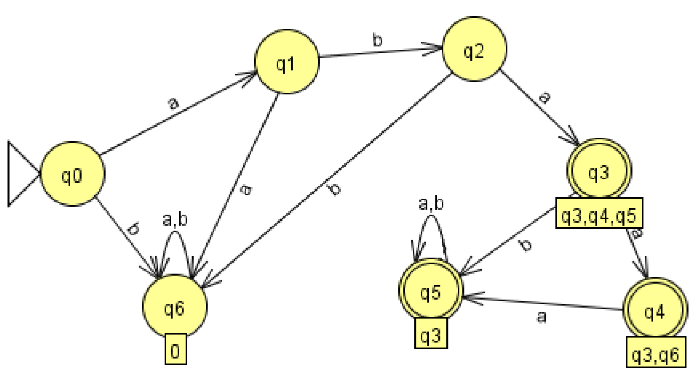
ii) Η πρώτη κανονική έκφραση δίνει το ακόλουθο ΜΠΑ



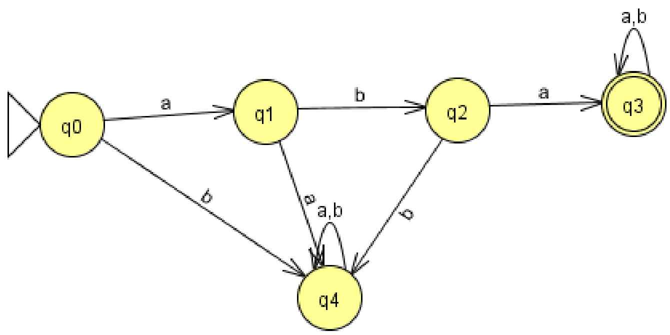
Ενώ η δεύτερη το ΜΠΑ:



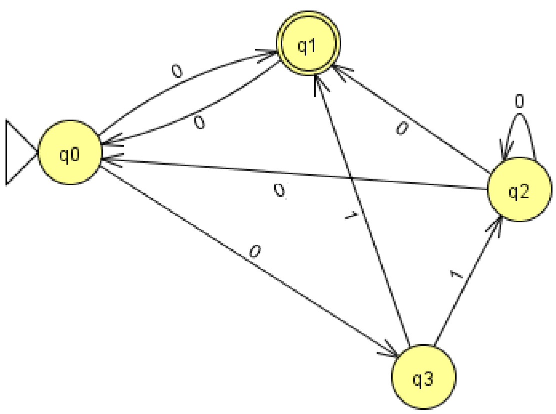
iii) Το ντετερμινιστικό αυτόματο της πρώτης κανονικής έκφρασης είναι:



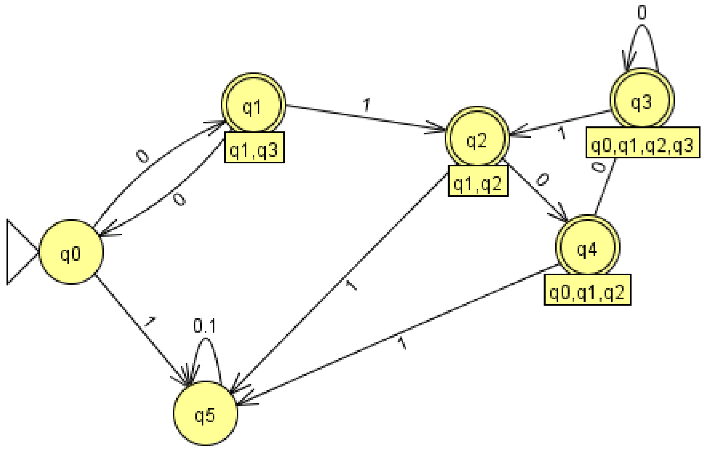
και της δεύτερης:



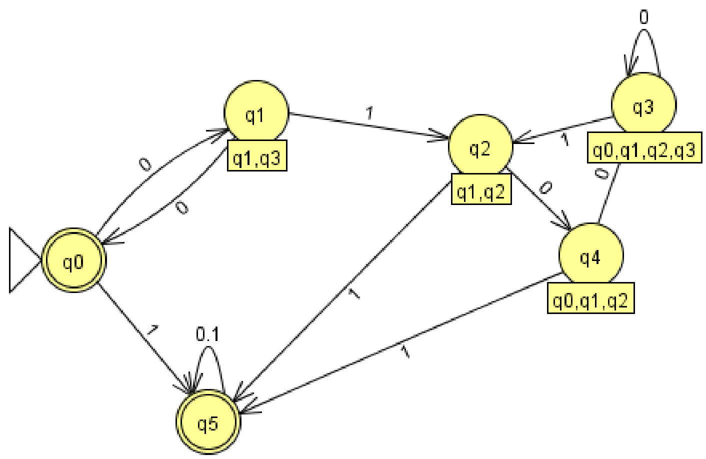
Να κατασκευαστεί το ντετερμινιστικό πεπερασμένο αυτόματο που δέχεται το συμπλήρωμα της γλώσσας του αυτομάτου:



Απάντηση



Το συμπλήρωμα που προκύπτει με εναλλαγή των τελικών και μη τελικών



καταστάσεων είναι:

Εξετάστε αν είναι κανονική η γλώσσα L των συμβολοσειρών στο αλφάβητο $\Sigma = \{0,1\}$ στις οποίες ο αριθμός των μηδενικών είναι μικρότερος από το διπλάσιο του αριθμού των άσσων.

Απάντηση.

Για κάθε θετικό φυσικό n η λέξη 0^n1^n ανήκει στην L (διότι $n < 2n$) και επομένως η L είναι άπειρη.
Υποθέτουμε ότι είναι κανονική και έστω p το μήκος άντλησης για τη γλώσσα.
Η λέξη $u = 0^p1^p$ ανήκει στην L και έχει μήκος $|u| \geq p$.

Από το Λήμμα Αντλησης προκύπτει ότι η u μπορει να γραφεί ως συνέννωση, $u = xyz$, έτσι ώστε

$$0 < |y| \le |xy| \le p$$

Επιπλέον, σύμφωνα με το Λήμμα Αντλησης, θα πρέπει να ισχύει ότι

$$(\forall i)\ (xy^i z \in L)$$

Επειδή $|xy| \le p = |0^p|$, θα είναι $xy = 0^q$, για κάποιο $q \le p$, ώστε $z = 0^{p-q}1^p$
Εειδή $|y|>0$, θα είναι $y = 0^t$, για κάποιο $0<t \le q$. Επομένως η u γράφεται ως

$$\underbrace{0^{q-t}}_{x}\ \underbrace{0^t}_{y}\ \underbrace{0^{p-q}1^p}_{z}$$

Θεωρούμε τη λέξη

$$u_* = xy^{(p+1)^2}z = \underbrace{0^{q-t}}_{x}\ \underbrace{0^{t(p+1)^2}}_{y^{(p+1)^2}}\ \underbrace{0^{p-q}1^p}_{z}$$

Επειδή η λέξη πρέπει να ανήκει στην L, έχουμε

$$(q-t)+t(p+1)^2+(p-q)<2p \Leftrightarrow$$
$$\Leftrightarrow t(p+1)^2 - t < p \Leftrightarrow$$
$$\Leftrightarrow tp(p+2) < p \Leftrightarrow$$
$$\Leftrightarrow t < \frac{1}{p+2} < 1 \Rightarrow$$
$$\Rightarrow t = 0$$

άτοπο, διότι $t = |y| > 0$.

Παρατήρηση: Μια εναλλακτική, ίσως απλούστερη απόδειξη, θα ξεκινούσε από τη λέξη $u = 0^{2p-1}1^p$.

A)(5) Δώστε γραμματική για τη γλώσσα που παράγεται από την κανονική έκφραση (000+11)*(01)*

B)(5) Τροποποιήστε την παρακάτω γραμματική έτσι ώστε να μην περιέχει κανόνες της μορφής $A \to \varepsilon$ ή $A \to B$

$S \rightarrow ABA$

$A \rightarrow \alpha A|\varepsilon$

$B \rightarrow bB|\varepsilon$

Γ)(10) Κατασκευάστε ντετερμινιστικό αυτόματο στοίβας για τη γλώσσα

$\{a^n b^{n+m} a^m | n,m \geq 0\}$

Απάντηση

Α) Η γραμματική που προκύπτει είναι:

$S \rightarrow BC$

$B \rightarrow AB|\varepsilon$

$A \rightarrow 000|11$

$C \rightarrow DC|\varepsilon$

$D \rightarrow 01$

Β) Η γραμματική που προκύπτει μετά την απαλοιφή των ε-κανόνων είναι:

$S \rightarrow ABA|AB|BA|AA|\alpha A|bB|\alpha|b$

$A \rightarrow \alpha A|\alpha$

$B \rightarrow bB|b$

Γ)

Αρ κίνησης	Κατάσταση	Σύμβολο εισόδου	Σύμβολο Σωρού	Κίνηση
1	q_0	a	Z_0	(q_1,aZ_0)
2	q_0	b	Z_0	(q_4, bZ_0)
3	q_1	a	a	(q_1, aa)
4	q_1	b	a	(q_2, ε)
5	q_2	b	a	(q_2, ε)
6	q_2	ε	Z_0	(q_3, Z_0)
7	q_3	b	Z_0	(q_4, bZ_0)

8	q_4	b	b	(q_4, bb)
9	q_4	a	b	(q_5, ε)
10	q_5	a	b	(q_5, ε)
11	q_5	ε	Z_0	(q_6, Z_0)
(Οι υπόλοιποι συνδυασμοί)				τίποτα

Το αυτόματο έχει αποδεκτές καταστάσεις τις q_0,q_3 και q_6. Στην κατάσταση q1 σπρώχνει τα a's στον σωρό, στην q_2 ταιριάζει τα a'ς του σωρού με τα b's της εισόδου, η q_4 σπρώχνει τα b's στο σωρό και η q_5 τα ταιριάζει με τα a's της εισόδου.

Δείξτε ότι η γλώσσα $L=\{a^ib^jc^k| \ i\neq j \text{ ή } i\neq k\}$ είναι Ανεξάρτητη Συμφραζομένων.

Απάντηση

Η L είναι η ένωση των γλωσσών $\{a^ib^jc^k| \ i\neq j\}$ και $L=\{a^ib^jc^k| \ i\neq k\}$ κάθε μία από τις οποίες είναι Ανεξάρτητη Συμφραζομένων, Οι γλώσσες ανεξάρτητες συμφραζομένων είναι κλειστές ως προς την ένωση οπότε και η L είναι ανεξάρτητη συμφραζομένων.

Εξετάστε αν είναι κανονική η γλώσσα L των συμβολοσειρών στο αλφάβητο Σ = {a,b,c} στις οποίες ο αριθμός των a είναι μικρότερος από το άθροισμα του αριθμού των b και c.

Απάντηση

Για κάθε φυσικό n ≥ 1, η λέξη $w = a^n b^n c^n$ ανήκει στην L, επομένως η L είναι άπειρη. Εστω ότι η L είναι κανονική και έστω p το μήκος άντλησης. Θεωρούμε τη λέξη $u = a^p b^p c^p$. Επειδή |u| ≥ p, προκύπτει από το Λήμμα Άντλησης ότι η u μπορει να γραφεί ως συνένωση u = xyz, για x,y,z που ικανοποιούν τους περιορισμούς

$$0 < |y| \leq |xy| \leq p$$

Επιπλέον, σύμφωνα με το Λήμμα Αντλησης, θα πρέπει να ισχύει ότι

$$(\forall i) \ (xy^iz \in L)$$

Επειδή $|xy| \le p = |a^p|$, θα είναι $xy = a^q$, για κάποιο $q \le p$, ώστε $z = a^{p-q}b^pc^p$. Επειδή $|y|>0$, θα είναι $y = a^t$, για κάποιο $0 < t \le q$. Επομένως η u γράφεται ως

$$u = \underbrace{a^{q-t}}_{x}\ \underbrace{a^{t}}_{y}\ \underbrace{a^{p-q}b^{p}c^{p}}_{z}$$

Θεωρούμε τη λέξη

$$u_* = xy^{p(p+1)}\ z = \underbrace{a^{q-t}}_{x}\ \underbrace{a^{tp(p+1)}}_{y^{p(p+1)}}\ \underbrace{a^{p-q}b^{p}c^{p}}_{z}$$

Η λέξη u_* πρέπει να ανήκει στην L. Έχουμε όμως

$$(q-t)+tp(p+1)+(p-q) < 2p \Leftrightarrow$$

$$(p^2+p-1)t < p \Leftrightarrow$$

$$t < \frac{p}{p^2+p-1} \quad (1)$$

Επειδή $0 < |y| \le |xy| \le p$, είναι $p \ge 1$, οπότε $p^2+p-1 \ge p^2$, οπότε προκύπτει $t < \frac{p}{p^2+p-1} \le \frac{1}{p} \le 1$, οπότε $t = 0$, άτοπο, επειδή $t = |y| = |a^t| > 0$. Συνεπώς η L δεν είναι κανονική.

Παρατήρηση:

Μπορούμε επίσης να επιλέξουμε τη λέξη $w = a^{2p-1}b^pc^p$. Με ανάλογη ανάλυση όπως και παραπάνω η w γράφεται ως

$$w = \underbrace{a^{q-t}}_{x}\ \underbrace{a^{t}}_{y}\ \underbrace{a^{2p-q-1}b^{p}c^{p}}_{z}$$

οπότε η λέξη $w_* = \underbrace{a^{q-t}}_{x}\ \underbrace{a^{2t}}_{y^2}\ \underbrace{a^{2p-q-1}b^{p}c^{p}}_{z}$ πρέπει μεν να ανήκει στην L, αλλά προκύπτει και πάλι $t < 1$, επομένως $t = 0$, άτοπο.

Αποδείξτε ότι η γλώσσα L επί του αλφάβητου $\Sigma = \{0, 1, 2\}$, η οποία περιλαμβάνει λέξεις στις οποίες το άθροισμα του πλήθους των 0 και 1 δεν είναι ίσο με το πλήθος των 2, δεν είναι κανονική.

Απάντηση

Έστω j, k, m το πλήθος των 0, 1, 2, αντίστοιχα στις λέξεις της L. Από τη εκφώνηση έχουμε ότι $j + k \neq m$.

Υποθέτουμε, αντίθετα προς το ζητούμενο, ότι η L είναι κανονική. Προφανώς η γλώσσα περιλαμβάνει άπειρο πλήθος λέξεων και επομένως μπορεί να εφαρμοστεί το Λήμμα της Άντλησης. Έστω λοιπόν $n \geq 1$ το μήκος της άντλησης και $x = 0^n 2^{n+n!}$. Για τη λέξη x είναι $j = n$, $k = 0$ και $m = n + n!$ Άρα είναι $j + k= n+0 =n \neq m = n + n!$ Για κάθε $n \geq 1$ και επομένως η λέξη x ανήκει στη γλώσσα L. Επίσης είναι $|x| = 2n + n! \geq 2n \geq n$ και επομένως σύμφωνα με το Λήμμα της Άντλησης η x μπορεί να γραφεί ως $x = uvw$ με

$|uv| \leq n$,

$|v| \geq 1$, έτσι ώστε

$uv^p w \in L$ για κάθε $p \geq 0$.

Λόγω της μορφής της λέξης x, θα είναι $u = 0^a$, $v = 0^b$, $w=0^c 2^{n+n!}$, όπου $a + b + c = n$ και $b > 0$. Δεδομένου ότι το $n!$ ($n! = n(n-1)(n-2)\ldots 1$) διαιρείται ακριβώς με b, μπορούμε να θέσουμε $p = 1 + n!/b$. Για την τιμή αυτή όμως έχουμε την εξής σχέση :

(πλήθος 0 και 1 στη λέξη $x = uv^p w$) = (πλήθος 0 στη λέξη $x = uv^p w$) = $a + b(1 + n!/b) + c = a + b + n! +c = n + n!$ = (πλήθος 2 στη λέξη $x = uv^p w$).

Η τελευταία σχέση δηλώνει ότι η λέξη x δεν ανήκει στη γλώσσα L, Φθάνοντας έτσι στη ζητούμενη αντίφαση με το αποτέλεσμα του Λήμματος της Άντλησης.