

**Αρχές Γλωσσών Προγραμματισμού και Μεταφραστών
ΣΕΤ ΑΣΚΗΣΕΩΝ #1 ΚΑΙ ΛΥΣΕΙΣ (V1.1)**

Λεκτική Ανάλυση

Γκόγκος Χρήστος

24/10/2019

Ερώτημα 1

Γράψτε κανονικές εκφράσεις που ανιχνεύουν τις ακόλουθες συμβολοσειρές δυαδικών ψηφίων χρησιμοποιώντας μόνο τα ψηφία 0 και 1, τους τελεστές * και | καθώς και παρενθέσεις.

1. Τη συμβολοσειρά 11 ή τη συμβολοσειρά 101, ή τη συμβολοσειρά 1011.
2. Συμβολοσειρές με κανένα, ένα ή περισσότερα μηδενικά και χωρίς μονάδες.
3. Όλες οι συμβολοσειρές με εξαίρεση την κενή συμβολοσειρά.
4. Όλες οι συμβολοσειρές που περιέχουν 4 συνεχόμενα 1.
5. Όλες οι συμβολοσειρές που περιέχουν 4 φορές το 1.
6. Όλες οι συμβολοσειρές με μήκος τουλάχιστον 3 και με τον τρίτο ψηφίο να είναι 0.
7. Όλες οι συμβολοσειρές με αριθμό 0 πολλαπλάσιο του 3.
8. Όλες οι συμβολοσειρές που ξεκινούν και τερματίζουν με το ίδιο ψηφίο.
9. Όλες οι συμβολοσειρές με άρτιο μήκος.
10. Όλες οι συμβολοσειρές με περιττό μήκος.

Λύσεις

1. $11|101|1011$
2. 0^*
3. $(0|1)(0|1)^*$
4. $(0|1)^*1111(0|1)^*$
5. $(0|1)^*1(0|1)^*1(0|1)^*1(0|1)^*1(0|1)^*$
6. $(0|1)(0|1)0(0|1)^*$
7. $1^*|(1^*01^*01^*01^*)^*$
8. $(0(0|1)^*0)|(1(0|1)^*1)$
9. $((0|1)(0|1))^*$
10. $(0|1)((0|1)(0|1))^*$

Ερώτημα 2

Γράψτε κανονικές εκφράσεις που ανιχνεύουν τις ακόλουθες συμβολοσειρές.

1. Ονόματα αρχείων εικόνων που τελειώνουν με 'jpg', 'png' και 'gif'.
2. Φράσεις που ξεκινούν και τελειώνουν με το κείμενο java ή JAVA.
3. Υποθετικός κωδικός που αποτελείται από 5 ζεύγη ψηφίου και λατινικού γράμματος (π.χ. 2a7w9u4z8k).
4. Μονοψήφιοι ή διψήφιοι ακέραιοι αριθμοί.

5. Ακέραιοι αριθμοί από το 0 μέχρι και το 255.
6. Αριθμοί κινητής υποδιαστολής (π.χ. 123.456, -1.2345678e-9).
7. Αριθμοί πιστωτικών καρτών που αποτελούνται από 4 ομάδες των 4 ψηφίων που χωρίζονται μεταξύ τους με κενό, παύλα, κόμμα ή τίποτε (π.χ. 1234 5678 9012 3456 ή 1234-5678-9012-3456 ή 1234,5678,9012,3456 ή 1234567890123456).
8. Διευθύνσεις email αποτελούμενες από το αναγνωριστικό χρήστη το σύμβολο @ και τον τομέα (θεωρείστε ότι επιτρεπόμενοι χαρακτήρες τόσο στο αναγνωριστικό όσο και στον τομέα είναι τα πεζά και κεφαλαία γράμματα της λατινικής αλφαβήτου και τα ψηφία. Το τελευταίο τμήμα του τομέα να αποτελείται από 2 ή 3 γράμματα του λατινικού αλφαβήτου).
9. Έγκυρες IP διευθύνσεις (4 ακέραιοι από το 0 μέχρι το 255 χωρισμένοι με τελείες, π.χ. 255.149.11.6).
10. Ημερομηνίες της μορφής yyyy-mm-dd (έτος, μήνας, ημέρα) από 1900-01-01 έως και 2099-12-31. Ως διαχωριστικό ανάμεσα στους αριθμούς να μπορεί να χρησιμοποιηθεί οποιοδήποτε από τα σύμβολα - / . και το κενό. Θεωρείστε ότι το έτος μπορεί να λάβει τιμή από το 1900 έως το 2099, ο μήνας τιμή από το 01 μέχρι το 12 και η ημέρα του μήνα μπορεί να λάβει τιμή από το 01 μέχρι το 31 (δηλαδή, για παράδειγμα η τιμή 30/02/2019 θεωρείται στα πλαίσια της άσκησης ως έγκυρη).

Μπορείτε να χρησιμοποιήσετε τις ακόλουθες επεκτάσεις των κανονικών εκφράσεων:

`\w` οποιοσδήποτε αλφαριθμητικός χαρακτήρας.

`\d` οποιοδήποτε ψηφίο.

`\s` οποιοδήποτε "λευκός" χαρακτήρας (π.χ. διάστημα, αλλαγή γραμμής, tab).

`\b` όρια λέξεων.

`.` οποιοσδήποτε χαρακτήρας.

`\.` η τελεία.

`[abc]` ο χαρακτήρας a ή ο χαρακτήρας b ή ο χαρακτήρας c.

`[^abc]` όχι ο χαρακτήρας a, ούτε ο b, ούτε ο c.

`[a-zA-Z]` οι χαρακτήρες από το a μέχρι το z, πεζοί και κεφαλαίοι.

`^a` ξεκινά με a.

`a$` τερματίζει με a.

`{m}` m επαναλήψεις.

`{m,n}` m έως n επαναλήψεις.

Λύσεις

1. `(\w+)\.(jpg|png|gif)\$`
2. `^(JAVA|java).* (JAVA|java)\$`
3. `(\d[a-z]){5}`
4. `[0-9]|[1-9][0-9]`
5. `[0-9]|[1-9][0-9]|1[0-9][0-9]|2[0-4][0-9]|25[0-5]`
6. `[-+]?[0-9]*\.[0-9]+([eE][-+]?[0-9]+)?`
7. `\d{4}[-,]?\d{4}[-,]?\d{4}[-,]\d{4}`
8. `[a-zA-Z0-9]+@[a-zA-Z0-9]+\.[a-zA-Z]{2,3}`

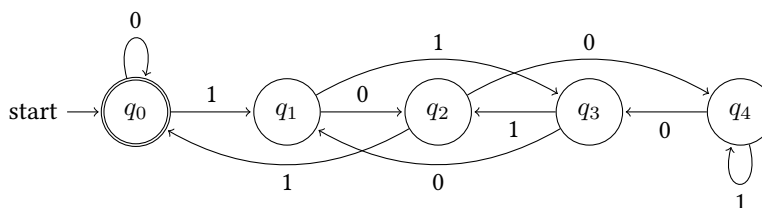
9. $\backslash b(25[0-5] | 2[0-4][0-9] | 1[0-9][0-9] | [1-9]?[0-9])\backslash .$
 $(25[0-5] | 2[0-4][0-9] | 1[0-9][0-9] | [1-9]?[0-9])\backslash .$
 $(25[0-5] | 2[0-4][0-9] | 1[0-9][0-9] | [1-9]?[0-9])\backslash .$
 $(25[0-5] | 2[0-4][0-9] | 1[0-9][0-9] | [1-9]?[0-9])\backslash b$
10. $(19 | 20)\backslash d\backslash d[- / .](0[1-9] | 1[012])[- / .](0[1-9] | [12][0-9] | 3[01])$

Μπορείτε να ελέγξετε την εγκυρότητα των παραπάνω κανονικών εκφράσεων χρησιμοποιώντας την ιστοσελίδα <http://regexr.com>.

Ερώτημα 3

Ελέγξτε αν οι ακόλουθες συμβολοσειρές αναγνωρίζονται ή όχι από το DFA του σχήματος 1. Για κάθε ακολουθία που αναγνωρίζεται επιτυχώς καταγράψτε την ακολουθία καταστάσεων που συμμετέχει στην αναγνώρισή της.

1. 101
2. 1010
3. 1011
4. 1111



Σχήμα 1: Ντετερμινιστικό πεπερασμένο αυτόματο (DFA)

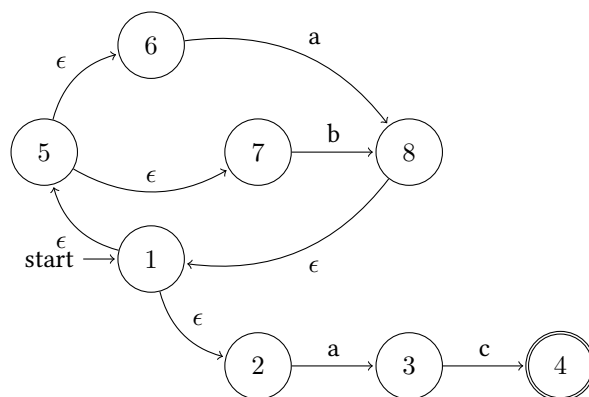
Λύση

1. $q_0 - q_1 - q_2 - q_0$ (NAI)
2. $q_0 - q_1 - q_2 - q_0 - q_0$ (NAI)
3. $q_0 - q_1 - q_2 - q_0 - q_1$ (OXI)
4. $q_0 - q_1 - q_3 - q_2 - q_0$ (NAI)

Το DFA αναγνωρίζει όλους τους δυαδικούς αριθμούς που διαιρούνται ακριβώς με το 5.

Ερώτημα 4

Να βρεθεί το ϵ -closure για καθένα από τους κόμβους του NFA στο σχήμα 2.



Σχήμα 2: Μη ντετερμινιστικό πεπερασμένο αυτόματο (NFA)

Λύση

Το ϵ -closure για το κόμβο 1 είναι το $\{1,2,5,6,7\}$.

Το ϵ -closure για το κόμβο 2 είναι το $\{2\}$.

Το ϵ -closure για το κόμβο 3 είναι το $\{3\}$.

Το ϵ -closure για το κόμβο 4 είναι το $\{4\}$.

Το ϵ -closure για το κόμβο 5 είναι το $\{5,6,7\}$.

Το ϵ -closure για το κόμβο 6 είναι το $\{6\}$.

Το ϵ -closure για το κόμβο 7 είναι το $\{7\}$.

Το ϵ -closure για το κόμβο 8 είναι το $\{8,1,2,5,6,7\}$.

Ερώτημα 5

Ελέγξτε αν οι ακόλουθες συμβολοσειρές αναγνωρίζονται ή όχι από το NFA του σχήματος 2. Για κάθε ακολουθία που αναγνωρίζεται επιτυχώς καταγράψτε την ακολουθία μεταβάσεων που την αναγνωρίζει.

1. aac
2. baac
3. aaca
4. ababc

Λύση

1. Η συμβολοσειρά aac αναγνωρίζεται μέσω της ακολουθίας μεταβάσεων: 1-5-6-8-1-2-3-4.
2. Η συμβολοσειρά baac αναγνωρίζεται μέσω της ακολουθίας μεταβάσεων: 1-5-7-8-1-5-6-8-1-2-3-4.
3. Η συμβολοσειρά aaca δεν αναγνωρίζεται.
4. Η συμβολοσειρά ababc δεν αναγνωρίζεται.

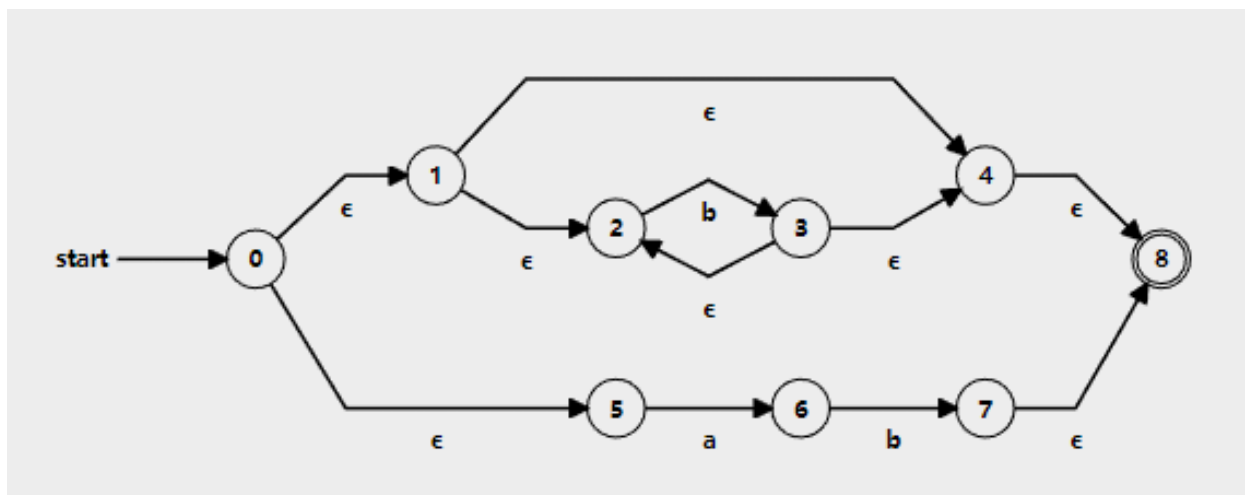
Ερώτημα 6

Μετατρέψτε τις ακόλουθες κανονικές εκφράσεις σε NFAs. Στη συνέχεια μετατρέψτε τα NFAs σε DFAs.

1. $b^*|ab$
2. $(a|b)^*$
3. $(a^*|b^*)^*$
4. $((\epsilon | a)b^*)^*$

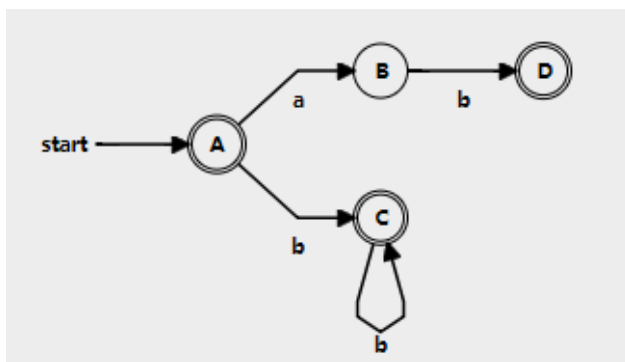
Λύση

1. $b^* | ab$



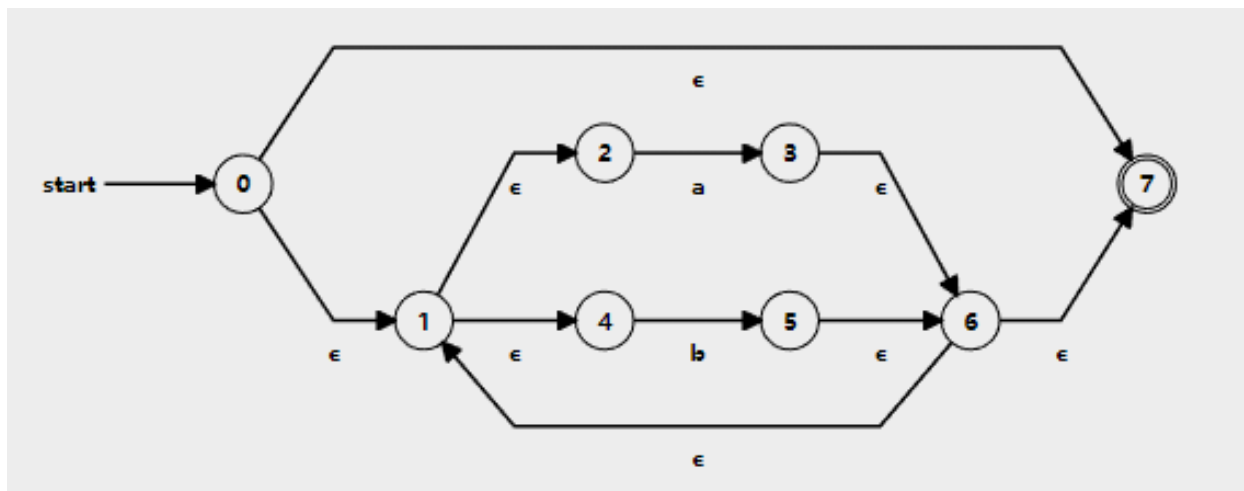
Σχήμα 3: NFA $b^*|ab$

NFA καταστάσεις	DFA καταστάσεις	a	b
{0,1,2,4,5,8}	A	B	C
{6}	B	\emptyset	D
{2,3,4,8}	C	\emptyset	C
{7,8}	D	\emptyset	\emptyset



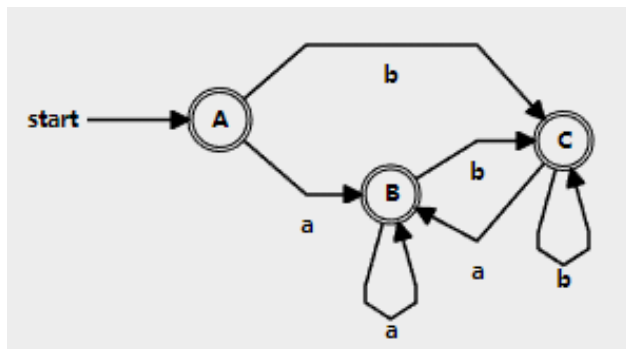
Σχήμα 4: DFA $b^*|ab$

2. $(a|b)^*$



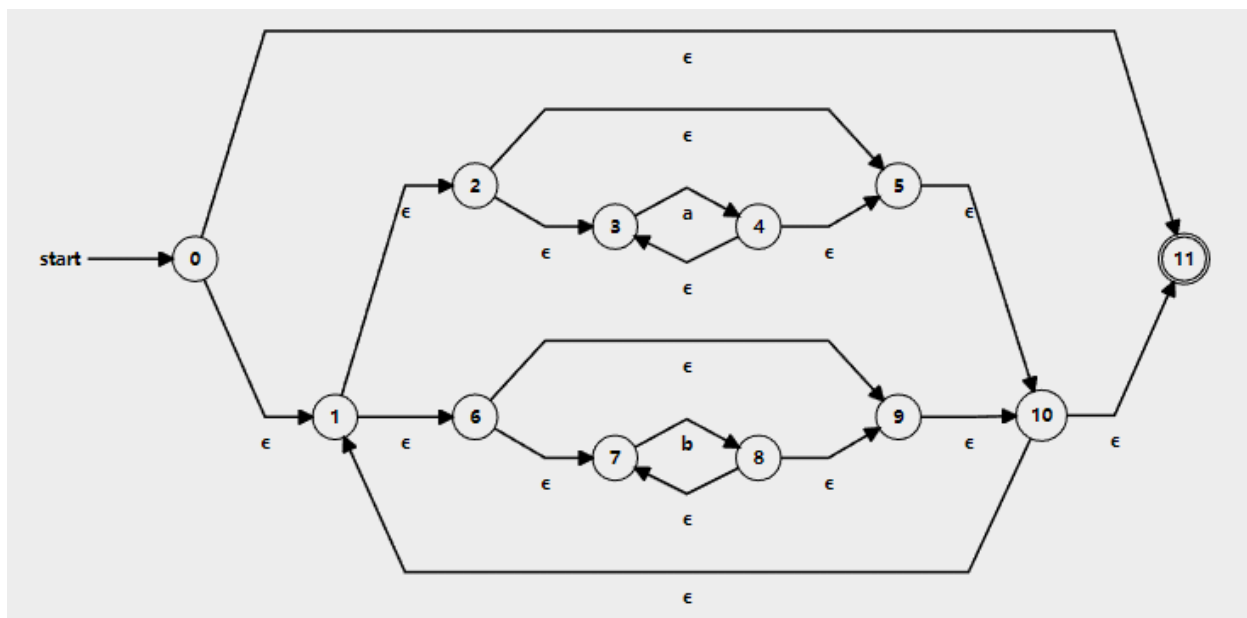
Σχήμα 5: NFA $(a|b)^*$

NFA καταστάσεις	DFA καταστάσεις	a	b
{0,1,2,4,7}	A	B	C
{1,2,3,4,6,7}	B	B	C
{1,2,4,5,6,7}	C	B	C



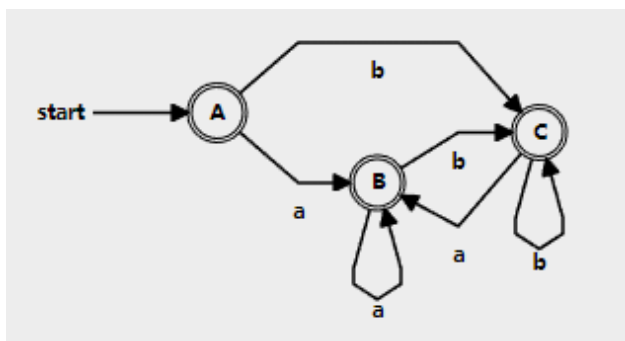
Σχήμα 6: DFA $(a|b)^*$

3. $(a^* | b^*)^*$



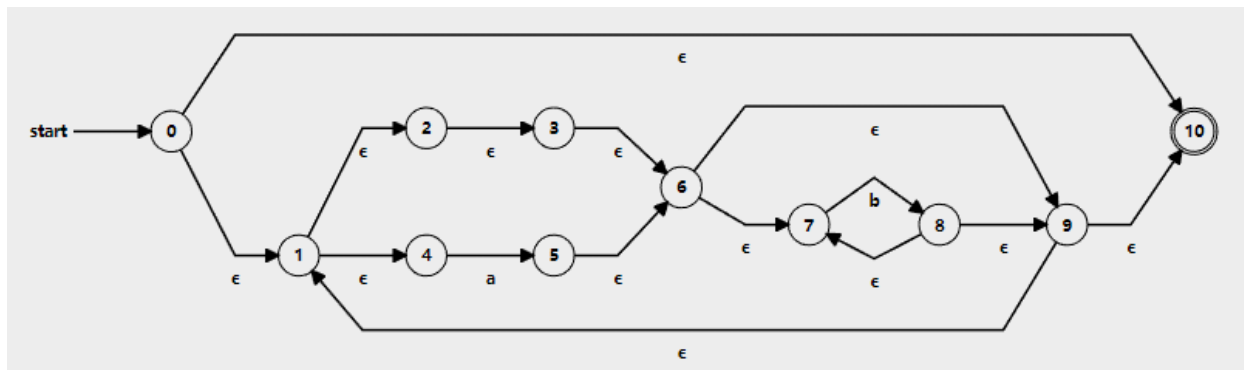
Σχήμα 7: NFA $(a^*|b^*)^*$

NFA καταστάσεις	DFA καταστάσεις	a	b
{0,1,2,3,5,6,7,9,10,11}	A	B	C
{1,2,3,4,5,6,7,9,10,11}	B	B	C
{1,2,3,5,6,7,8,9,10,11}	C	B	C



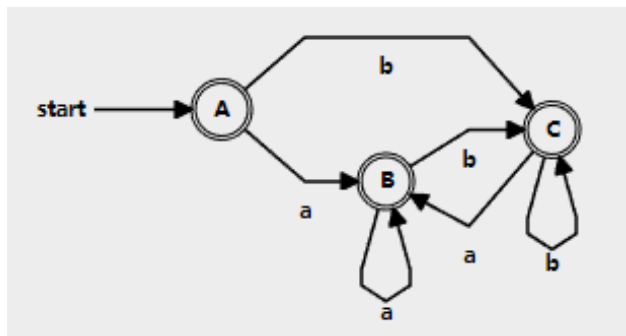
Σχήμα 8: DFA $(a^*|b^*)^*$

4. $((\epsilon|a)b^*)^*$



Σχήμα 9: NFA $((\epsilon|a)b^*)^*$

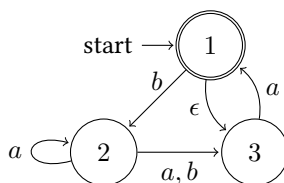
NFA καταστάσεις	DFA καταστάσεις	a	b
{0,1,2,3,4,6,7,9,10}	A	B	C
{1,2,3,4,5,6,7,9,10}	B	B	C
{1,2,3,4,6,7,8,9,10}	C	B	C



Σχήμα 10: DFA $((\epsilon|a)b^*)^*$

Ερώτημα 7

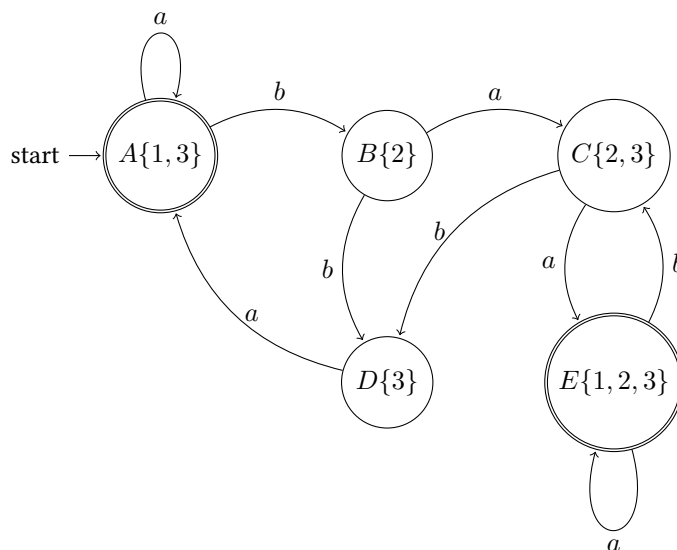
Μετατρέψτε το NFA του σχήματος 11 σε DFA.



Σχήμα 11: Μη ντετερμινιστικό πεπερασμένο αυτόματο (NFA)

Λύση

NFA καταστάσεις	DFA καταστάσεις	a	b
{ <u>1</u> ,3}	A	A	B
{2}	B	C	D
{2,3}	C	E	D
{3}	D	A	\emptyset
{ <u>1</u> ,2,3}	E	E	C

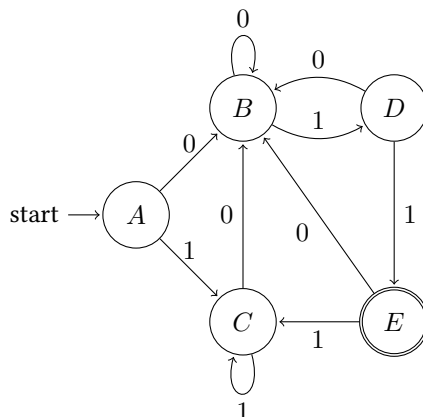


1. Η κατάσταση A του DFA είναι η κατάσταση που περιέχει το ϵ -closure της αρχικής κατάστασης του NFA. Άρα $A=\{1,3\}$.
2. Στο NFA η μετάβαση από τις καταστάσεις του NFA $\{1,3\}$ (κατάσταση A στο DFA) με την ενέργεια a οδηγεί στις καταστάσεις του NFA $\{1,3\}$ (δηλαδή στην κατάσταση A του DFA).
3. Στο NFA η μετάβαση από τις καταστάσεις του NFA $\{1,3\}$ (κατάσταση A στο DFA) με την ενέργεια b οδηγεί στην κατάσταση του NFA $\{2\}$ (δηλαδή στη νέα κατάσταση B= $\{2\}$ του DFA).
4. Στο NFA η μετάβαση από την κατάσταση του NFA $\{2\}$ (κατάσταση B στο DFA) με την ενέργεια a οδηγεί στις καταστάσεις του NFA $\{2,3\}$ (δηλαδή στη νέα κατάσταση C του DFA).
5. Στο NFA η μετάβαση από την κατάσταση του NFA $\{2\}$ (κατάσταση B στο DFA) με την ενέργεια b οδηγεί στην κατάσταση του NFA $\{3\}$ (δηλαδή στη νέα κατάσταση D του DFA).
6. Στο NFA η μετάβαση από τις καταστάσεις του NFA $\{2,3\}$ (κατάσταση C στο DFA) με την ενέργεια a οδηγεί στις καταστάσεις του NFA $\{1,2,3\}$ (δηλαδή στη νέα κατάσταση E του DFA).
7. Στο NFA η μετάβαση από τις καταστάσεις του NFA $\{2,3\}$ (κατάσταση C στο DFA) με την ενέργεια b οδηγεί στην κατάσταση του NFA $\{3\}$ (δηλαδή στην κατάσταση D του DFA).
8. Στο NFA η μετάβαση από την κατάσταση του NFA $\{3\}$ (κατάσταση D στο DFA) με την ενέργεια a οδηγεί στις καταστάσεις του NFA $\{1,3\}$ (δηλαδή στην κατάσταση A του DFA).
9. Στο NFA η μετάβαση από την κατάσταση του NFA $\{3\}$ (κατάσταση C στο DFA) με την ενέργεια b δεν οδηγεί σε καταστάσεις του NFA.

10. Στο NFA η μετάβαση από τις καταστάσεις του NFA {1,2,3} (κατάσταση E στο DFA) με την ενέργεια a οδηγεί στις καταστάσεις του NFA {1,2,3} (δηλαδή στην κατάσταση E του DFA).
11. Στο NFA η μετάβαση από τις καταστάσεις του NFA {1,2,3} (κατάσταση E στο DFA) με την ενέργεια b οδηγεί στις καταστάσεις του NFA {2,3} (δηλαδή στην κατάσταση C του DFA).

Ερώτημα 8

Κατασκευάστε DFA ελάχιστων καταστάσεων για το DFA του σχήματος 12.



Σχήμα 12: Ντετερμινιστικό πεπερασμένο αυτόματο (DFA) με 5 καταστάσεις

Λύση

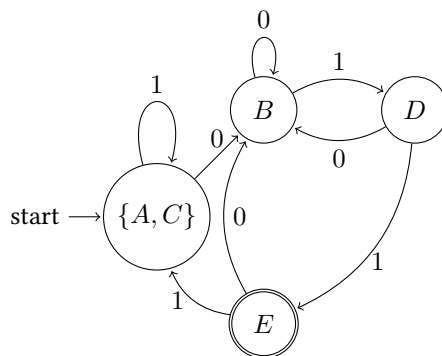
Από το DFA του σχήματος 12 σχηματίζεται ο ακόλουθος πίνακας μετάβασης (transition table).

	0	1
A	B	C
B	B	D
C	B	C
D	B	E
<u>E</u>	B	C

0-equivalence	{A,B,C,D}, {E}	A,B → OK, A,C → OK, C,D → NOT OK
1-equivalence	{A,B,C}, {D}, {E}	A,B → NOT OK, A,C → OK
2-equivalence	{A,C}, {B}, {D}, {E}	A,C → OK
3-equivalence	{A,C}, {B}, {D}, {E}	

Αναλυτική περιγραφή συμπλήρωσης του πίνακα.

- 0-equivalence: Αρχικά δημιουργείται μια διαμέριση που περιέχει 2 σύνολα, το σύνολο που περιέχει όλες τις μη τελικές καταστάσεις {A,B,C,D} και το σύνολο που περιέχει τις τελικές καταστάσεις {E}.
- 1-equivalence: Τα στοιχεία του συνόλου {A,B,C,D} εξετάζονται ανά ζεύγη προκειμένου να διαπιστωθεί αν για όλες τις πιθανές μεταβάσεις οι καταστάσεις που προκύπτουν ανήκουν στα ίδια σύνολα της προηγούμενης σειράς (0-equivalence) στον πίνακα.



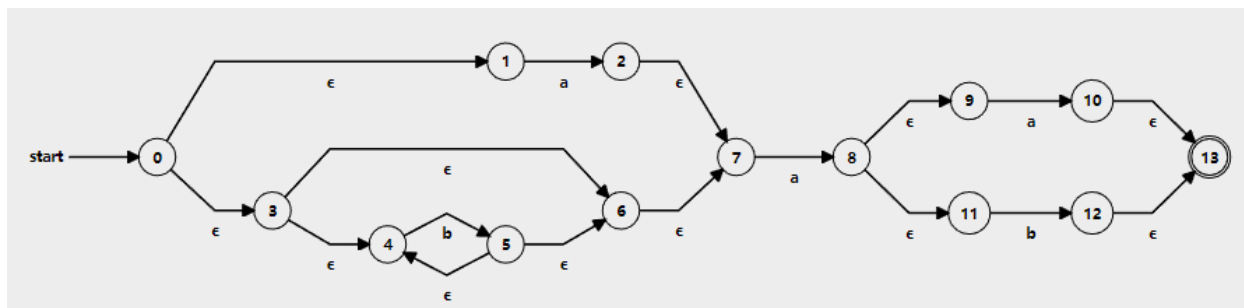
Σχήμα 13: Ντετερμινιστικό πεπερασμένο αυτόματο (DFA) με 4 καταστάσεις

- Για το ζεύγος (A,B) του συνόλου {A,B,C,D} η μετάβαση για το A με το 0 γίνεται στο B που ανήκει στο {A,B,C,D}, η μετάβαση για το B με το 0 γίνεται στο B που επίσης ανήκει στο {A,B,C,D}, η μετάβαση για το A με το 1 γίνεται στο C που ανήκει στο {A,B,C,D} και η μετάβαση για το B με το 1 γίνεται στο D που επίσης ανήκει στο {A,B,C,D}. Άρα (A,B) είναι 1-ισοδύναμα.
- Για το ζεύγος (A,C) του συνόλου {A,B,C,D}, η μετάβαση για το A με το 0 γίνεται στο B που ανήκει στο {A,B,C,D}, η μετάβαση για το C με το 0 γίνεται στο B που επίσης ανήκει στο {A,B,C,D}, η μετάβαση για το A με το 1 γίνεται στο C που ανήκει στο {A,B,C,D} και η μετάβαση για το C με το 1 γίνεται στο C που επίσης ανήκει στο {A,B,C,D}. Άρα (A,C) είναι 1-ισοδύναμα.
- Για το ζεύγος (C,D) του συνόλου {A,B,C,D}, η μετάβαση για το C με το 0 γίνεται στο B που ανήκει στο {A,B,C,D}, η μετάβαση για το D με το 0 γίνεται στο B που επίσης ανήκει στο {A,B,C,D}, η μετάβαση για το C με το 1 γίνεται στο C που ανήκει στο {A,B,C,D} και η μετάβαση για το D με το 1 γίνεται στο E που δεν ανήκει στο {A,B,C,D} αλλά στο {E}. Άρα τα C και D δεν είναι 1-ισοδύναμα και δημιουργούνται τα νέα σύνολα {A,B,C} και {D}.
- 2-equivalence: Τα στοιχεία του συνόλου {A,B,C} εξετάζονται ανά ζεύγη προκειμένου να διαπιστωθεί αν για όλες τις πιθανές μεταβάσεις οι καταστάσεις που προκύπτουν ανήκουν στα ίδια σύνολα της προηγούμενης σειράς (1-equivalence) στον πίνακα.
 - Για το ζεύγος (A,B) του συνόλου {A,B,C} η μετάβαση για το A με το 0 γίνεται στο B που ανήκει στο {A,B,C}, η μετάβαση για το B με το 0 γίνεται στο B που επίσης ανήκει στο {A,B,C}, η μετάβαση για το A με το 1 γίνεται στο C που ανήκει στο {A,B,C} και η μετάβαση για το B με το 1 γίνεται στο D που δεν ανήκει στο {A,B,C} αλλά στο {D}. Άρα (A,B) δεν είναι 2-ισοδύναμα και δημιουργούνται τα σύνολα {A,C} και {B}.
 - Για το ζεύγος (A,C) του συνόλου {A,B,C}, η μετάβαση για το A με το 0 γίνεται στο B που ανήκει στο {A,B,C}, η μετάβαση για το C με το 0 γίνεται στο B που επίσης ανήκει στο {A,B,C}, η μετάβαση για το A με το 1 γίνεται στο C που ανήκει στο {A,B,C} και η μετάβαση για το C με το 1 γίνεται στο C που επίσης ανήκει στο {A,B,C}. Άρα (A,C) είναι 2-ισοδύναμα.
- 3-equivalence: Τα στοιχεία του συνόλου {A,C} εξετάζονται ανά ζεύγη προκειμένου να διαπιστωθεί αν για όλες τις πιθανές μεταβάσεις οι καταστάσεις που προκύπτουν ανήκουν στα ίδια σύνολα της προηγούμενης σειράς (2-equivalence) στον πίνακα.
 - Για το ζεύγος (A,C) του συνόλου {A,C}, η μετάβαση για το A με το 0 γίνεται στο B που ανήκει στο {B}, η μετάβαση για το C με το 0 γίνεται στο B που επίσης ανήκει στο {B}, η μετάβαση για το A με το 1 γίνεται στο C που ανήκει στο {A,C} και η μετάβαση για το C με το 1 γίνεται στο C που επίσης ανήκει στο {A,C}. Άρα (A,C) είναι 3-ισοδύναμα. Εφόσον δεν προκύπτει νέα διαμέριση ο αλγόριθμος τερματίζει.

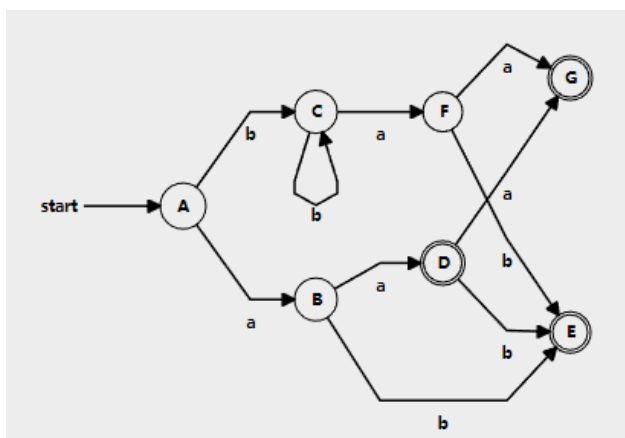
Ερώτημα 9

Κατασκευάστε DFA ελάχιστων καταστάσεων για την κανονική έκφραση $(a|b^*)a(a|b)$. Πρώτα μετατρέψτε την κανονική έκφραση σε NFA, μετά το NFA σε DFA και τέλος το DFA σε DFA ελάχιστων καταστάσεων.

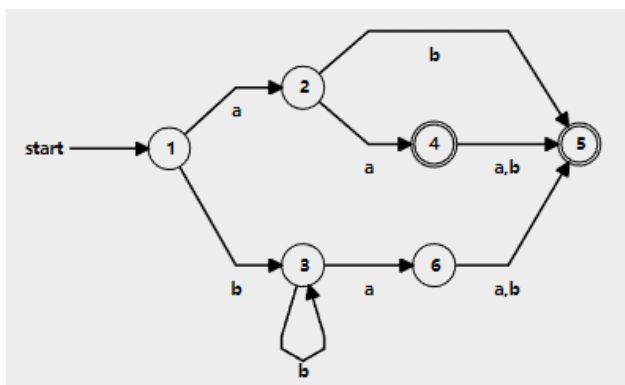
Λύση



Σχήμα 14: NFA $(a|b^*)a(a|b)$



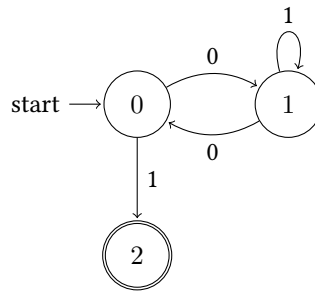
Σχήμα 15: DFA $(a|b^*)a(a|b)$



Σχήμα 16: minDFA $(a|b^*)a(a|b)$

Ερώτημα 10

Για το DFA του σχήματος 17 κατασκευάστε ισοδύναμο κώδικα σε C.



Σχήμα 17: Ντετερμινιστικό πεπερασμένο αυτόματο (DFA) που αναγνωρίζει συμβολοσειρές αποτελούμενες μόνο από 0 και 1 και οι οποίες ξεκινούν με άρτιο πλήθος 0 και τελειώνουν με 1

Λύση

```
int match(char *next)
{
S0: if (*next == '\0')
    return 0;
    if (*next == '0')
    {
        next++; goto S1;
    }
    if (*next == '1')
    {
        next++; goto S2;
    }
    return 0;
S1: if (*next == '\0')
    return 0;
    if (*next == '0')
    {
        next++; goto S0;
    }
    if (*next == '1')
    {
        next++; goto S1;
    }
    return 0;

S2:
    if (*next == '\0')
        return 1;
    return 0;
}
```

Listing 1: Η συνάρτηση match