1η ΔΙΑΛΕΞΗ
ΤΥΠΙΚΕΣ ΓΛΩΣΣΕΣ - ΑΥΤΟΜΑΤΑ

- **Τυπικές γλώσσες**
  - **Αλφάβητο, Λέξεις**
  - **Σύζευξη λέξεων**
  - **Τυπικές γλώσσες**
- D-**Αυτόματα**
  - **Λέξεις και αυτόματα**
  - **Γλώσσες και αυτόματα**

Αλφάβητο

Αλφάβητο είναι ένα μη κενό, πεπερασμένο σύνολο συμβόλων $\mathcal{E}$.

Παραδείγματα

1. Το ελληνικό αλφάβητο $\mathcal{E}_1 = \{\mathrm{A}, \mathrm{B}, \Gamma, \ldots, \Omega\}$.
2. Το λατινικό αλφάβητο $\mathcal{E}_2 = \{\mathrm{A}, \mathrm{B}, \mathrm{C}, \ldots, \mathrm{Z}\}$.
3. Το δυαδικό αλφάβητο $\mathcal{E}_3 = \{0, 1\}$.
4. Το αλφάβητο της λογικής $\mathcal{E}_4 = \{a, b, c, \ldots, z, (,), \vee, \wedge, \rightarrow, \leftrightarrow, \neg\}$.

Τα στοιχεία ενός αλφαβήτου ονομάζονται **γράμματα**.

Λέξεις

Κάθε πεπερασμένη ακολουθία γραμμάτων ονομάζεται **λέξη**.

Παραδείγματα

1. Για το $\mathcal{E}_1$, λέξεις είναι οι ΝΙΚΟΣ, ΑΑΑΒΑΩ, Β, ΤΡΑΠΕΖ, κ.λπ.
2. Για το $\mathcal{E}_2$, λέξεις είναι οι BOAT, B, CCCCQ, κ.λπ.
3. Για το $\mathcal{E}_3$, λέξεις είναι οι 10010, 11, 1010, κ.λπ.
4. Για το $\mathcal{E}_4$, λέξεις είναι οι $(a \to b)$, $(a \vee b) \wedge (c \vee d)$, $a \vee (\neg a)$, κ.λπ.

Ορίζουμε επίσης μια λέξη χωρίς γράμματα, την **κενή λέξη** $\Box$ ή ϵ ή λ.

Ο αριθμός των γραμμάτων μιας λέξης w λέγεται **μήκος** της λέξης και συμβολίζεται με $l(w)$ ή $|w|$.
Έτσι η πρώτη λέξη καθενός από τα τέσσερα προηγούμενα παραδείγματα έχει μήκος $5, 4, 5$ και 5 αντίστοιχα.

Το σύνολο όλων των λέξεων

Το σύνολο **όλων των λέξεων** ενός αλφαβήτου $\mathcal{E}$ συμβολίζεται με $\mathcal{E}^*$.

Παραδείγματα

1. Για το αλφάβητο $\mathcal{E}_1 = \{A, B, \Gamma, \ldots, \Omega\}$ παίρνουμε ως $\mathcal{E}_1^*$ το σύνολο όλων των λέξεων που χρησιμοποιούν οσαδήποτε, οποιαδήποτε και με οποιαδήποτε σειρά γράμματα του ελληνικού αλφαβήτου (πεπερασμένα σε πλήθος) καθώς επίσης και την $\square$.
2. Για το αλφάβητο $\mathcal{E}_3 = \{0, 1\}$ έχουμε $\mathcal{E}_3^* = \{\square, 0, 1, 00, 01, 10, 11, 000, 001, 010, \ldots\}$.

Σύζευξη λέξεων

Ορίζουμε μια εσωτερική πράξη στο $\mathcal{E}^*$ ως εξής:
Για κάθε $w_1, w_2 \in \mathcal{E}^*$ παίρνουμε την $w = w_1 w_2$ που προκύπτει από τα γράμματα της w_1 ακολουθούμενα από τα γράμματα της w_2. Η πράξη αυτή λέγεται **ζεύξη** ή **σύζευξη**.

Παραδείγματα
Αν $w_1 = 1011$ και $w_2 = 11$ τότε $w_1 w_2 = 101111$ και $w_2 w_1 = 111011$.
Αν $w_1 = \mathrm{ANTI}$ και $w_2 = \Gamma\mathrm{PA}\Phi\Omega$ τότε $w_1 w_2 = \mathrm{ANTI}\Gamma\mathrm{PA}\Phi\Omega$.

- Ισχύει η προσεταιριστικότητα αλλά όχι η αντιμεταθετικότητα, όπως φαίνεται και από το πρώτο παράδειγμα.
- Επίσης, υπάρχει ουδέτερο στοιχείο (η λέξη $\square$, αφού $w\square = \square w = w$).

Για την λέξη w, γράφουμε w^ν αντί $\underbrace{ww\cdots w}_{\nu \text{ φορές}}$. Θεωρούμε ότι $w^0 = \square$.

Η αλγεβρική δομή $\mathcal{E}^*$ με την πράξη της σύζευξης λέγεται **ελεύθερο μονοειδές** (με γεννήτορες τα στοιχεία του $\mathcal{E}$).

Τυπική γλώσσα

Κάθε υποσύνολο L του $\mathcal{E}^*$ ονομάζεται **τυπική γλώσσα του** $\mathcal{E}^*$.

Παραδείγματα

1. $\mathcal{E}$ το ελληνικό αλφάβητο και L όλες οι λέξεις της ελληνικής γλώσσας.
2. $\mathcal{E} = \{0, 1, 2, 3, \dots, 9\}$ και L όλοι οι αριθμοί x με $100 \leq x \leq 355$.

Το κενό υποσύνολο $\emptyset$ του $\mathcal{E}^*$ ονομάζεται **κενή γλώσσα**.

Ασκήσεις

Άσκηση 1

Να γραφεί το σύνολο των λέξεων μήκους 4 που αρχίζουν από a όταν $\mathcal{E} = \{a, b, c\}$.

Απάντηση. $\{aaaa, aaab, aaac, aaba, aabb,$
$aabc, aaca, aacb, aacc, abaa, abab, abac,$
$abba, abbb, abbc, abca, abcb, abcc, acaa, acab,$
$acac, acba, acbb, acbc, acca, accb, accc\}$

Άσκηση 2

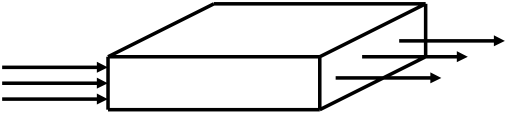
Να γραφεί το σύνολο των λέξεων, μήκους μικρότερου ή ίσου του 4, στις οποίες το a έχει περιττό αριθμό εμφανίσεων όταν $\mathcal{E} = \{a, b\}$.

Άσκηση 3

Να γραφεί το σύνολο των λέξεων, μήκους μικρότερου ή ίσου του 4, στις οποίες δεν εμφανίζονται τρία διαδοχικά b όταν $\mathcal{E} = \{a, b\}$.

Αυτόματα

Ένα κουτί με διαύλους εισόδου και εξόδου:



Είσοδος Κουτί Έξοδος

Χρήστης $\longrightarrow$ Αυτόματο Αυτόματο $\longrightarrow$ Χρήστης

Αυτόματα με πεπερασμένο αριθμό καταστάσεων (θέσεων) (Finite state automata, πεπερασμένα αυτόματα).
Το σύνολο των καταστάσεων καθώς και η διαδικασία που επιτρέπει τη μετάβαση από τη μια κατάσταση στην άλλη (με τη βοήθεια ''σημάτων'') είναι τα βασικά στοιχεία ενός αυτόματου.

D-αυτόματα

Στη συνέχεια θα ασχοληθούμε με τα D-αυτόματα:

Κάθε πεντάδα $a = (S, \mathcal{E}, T, s_0, f)$ όπου
S: ένα πεπερασμένο σύνολο καταστάσεων,
$\mathcal{E}$: ένα αλφάβητο (το αλφάβητο εισόδου),
$T \subseteq S$: το σύνολο των τελικών καταστάσεων,
$s_0 \in S$: η αρχική κατάσταση,
$f : S \times \mathcal{E} \to S$ (η συνάρτηση μετάβασης),
ονομάζεται **(πεπερασμένο)** D-**αυτόματο (deterministic finite state automaton)**.
Η συνάρτηση μετάβασης f, εκφράζει τον ‘‘εσωτερικό μηχανισμό’’ του D-αυτόματου, και **σε κάθε ζεύγος του** $S \times \mathcal{E}$ αντιστοιχεί **ακριβώς ένα στοιχείο του** S.

Σε κάθε D-αυτόματο αντιστοιχεί ένα γράφημα τόξων με ετικέτες στα τόξα.

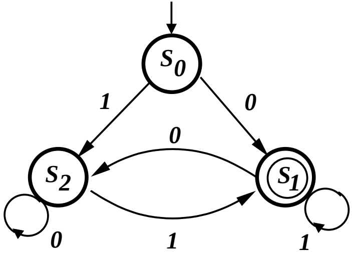
- Κορυφές του γραφήματος είναι οι καταστάσεις του αυτόματου,
- Σε κάθε τιμή $f(s_i, k) = s_j$ της f αντιστοιχεί το τόξο (s_i, s_j) με ετικέτα k.
- Στην κορυφή της αρχικής κατάστασης s_0 υπάρχει ένα τόξο εισόδου,
- Για τις κορυφές των τελικών καταστάσεων χρησιμοποιούνται διπλοί κύκλοι, αντί για τον ένα που χρησιμοποιείται για τις υπόλοιπες κορυφές.

Παραδείγματα

1. $S = \{s_0, s_1, s_2\}$, $\mathcal{E} = \{0, 1\}$, $T = \{s_1\}$ και $f : S \times \mathcal{E} \to S$ με

   $f(s_0, 0) = s_1 \quad f(s_0, 1) = s_2$
   $f(s_1, 0) = s_2 \quad f(s_1, 1) = s_1$
   $f(s_2, 0) = s_2 \quad f(s_2, 1) = s_1$

   Το αντίστοιχο γράφημα (που περιγράφει αυτό το D-αυτόματο) είναι το εξής:

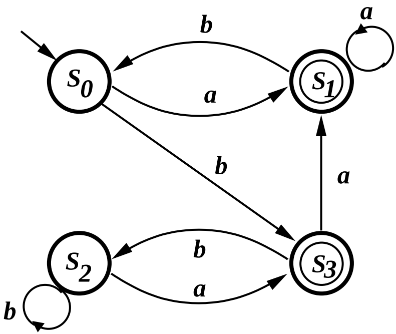


Παραδείγματα

1. $S = \{s_0, s_1, s_2, s_3\}$, $\mathcal{E} = \{a, b\}$, $T = \{s_1, s_3\}$ και $f : S \times \mathcal{E} \to S$ με

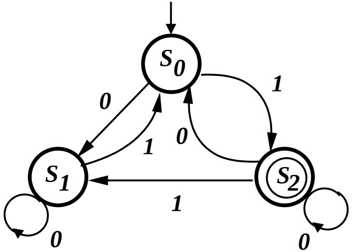
$$
\begin{array}{ll}
f(s_0, a) = s_1 & f(s_0, b) = s_3 \\
f(s_1, a) = s_1 & f(s_1, b) = s_0 \\
f(s_2, a) = s_3 & f(s_2, b) = s_2 \\
f(s_3, a) = s_1 & f(s_3, b) = s_2.
\end{array}
$$

Το αντίστοιχο γράφημα (που περιγράφει αυτό το D-αυτόματο) είναι το εξής:



Παραδείγματα

1. Το αυτόματο που περιγράφει το επόμενο γράφημα **δεν είναι** ντετερμινιστικό (αφού η $f(s_2, 0)$ δεν είναι μοναδικά καθορισμένη).



Λέξεις και αυτόματα

Για να εξετάσουμε αν ένα D-αυτόματο "δέχεται" (αναγνωρίζει) μια λέξη, ορίζουμε πρώτα μια καινούργια συνάρτηση $f^* : S \times \mathcal{E}^* \to S$ ως εξής:

$$f^*(s_i, \square) = s_i, \;\; \forall s_i \in S$$

$$f^*(s_i, m_1 m_2) = f(f^*(s_i, m_1), m_2), \;\; \forall s_i \in S, m_1 \in \mathcal{E}^*, m_2 \in \mathcal{E}.$$

Μια λέξη m **αναγνωρίζεται** από ένα D-αυτόματο, όταν $f^*(s_0, m) \in T$.

Δηλαδή πρέπει, αν ξεκινήσουμε από το s_0 και προχωρήσουμε "βήμα-βήμα", ακολουθώντας τα "γράμματα" της λέξης m, να καταλήξουμε σε κάποιο "τελικό" στοιχείο, δηλαδή σε κάποιο στοιχείο του T.

Παραδείγματα

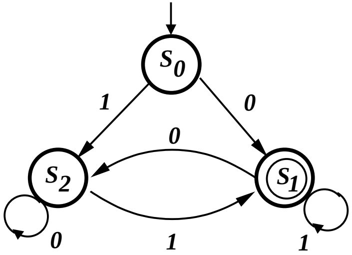
1. Το D-αυτόματο του παραδείγματος 1:
   $S = \{s_0, s_1, s_2\}$, $\mathcal{E} = \{0, 1\}$, $T = \{s_1\}$ και $f : S \times \mathcal{E} \to S$ με

   $$\begin{array}{ll} f(s_0, 0) = s_1 & f(s_0, 1) = s_2 \\ f(s_1, 0) = s_2 & f(s_1, 1) = s_1 \\ f(s_2, 0) = s_2 & f(s_2, 1) = s_1 \end{array}$$

   αναγνωρίζει τη λέξη $m = 1001101$, αφού

   $$f^*(s_0, 1) = f^*(s_0, \square 1) = f(f^*(s_0, \square), 1) = f(s_0, 1) = s_2$$
   $$f^*(s_0, 10) = f(f^*(s_0, 1), 0) = f(s_2, 0) = s_2$$
   $$f^*(s_0, 100) = f(f^*(s_0, 10), 0) = f(s_2, 0) = s_2$$
   $$f^*(s_0, 1001) = f(f^*(s_0, 100), 1) = f(s_2, 1) = s_1$$
   $$f^*(s_0, 10011) = f(f^*(s_0, 1001), 1) = f(s_1, 1) = s_1$$
   $$f^*(s_0, 100110) = f(f^*(s_0, 10011), 0) = f(s_1, 0) = s_2$$
   $$f^*(s_0, m) = f^*(s_0, 1001101) = f(f^*(s_0, 100110), 1) = f(s_2, 1) = s_1 \in T.$$

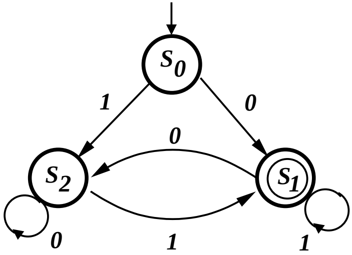
Παρατήρηση. Πρακτικός τρόπος: Ξεκινάμε από το s_0 και ακολουθούμε τα βήματα της $m = 1001101$



$$s_0 \xrightarrow{1} s_2 \xrightarrow{0} s_2 \xrightarrow{0} s_2 \xrightarrow{1} s_1 \xrightarrow{1} s_1 \xrightarrow{0} s_2 \xrightarrow{1} s_1 \in T.$$

Παραδείγματα

1. Το ίδιο D-αυτόματο **δεν** αναγνωρίζει τη λέξη $w = 11110010$ διότι

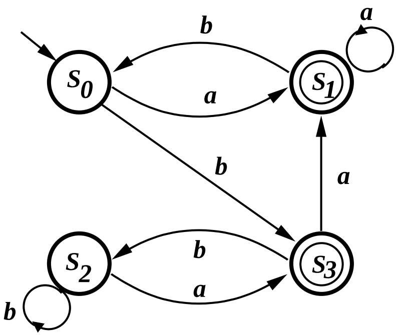


$$s_0 \xrightarrow{1} s_2 \xrightarrow{1} s_1 \xrightarrow{1} s_1 \xrightarrow{1} s_1 \xrightarrow{0} s_2 \xrightarrow{0} s_2 \xrightarrow{1} s_1 \xrightarrow{0} s_2 \notin T.$$

Παραδείγματα

1. Το D-αυτόματο του παραδείγματος 2:
$S = \{s_0, s_1, s_2, s_3\}$, $\mathcal{E} = \{a, b\}$, $T = \{s_1, s_3\}$ και $f : S \times \mathcal{E} \to S$ με

$$\begin{array}{ll} f(s_0, a) = s_1 & f(s_0, b) = s_3 \\ f(s_1, a) = s_1 & f(s_1, b) = s_0 \\ f(s_2, a) = s_3 & f(s_2, b) = s_2 \\ f(s_3, a) = s_1 & f(s_3, b) = s_2. \end{array}$$



αναγνωρίζει τη λέξη $m = abbbbabaa$ διότι

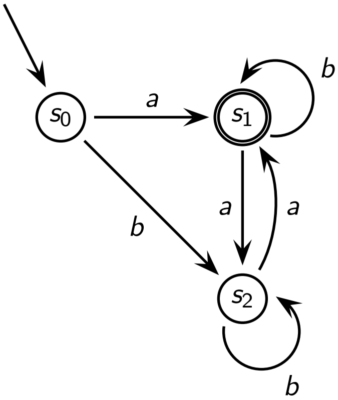
$$s_0 \xrightarrow{a} s_1 \xrightarrow{b} s_0 \xrightarrow{b} s_3 \xrightarrow{b} s_2 \xrightarrow{b} s_2 \xrightarrow{a} s_3 \xrightarrow{b} s_2 \xrightarrow{a} s_3 \xrightarrow{a} s_1 \in T.$$

Γλώσσες και αυτόματα

Ένα D-αυτόματο **αναγνωρίζει** μια γλώσσα L (ή, ισοδύναμα μια γλώσσα L **αναγνωρίζεται** από ένα D-αυτόματο), αν το D-αυτόματο αναγνωρίζει όλες τις λέξεις της L και μόνον αυτές.

Παράδειγμα

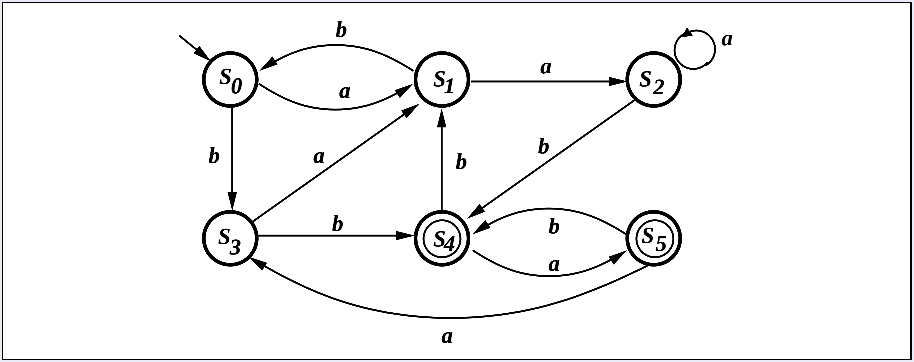
Το παρακάτω D-αυτόματο αναγνωρίζει τη γλώσσα του $\{a, b\}^*$ η οποία αποτελείται από τις λέξεις στις οποίες το a έχει περιττό αριθμό εμφανίσεων.



Ασκήσεις

Άσκηση 1

Να εξετασθεί αν το D-αυτόματο



αναγνωρίζει τις παρακάτω λέξεις: $w_1 = abbaabbab$, $w_2 = bbaabbaab$, $w_3 = babababa$.

Απάντηση. Το αυτόματο αναγνωρίζει τις w_1, w_2 αλλά όχι την w_3.

Άσκηση 2

1. *Να δοθεί το γράφημα που περιγράφει το D-αυτόματο* $(S, \mathcal{E}, T, s_0, f)$ *όπου*

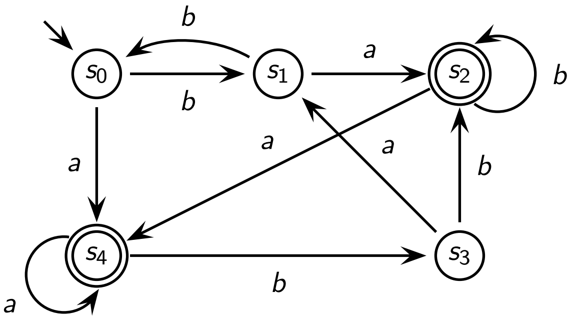
$$S = \{s_0, s_1, s_2, s_3, s_4\}, \mathcal{E} = \{a, b\},$$
$$T = \{s_2, s_4\} \text{ και } f : S \times \mathcal{E} \to S$$

με

$$\begin{array}{ll} f(s_0, a) = s_4 & f(s_0, b) = s_1 \\ f(s_1, a) = s_2 & f(s_1, b) = s_0 \\ f(s_2, a) = s_4 & f(s_2, b) = s_2 \\ f(s_3, a) = s_1 & f(s_3, b) = s_2 \\ f(s_4, a) = s_4 & f(s_4, b) = s_3 \end{array}$$

2. *Να εξετασθεί αν το αυτόματο αυτό αναγνωρίζει τις λέξεις:* $w_1 = abababab$, $w_2 = baaabbabaaab$ *και* $w_3 = baabbaabaa$.
3. *Έστω* u *μια λέξη που αποτελείται από* m *σε πλήθος* a *και* v *μια λέξη που αποτελείται από* n *σε πλήθος* b $(m, n \in \mathbb{N}^*)$. *Να δειχθεί ότι το παραπάνω D-αυτόματο αναγνωρίζει τη λέξη* $w = vu$.
4. *Κάτω από ποιες προϋποθέσεις αναγνωρίζει τη λέξη* $w' = uv$;

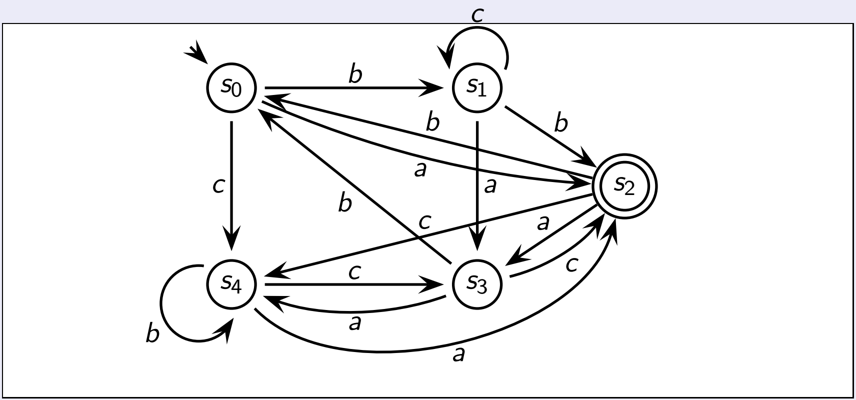
Απάντηση.

1. 

2. $f^*(w_1) = s_0 \notin T$ (άρα η w_1 δεν αναγνωρίζεται).
$f^*(w_2) = s_3 \notin T$ (άρα η w_2 δεν αναγνωρίζεται).
$f^*(w_3) = s_2 \in T$ (άρα η w_3 αναγνωρίζεται).
3. Αν ο n είναι άρτιος τότε $f^*(b^n) = s_0$, οπότε $f^*(b^n a^m) = s_4 \in T$.
Αν ο n είναι περιττός τότε $f^*(b^n) = s_1$, οπότε αν $m = 1$, τότε $f^*(b^n a) = s_2 \in T$, ενώ αν $m \geq 2$ τότε $f^*(b^n a^m) = s_4 \in T$.
Αφού λοιπόν σε κάθε περίπτωση $f^*(w) \in T$, τότε η $w = b^n a^m$ αναγνωρίζεται από το αυτόματο.

Άσκηση 3

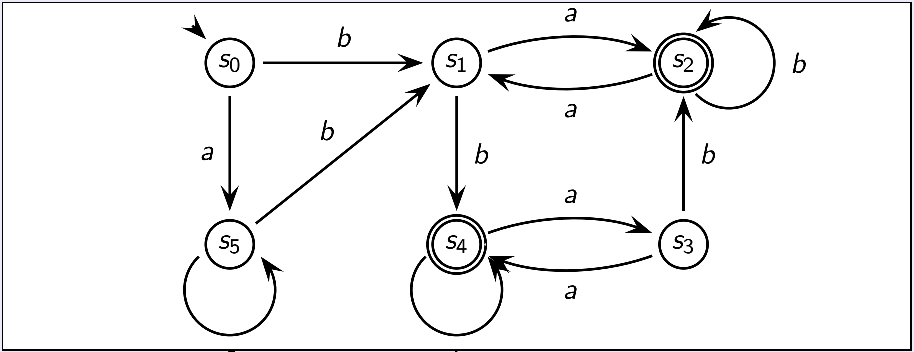
1. *Να ορισθεί αναλυτικά η συνάρτηση f του D-αυτόματου $(S, \mathcal{E}, T, s_0, f)$ το οποίο περιγράφεται από το επόμενο γράφημα:*



2. *Να εξετασθεί αν το παραπάνω D-αυτόματο αναγνωρίζει τις λέξεις:* $w_1 = abcabcabc$, $w_2 = aaabbbccc$, $w_3 = baccabbabcc$.
3. *Να δοθούν όλες οι δυνατές τιμές των $\mu, \nu \in \mathbb{N}^*$, ώστε το παραπάνω αυτόματο να αναγνωρίζει τη λέξη $w = a^3 b^\mu c^\nu$.*

Άσκηση 4

Να ορισθούν αναλυτικά τα σύνολα $S, \mathcal{E}, T$ *και η συνάρτηση* f *του D-αυτόματου* $(S, \mathcal{E}, T, s_0, f)$ *που περιγράφεται από το επόμενο γράφημα:*



Να εξετασθεί αν το αυτόματο αυτό αναγνωρίζει τις λέξεις:
$w_1 = baaabbabbba$ *και* $w_2 = aaaabbaaaaba$.
Να εξηγηθεί γιατί το αυτόματο αυτό δεν αναγνωρίζει οποιαδήποτε ζεύξη $w_1 w_2$, *με* $w_2 = baba$.

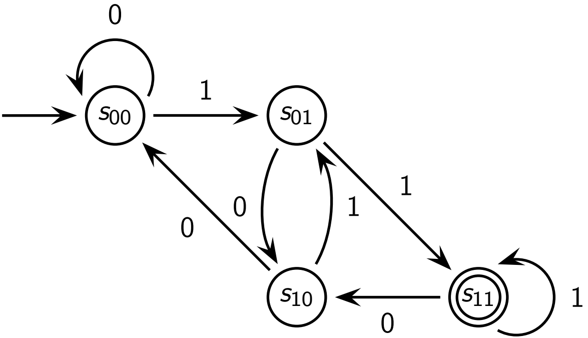
Άσκηση 5

Να παρασταθεί με γράφημα το D-αυτόματο για το οποίο

$$S = \{s_{00}, s_{01}, s_{10}, s_{11}\},$$
$$\mathcal{E} = \{0, 1\}, s_0 = s_{00}, T = \{s_{11}\}$$

και με f τέτοια ώστε $f(s_{ij}, k) = s_{jk}$.

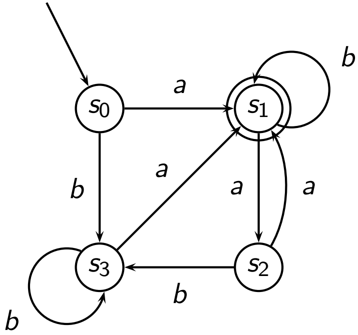
Απάντηση



Άσκηση 6

Να βρεθεί ένα D-αυτόματο με $\mathcal{E} = \{a, b\}$ *και* $|S| = 4$, *που να αναγνωρίζει τη γλώσσα η οποία αποτελείται από τις λέξεις στις οποίες το* a *έχει περιττό αριθμό εμφανίσεων.*

Απάντηση. Ένα τέτοιο D-αυτόματο είναι το παρακάτω:



2η ΔΙΑΛΕΞΗ
ΤΥΠΙΚΕΣ ΓΛΩΣΣΕΣ - ΑΥΤΟΜΑΤΑ

- Πράξεις στις γλώσσες
  - Ένωση
  - Σύζευξη
  - Δυνάμεις
  - Κλειστότητα του Kleene
  - Συμπλήρωμα
- Κανονικές γλώσσες
  - Κανονικές εκφράσεις
  - Κανονικές γλώσσες
    - Θεώρημα του Kleene
  - Pumping lemma
- Κανόνες αντικατάστασης
  - Το πρόβλημα των λέξεων

Πράξεις στις γλώσσες

Αν A, B είναι δύο τυπικές γλώσσες του $\mathcal{E}^*$, τότε ορίζουμε

- Την **ένωση** $A \cup B$ των A και B ως το σύνολο όλων των λέξεων x όπου $x \in A$ ή/και $x \in B$.
  Παράδειγμα Αν $A = \{\square, 0, 11\}$ και $B = \{1, 10, 110\}$ τότε

$$A \cup B = \{\square, 0, 11, 1, 10, 110\}$$

- Την **σύζευξη** AB των A και B ως το σύνολο όλων των λέξεων xy όπου $x \in A$ και $y \in B$.
  Παράδειγμα Αν $A = \{\square, 0, 11\}$ και $B = \{1, 10, 110\}$ τότε

$$AB = \{1, 10, 110, 01, 010, 0110, 111, 1110, 11110\}$$

$$BA = \{1, 10, 111, 10, 100, 1011, 110, 1100, 11011\}$$

  Αν $B = \emptyset$ τότε $AB = BA = \emptyset$.

Πράξεις στις γλώσσες

- Αν A είναι μια τυπική γλώσσα τότε ορίζουμε επαγωγικά την n-**οστή δύναμη της** A ως εξής:

$$A^0 = \{\square\}$$

και

$$A^{n+1} = A^n A, \text{ για κάθε } n \in \mathbb{N}$$

Παράδειγμα Έστω $A = \{0, 11\}$ να βρεθούν οι γλώσσες A^n για $n = 0, 1, 2, 3$.
$A^0 = \{\square\}$
$A^1 = A^0 A = \{\square\}\{0, 11\} = \{0, 11\} = A$
$A^2 = A^1 A = \{0, 11\}\{0, 11\} = \{00, 011, 110, 1111\}$
$A^3 = A^2 A = \{00, 011, 110, 1111\}\{0, 11\} =$
$\{000, 0011, 0110, 01111, 1100, 11011, 11110, 111111\}$

- Το **άστρο του Kleene** ή **κλειστότητα του Kleene** L^* μιας τυπικής γλώσσας L είναι το σύνολο που αποτελείται από τις συζεύξεις οσονδήποτε λέξεων της γλώσσας L, δηλαδή

$$L^* = L^0 \cup L^1 \cup L^2 \cup \cdots = \bigcup_{k=0}^{\infty} L^k$$

  Παραδείγματα
  Αν $L = 0$ τότε $L^* = \{0^n : n \in \mathbb{N}\}$
  Αν $L = 11$ τότε $L^* = \{(11)^n : n \in \mathbb{N}\}$ (το σύνολο όλων των λέξεων που αποτελούνται από άρτιο αριθμό 1.
  Αν $L = \{00, 11\}$ τότε L^* είναι το σύνολο όλων των δυαδικών λέξεων οι οποίες δεν περιέχουν τις υπολέξεις 101 και 010. Επίσης, στο L δεν περιλαμβάνονται οι λέξεις 0 και 1.

- Επίσης με L^+ συμβολίζεται το σύνολο $L^* \setminus \{\square\}$.

Πράξεις στις γλώσσες

- **Συμπλήρωμα** μιας γλώσσας $L \subseteq \mathcal{E}^*$ ονομάζεται το σύνολο $\overline{L} = \{w : w \in \mathcal{E}^* \text{ και } w \notin L\}$.

  Παράδειγμα
  Αν $\mathcal{E} = \{0, 1\}$ και L είναι όλα τα στοιχεία του $\mathcal{E}^*$ που τελειώνουν σε 0, τότε $\overline{L}$ είναι όλα τα στοιχεία του $\mathcal{E}^*$ που τελειώνουν σε 1, καθώς και η $\square$.

  Προφανώς, ισχύει ότι $\overline{\overline{L}} = L$. Το $\overline{\mathcal{E}^*}$ δεν περιέχει καμία λέξη.

Κανονικές εκφράσεις

Οι κανονικές εκφράσεις (regular expressions) ορίζονται αναδρομικά.
Έστω $\mathcal{E}$ ένα αλφάβητο. Τα επόμενα σύμβολα ονομάζονται **κανονικές εκφράσεις του** $\mathcal{E}$.

- Το κενό σύνολο $\emptyset$ είναι κανονική έκφραση του $\mathcal{E}$.
- Η κενή λέξη $\square$ είναι κανονική έκφραση του $\mathcal{E}$.
- Κάθε γράμμα a του $\mathcal{E}$ είναι κανονική έκφραση του $\mathcal{E}$.
- Αν A, B είναι κανονικές εκφράσεις του $\mathcal{E}$, τότε και τα $(A \cup B)$, (AB), $A^* = (\bigcup_{k=0}^{\infty} A^k)$ είναι κανονικές εκφράσεις του $\mathcal{E}$.
- Δεν υπάρχουν άλλες κανονικές εκφράσεις του $\mathcal{E}$.

Παραδείγματα

Στο αλφάβητο $\mathcal{E} = \{0, 1\}$ οι παρακάτω εκφράσεις είναι κανονικές.

- 1^*
- 01^*0^*
- $(0^*1 \cup 01^*)1^*$
- $(01)(01)^*10^* \cup (00)^*$

Κάθε κανονική έκφραση του $\mathcal{E}$ αναπαριστά ένα σύνολο λέξεων του $\mathcal{E}^*$. Συγκεκριμένα,

- Το κενό σύνολο $\emptyset$ αναπαριστά το κενό σύνολο!
- Η κενή λέξη $\square$ αναπαριστά το μονοσύνολο $\{\square\}$.
- Κάθε γράμμα a του $\mathcal{E}$ αναπαριστά το μονοσύνολο $\{a\}$.
- Η $(A \cup B)$ αναπαριστά την ένωση των συνόλων που αναπαριστούν τα A και B
- Η (AB) αναπαριστά την σύζευξη των συνόλων που αναπαριστούν τα A και B.
- Η A^* αναπαριστά την κλειστότητα του Kleene του συνόλου που αναπαριστά το A.

Παραδείγματα

- $1^* = \{\square, 1, 11, 111, \ldots\} = \{1^n : n \in \mathbb{N}\}$
- $01^*0^* = \{0\}\{1^n : n \in \mathbb{N}\}\{0^k : k \in \mathbb{N}\} = \{01^n0^k : k, n \in \mathbb{N}\}$
- $0(1^*0^*)^*$ = Το σύνολο των δυαδικών λέξεων που αρχίζουν με 0.
- $(0^*1^*)^*(101)(0^*1^*)^*(101)(0^*1^*)^*$ = Το σύνολο των δυαδικών λέξεων που περιέχουν δύο τουλάχιστον εμφανίσεις του 101.
- $(0^*1^*)^*(101 \cup 1100)$ = Το σύνολο των δυαδικών λέξεων που τελειώνουν είτε σε 101, είτε σε 1100.
- $(0^*1 \cup 01^*)1^* = \{0^n1 : n \in \mathbb{N}\} \cup \{01^k : k \in \mathbb{N}\}\{1^\lambda : \lambda \in \mathbb{N}\}$.
- $(01)(01)^*10^* \cup (00)^* = \{01(01)^n10^k : n, k \in \mathbb{N}\} \cup \{(00)^\lambda : \lambda \in \mathbb{N}\}$.

Μια γλώσσα L του $\mathcal{E}$ ονομάζεται **κανονική** (regular) αν ορίζεται από κανονική έκφραση του $\mathcal{E}$, δηλαδή υπάρχει κανονική έκφραση R του $\mathcal{E}$ που αναπαριστά την γλώσσα L.

Παραδείγματα Έστω $\mathcal{E} = \{0, 1\}$.

- Η γλώσσα L των δυαδικών λέξεων είναι κανονική διότι $L = (0 \cup 1)^*$
- Η γλώσσα L των δυαδικών λέξεων που αρχίζουν με 01 είναι κανονική διότι $L = (01)(0 \cup 1)^*$.
- Η γλώσσα L των δυαδικών λέξεων που περιέχουν άρτιο αριθμό 0 είναι κανονική διότι $L = 1^*(01^*0)^*$.
- Αποδεικνύεται ότι η γλώσσα $L = \{0^n 1^n : n \in \mathbb{N}\}$ δεν είναι κανονική.

Θεώρημα του Kleene (1956)

Μια γλώσσα είναι κανονική αν και μόνο αν αναγνωρίζεται από κάποιο D-αυτόματο.

Με άλλα λόγια, τα D-αυτόματα αναγνωρίζουν μόνο κανονικές γλώσσες και για κάθε κανονική γλώσσα υπάρχει ένα D-αυτόματο που την αναγνωρίζει.

Ένα χρήσιμο κριτήριο για το πότε μια γλώσσα είναι κανονική είναι το λεγόμενο pumping lemma:

Pumping lemma

Αν μια γλώσσα L είναι κανονική τότε υπάρχει ένας ακέραιος $p \geq 1$ τέτοιος ώστε κάθε λέξη a με μήκος $|a| \geq p$
α) γράφεται στην μορφή $a = uvw$ όπου $v \neq \square$
β) $|uv| \leq p$
γ) $uv^i w \in L$ για κάθε $i \in \mathbb{N}^*$.

Δηλαδή στις κανονικές γλώσσες, κάθε αρκετά μακριά λέξη a μπορεί να χωριστεί σε 3 μέρη, δηλαδή $a = uvw$, έτσι ώστε όλες οι λέξεις $uv^i w$, $i \in \mathbb{N}$ να περιέχονται επίσης στην γλώσσα: Το μεσαίο τμήμα όλων των αρκετά μακριών λέξεων μια κανονικής γλώσσας μπορεί να επαναληφθεί οσεσδήποτε φορές (φουσκώνοντας (pumping) έτσι την λέξη) ώστε να παράγει νέες λέξεις που επίσης περιέχονται στην γλώσσα.
Στην Ελληνική βιβλιογραφία το pumping lemma έχει μεταφραστεί ως λήμμα της άντλησης! Αν η γλώσσα είναι πεπερασμένη και έχει μέγιστο μήκος λέξεων ίσο με n, τότε το λήμμα ισχύει για $p = n + 1$.

Παράδειγμα Η γλώσσα $L = \{0^n 1^n : n \in \mathbb{N}\}$ δεν είναι κανονική.
Απόδειξη Έστω ότι η L είναι κανονική. Τότε από το pumping lemma υπάρχει ακέραιος p ώστε κάθε λέξη $a \in L$ με $|a| > p$ να γράφεται στην μορφή

$$a = uvw$$

όπου $|uv| \leq p$ και $uv^i w \in L$.
Θεωρούμε την λέξη $a = 0^p 1^p$. Το μήκος της a είναι μεγαλύτερο από το p. Από το pumping lemma υπάρχουν λέξεις u, v, w ώστε $a = uvw$ όπου $|uv| \leq p$. Επειδή $|uv| < p$ έπεται ότι $u = 0^k$, $v = 0^\lambda$, $\lambda > 0$, όπου $k + \lambda \leq p$, και $w = 0^{p-k-\lambda} 1^p$.
Από το pumping lemma πρέπει όλες οι λέξεις $uv^i w$ επίσης να ανήκουν στην L. Όμως,

$$uv^i w = 0^k 0^{i\lambda} 0^{p-k-\lambda} 1^p = 0^{p+(i-1)\lambda} 1^p$$

για $i = 2$ προκύπτει ότι η λέξη $0^{p+\lambda} 1^p \in L$, $\lambda > 0$, το οποίο είναι άτοπο.

Κανόνες αντικατάστασης

Σύστημα σχέσεων του Thue είναι ένα σύστημα κανόνων αντικαταστάσεων στο $\mathcal{E}^*$, δηλαδή η παραδοχή ότι κάποιες λέξεις είναι ισοδύναμες (και άρα η μια μπορεί να αντικαθιστά την άλλη όταν τις συναντάμε μέσα σε άλλες λέξεις).
Γράφουμε

$$w_i \sim w_i'$$

$i = 1, 2, \ldots, n$ γι᾽ αυτές τις συνθήκες.
Δύο λέξεις λέγονται **γειτονικές** αν η μια προκύπτει από την άλλη με μια μόνο αντικατάσταση. Αν η μια προκύπτει από την άλλη μέσω περισσότερων διαδοχικών αντικαταστάσεων, τότε οι λέξεις λέγονται **ισοδύναμες** και γράφουμε $w_1 \approx w_2$.

Παραδείγματα

1. Θεωρώντας το σύστημα των σχέσεων του Thue:

$$ab \sim ba, aba \sim b, bab \sim a$$

   τότε οι λέξεις $w_1 = aabbacca$, $w_2 = acca$ είναι ισοδύναμες.
   Πραγματικά,

$$\begin{aligned} w_1 &\approx a\underline{ab}bacca \\ &\approx ab\underline{aba}cca \\ &\approx \underline{ab}bcca \\ &\approx \underline{bab}cca \\ &\approx acca \approx w_2. \end{aligned}$$

2. Θεωρώντας το σύστημα των σχέσεων του Thue:

$$\omega \sim \mathrm{o},\ \alpha\iota \sim \varepsilon,\ \varepsilon\iota \sim \iota$$

   έχουμε την ισοδυναμία

$$\Pi\varepsilon\iota\rho\alpha\iota\underline{\omega\varsigma} \approx \Pi\varepsilon\iota\rho\underline{\alpha\iota}\mathrm{o}\varsigma \approx \Pi\underline{\varepsilon\iota}\rho\varepsilon\mathrm{o}\varsigma \approx \Pi\iota\rho\varepsilon\mathrm{o}\varsigma.$$

Το πρόβλημα των λέξεων

Έστω $\mathcal{E}$ ένα αλφάβητο και ένα σύστημα σχέσεων του Thue στο $\mathcal{E}^*$.
Το **πρόβλημα των λέξεων** είναι να κατασκευασθεί ένας αλγόριθμος απαντάει στο ερώτημα αν δύο λέξεις w_1 και w_2 του $\mathcal{E}^*$ είναι ισοδύναμες ή όχι.

Παράδειγμα Έστω $\mathcal{E} = \{a, b, c, d, e\}$ και το σύστημα σχέσεων του Thue

$$
\begin{aligned}
ac &\sim ca \\
ad &\sim da \\
bc &\sim cb \\
bd &\sim db \\
eca &\sim ce \\
edb &\sim de \\
cdca &\sim cdcae \\
caaa &\sim aaa \\
daaa &\sim aaa
\end{aligned}
$$

Να κατασκευασθεί αλγόριθμος που να απαντάει στο ερώτημα αν μια λέξη w είναι ισοδύναμη με την λέξη aaa.

Το 1947 ο Markov απέδειξε ότι το πρόβλημα των λέξεων είναι μη επιλύσιμο.

Θεώρημα Markov

Υπάρχουν συστήματα σχέσεων του Thue και λέξεις w για τις οποίες **δεν υπάρχει** αλγόριθμος που να απαντά σε όλες τις περιπτώσεις αν μια λέξη u είναι ισοδύναμη με την w.

Το προηγούμενο παράδειγμα (από τον Tseitin) είναι ένα από αυτά τα συστήματα Thue για το οποίο έχει αποδειχθεί ότι δεν υπάρχει αλγόριθμος που να απαντά αν μια λέξη w είναι ισοδύναμη με την aaa ή όχι.

Άσκηση 1

Να συγκριθούν οι λέξεις $w_1 = aabaca$ *και* $w_2 = abc$ *του* $\mathcal{E}^*$, *που έχει αλφάβητο* $\mathcal{E} = \{a, b, c\}$ *και σύστημα σχέσεων του Thue:*

$$ab \sim ba, ac \sim ca, aaa \sim \square.$$

Απάντηση.

Είναι ισοδύναμες. $\square$