Τυπικές γλώσσες

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
   $\mathcal{E}_4 = \{a, b, c, \ldots, z, (,), \vee, \wedge, \rightarrow, \neg\}$.

Τα στοιχεία ενός αλφάβητου ονομάζονται **γράμματα**. Κάθε πεπερασμένη διαδοχή 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 $(a \rightarrow b)$, $(a \vee b) \wedge (c \vee d)$, $a \vee (\neg a)$, κ.λπ.

Ορίζουμε επίσης μια λέξη χωρίς γράμματα, την **κενή λέξη** $\square$ (ή ϵ).

Ο αριθμός των γραμμάτων μιας λέξης m λέγεται **μήκος** ή **βαθμός** της λέξης και συμβολίζεται με $l(m)$.

Έτσι για την πρώτη λέξη καθενός από τα τέσσερα προηγούμενα παραδείγματα έχουμε $l(m) = 5, 4, 5$ και 3 αντίστοιχα.

Η λέξη που προκύπτει από την αντιστροφή της σειράς των γραμμάτων μιας λέξης m ονομάζεται **κατοπτρική εικόνα** της m και συμβολίζεται με $\widetilde{m}$.

Παραδείγματος χάριν για τις 4 πρώτες λέξεις καθενός από τα 4 προηγούμενα παραδείγματα έχουμε $\widetilde{m} =$ ΣΟΚΙΝ, ΤΑΟΒ, 01001 και $b \to a$ αντίστοιχα.

Μια λέξη m για την οποία $\widetilde{m} = m$ ονομάζεται **συμμετρική**.

Παραδείγματα ΑΝΝΑ, 1001001, ΝΟΜΙΜΟΝ, κ.λπ.

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

Παραδείγματα

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

Κάθε οικογένεια λέξεων του $\mathcal{E}^*$ που περιέχει τουλάχιστον μια φορά την $\square$, λέγεται **φράση**.

Παράδειγμα ΕΙΜΑΙ$\square$ΜΙΑ$\square$ΦΡΑΣΗ.

Ορίζουμε μια εσωτερική πράξη στο $\mathcal{E}^*$ ως εξής: Για κάθε $m_1, m_2 \in \mathcal{E}^*$ παίρνουμε την $m = m_1 m_2$ που προκύπτει από τα τα γράμματα της m_1 ακολουθούμενα από τα γράμματα της m_2. Η πράξη αυτή λέγεται **ζεύξη**.

Παραδείγματα

Αν $m_1 = 1011$ και $m_2 = 11$ τότε $m_1 m_2 = 101111$ και $m_2 m_1 = 111011$.

Αν $m_1 =$ ΑΝΤΙ και $m_2 =$ ΓΡΑΦΩ τότε $m_1 m_2 =$ ΑΝΤΙΓΡΑΦΩ.

Προφανώς ισχύει η προσεταιριστικότητα αλλά όχι η αντιμεταθετικότητα, όπως φαίνεται και από το πρώτα παράδειγμα. Επίσης, υπάρχει ουδέτερο στοιχείο (η λέξη $\square$, αφού $m\square = \square m = m$).

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

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. $\mathcal{E} = \{0, 1, 2, 3, \ldots, 9\}$ και L όλοι οι αριθμοί x με $100 \leq x \leq 355$.

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

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. Συμβολίζεται με $\emptyset$ και λέγεται **κενή γλώσσα**.

Αντικατάσταση

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

$$m_i \sim m_i'$$

$i = 1, 2, \ldots, n$ 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 αντικαταστάσε
τότε οι λέξεις λέγονται **ισοδύναμες** και γράφουμε $m_1 \approx m_2$.

Παραδείγματα

1. Θεωρώντας το σύστημα των σχέσεων του Thue:

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

   τότε οι λέξεις $m_1 = aabbacc$, $m_2 = acca$ είναι ισοδύναμες. Πραγματικά,

   $$\begin{aligned} m_1 &\approx a\underline{ab}bacca \\ &\approx ab\underline{aba}cca \\ &\approx \underline{ab}bcca \\ &\approx \underline{bab}cca \\ &\approx acca \approx m_2 \end{aligned}$$

2. Θεωρώντας το σύστημα των σχέσεων του Thue:

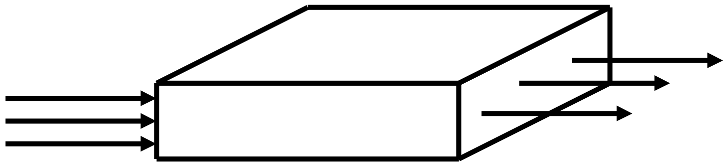
   $$\text{ω} \sim \text{ο},\ \text{αι} \sim \text{ε},\ \text{ει} \sim \text{ι}$$

   έχουμε την ισοδυναμία

   Πειραι<u>ω</u>ς ≈ Πειρ<u>αι</u>ος ≈ Π<u>ει</u>ρεος ≈ Πιρεος

Αυτόματα

Ένα κουτί με διαύλους εισόδου και εξόδου:



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.

Στη συνέχεια θα ασχοληθούμε με τα D-αυτόματα:

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).

$T \subset S$: το σύνολο των τελικών καταστάσεων.

$s_0 \in S$: η αρχική κατάσταση.

$f : S \times E \to S$ (η συνάρτηση μετάβασης).

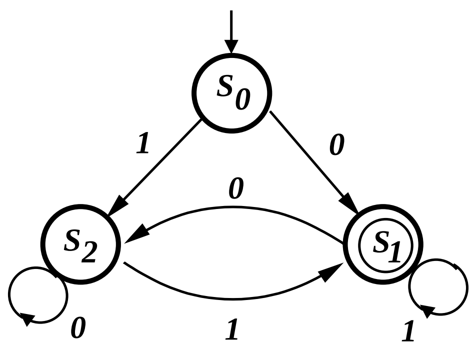
ονομάζεται **(πεπερασμένο) D-αυτόματο (deterministic finite state automaton)**. Η συνάρτηση μετάβασης f, εκφράζει τον ‘‘εσωτερικό μηχανισμό’’ του D-αυτόματου, και **σε κάθε ζεύγος του $S \times \mathcal{E}$** αντιστοιχεί **ακριβώς ένα στοιχείο του S**.

Σε κάθε D-αυτόματο αντιστοιχεί ένα προσανατολισμένο p-γράφημα με κορυφές τις καταστάσεις του αυτόματου και τόξα τα διατεταγμένα ζεύγη που προκύπτουν από τη συνάρτηση μετάβασης. Στην κορυφή της αρχικής κατάστασης s_0 υπάρχει ένα τόξο εισόδου, ενώ για τις κορυφές των τελικών καταστάσεων χρησιμοποιούνται διπλοί κύκλοι, αντί για τον ένα που χρησιμοποιείται για τις υπόλοιπες κορυφές.

Παραδείγματα

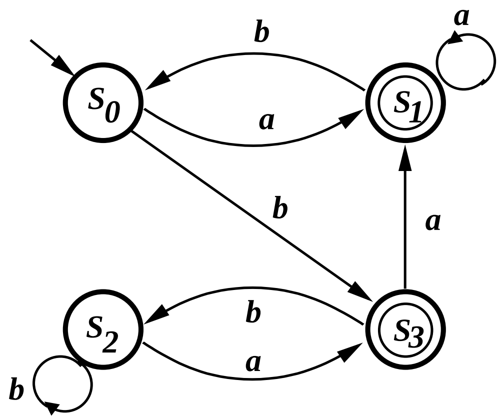
1. $S = \{s_0, s_1, s_2\}$, $\mathcal{E} = \{0, 1\}$, $T = \{s_1\}$ και $f : S \times \mathcal{E} \to S$ με $f(s_0, 0) = s_1$, $f(s_0, 1) = s_2$, $f(s_1, 0) = s_2$, $f(s_1, 1) = s_1$, $f(s_2, 0) = s_2$ και $f(s_2, 1) = s_1$.

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

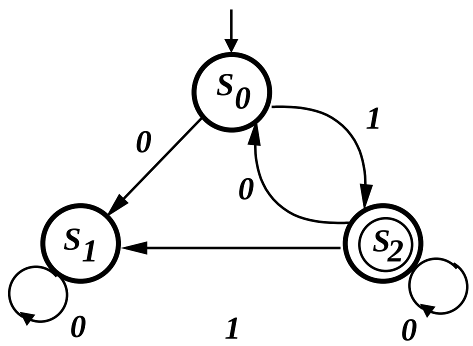


2. $S = \{s_0, s_1, s_2, s_3\}$, $\mathcal{E} = \{a, b\}$, $T = \{s_1, s_2\}$ και $f : S \times \mathcal{E} \to S$ με $f(s_0, a) = s_1$, $f(s_0, b) = s_3$, $f(s_1, a) = s_1$, $f(s_1, b) = s_0$, $f(s_2, a) = s_3$, $f(s_2, b) = s_2$, $f(s_3, a) = s_1$, $f(s_3, b) = s_2$.

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



3. 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 $f^* : S \times \mathcal{E}^* \rightarrow S$ ως εξής:

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

$$f^*(s_i, m_1 m_2) = f(f^*(s_i, m_1), m_2),$$

$\forall s_i \in S, m_1 \in \mathcal{E}^*, m_2 \in E.$

Μια λέξη μ **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 αναγνωρίζει τη λέξη $m = 1001101$, αφού

$$
\begin{aligned}
f^*(s_0, 1) &= f(f^*(s_0, \square), 1) = f(s_0, 1) = s_2 \\
f^*(s_0, 10) &= f(f^*(s_0, 1), 0) = f(s_2, 0) = s_2 \\
f^*(s_0, 100) &= f(f^*(s_0, 10), 0) = f(s_2, 0) = s_2 \\
f^*(s_0, 1001) &= f(f^*(s_0, 100), 1) = f(s_2, 1) = s_1 \\
f^*(s_0, 10011) &= f(f^*(s_0, 1001), 1) = f(s_1, 1) = s_1 \\
f^*(s_0, 100110) &= f(f^*(s_0, 10011), 0) = f(s_1, 0) = s_2 \\
f^*(s_0, m) &= f(f^*(s_0, 100110), 1) = f(s_2, 1) = s_1 \in T
\end{aligned}
$$

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

$$s_0 \to^1 s_2 \to^0 s_2 \to^0 s_2 \to^1 s_1 \to^1 s_1 \to^0 s_2 \to^1 s_1 \in T$$

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

$$s_0 \to^1 s_2 \to^1 s_1 \to^1 s_1 \to^1 s_1 \to^0 s_2 \to^0 s_2 \to^1 s_1 \to^0 s_2 \notin T$$

3. Το D-αυτόματατο του παραδείγματος 2 αναγνωρίζει τη λέξη $m = abbbbabaa$ διότι

$$s_0 \to^a s_1 \to^b s_0 \to^b s_3 \to^b s_2 \to^b s_2 \to^a s_3 \to^b s_2 \to^a s_3 \to^a s_1 \in T$$

Ασκήσεις

1. Να συγκριθούν οι λέξεις $m_1 = aabaca$ και $m_2 = abc$ του $\mathcal{E}^*$, που έχει αλφάβητο $\mathcal{E} = \{a, b, c\}$ και σύστημα
$$ab \sim ba, ac \sim ca, aaa \sim \square.$$

2. Να ορισθεί το σύνολο των λέξεων, μήκους μικρότερου ή ίσου του 4, που αρχίζουν από a όταν $\mathcal{E} = \{a, b, c\}$.

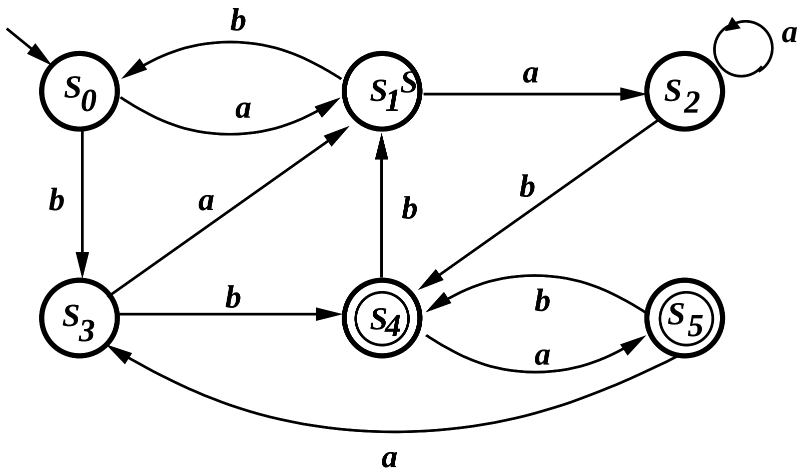
3. Να ορισθεί το σύνολο των λέξεων, μήκους μικρότερου ή ίσου του 4, στις οποίες το a έχει περιττό αριθμό εμφανίσεων όταν $\mathcal{E} = \{a, b\}$.

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

5. Να ευρεθεί ένα D-αυτόματο με $\mathcal{E} = \{a, b\}$ που να αναγνωρίζει τις λέξεις στις οποίες το a έχει περιττό αριθμό εμφανίσεων.

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

7. Να εξετασθεί αν το D-αυτόματο



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

8. (i) Να δοθεί το γράφημα που περιγράφει το D-αυτόματο $(S, \mathcal{E}, T, s_0, f)$ όπου $S = \{s_0, s_1, s_2, s_3, s_4\}$, $\mathcal{E} = \{a, b\}$, $T = \{s_2, s_4\}$ και $f : S \times \mathcal{E} \to S$ με $f(s_0, a) = s_4$, $f(s_0, b) = s_1$, $f(s_1, a) = s_2$, $f(s_1, b) = s_0$, $f(s_2, a) = s_4$, $f(s_2, b) = s_2$, $f(s_3, a) = s_1$, $f(s_3, b) = s_2$, $f(s_4, a) = s_4$, $f(s_4, b) = s_3$.

(ii) Να εξετασθεί αν το αυτόματο αυτό αναγνωρίζει τις λέξεις:
$w_1 = abababab$, $w_2 = baaabbabaaab$ και $w_3 = baabbaab$.

(iii) Έστω u μια λέξη που αποτελείται από m σε πλήθος a και v μια λέξη που αποτελείται απο m σε πλήθος b $(m, n \in \mathbb{N}^*)$. Να δειχθεί ότι το παραπάνω D-αυτόματο αναγνωρίζει τη λέξη $w = vu$. Κάτω από ποιες προϋποθέσεις αναγνωρίζει τη λέξη $w' = uv$;