ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ
Β΄ ΕΞΑΜΗΝΟ ΣΠΟΥΔΩΝ

ΔΙΑΚΡΙΤΑ ΜΑΘΗΜΑΤΙΚΑ

Αυτόματα Πεπερασμένων Καταστάσεων

Κανονικές Εκφράσεις και Κανονικές Γλώσσες

Διδάσκων: ΣΤΑΥΡΟΣ ΑΔΑΜ

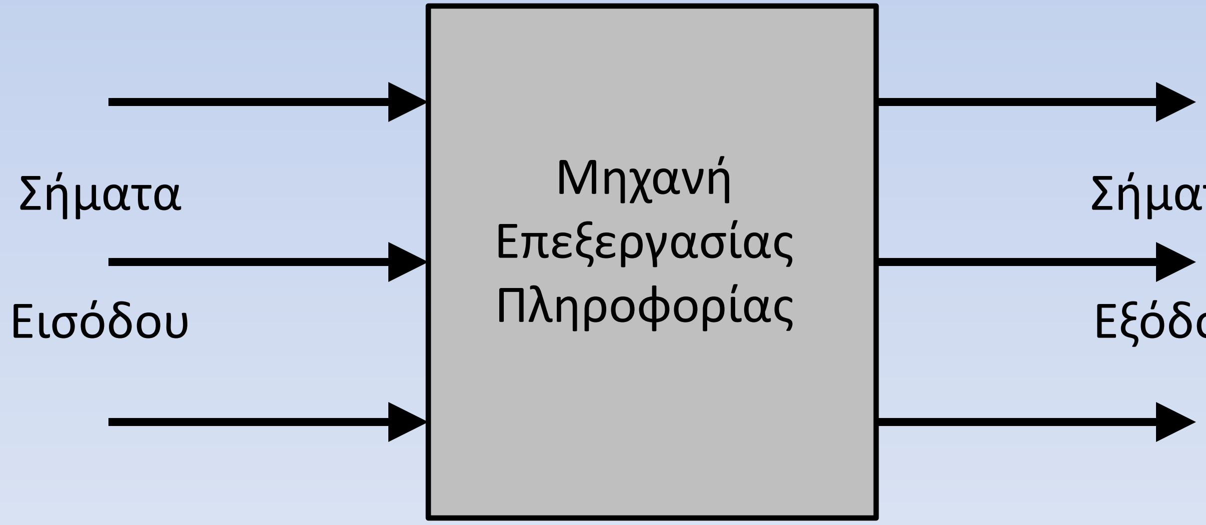
ΕΑΡΙΝΟ ΕΞΑΜΗΝΟ 2023-2024

Ορισμοί και Ιδιότητες

Ορισμοί

Μια μηχανή επεξεργασίας της πληροφορίας είναι ένας οποιοσδήποτε μηχανισμός ο οποίος είναι δυνατόν να θεωρηθεί ότι δέχεται ένα σύνολο σημάτων εισόδου και παράγει ένα αντίστοιχο σύνολο σημάτων εξόδου.

Σχηματικά:



Ορισμοί και Ιδιότητες

Παραδείγματα:

Μια λάμπα γραφείου είναι δυνατόν να θεωρηθεί ως ένα σύστημα επεξεργασίας πληροφοριών με σήματα εισόδου την κατάσταση του διακόπτη ON ή OFF και σήματα εξόδου την κατάσταση της λάμπας ΦΩΣ ή ΣΚΟΤΑΔΙ. Το ενδιαφέρον μας εδώ δεν είναι στην ηλεκτρική σχεδίαση ή τη δομή της συσκευής αλλά στο πώς η συσκευή λειτουργεί όταν δέχεται ως σήματα εισόδου την κατάσταση του διακόπτη και τι παράγει στην έξοδο δηλαδή ποιές είναι οι πληροφορίες ή τα σήματα που παράγει στην έξοδο.

Ένα συνδυαστικό κύκλωμα, όπως, μια πύλη AND είναι ένα σύστημα επεξεργασίας των σημάτων εισόδου με βάση τα οποία παράγει ένα σήμα εξόδου.

Άλλα παραδείγματα;

Ορισμοί και Ιδιότητες

Γενικά τα σήματα εισόδου μεταβάλλονται με το χρόνο, και κατά συνέπεια και τα σήματα εξόδου που παράγει η μηχανή μεταβάλλονται με το χρόνο. Σε κάποιες μηχανές τα σήματα εξόδου εξαρτώνται από την άμεσα προηγούμενη χρονικά κατάσταση των σημάτων εξόδου ενώ σε κάποιες άλλες για την παραγωγή των σημάτων εξόδου λαμβάνεται υπόψη η κατάσταση των σημάτων εισόδου αρκετές χρονικές στιγμές πριν.

Για παράδειγμα ο μηχανισμός κλειδώματος ενός χρηματοκιβωτίου για να δώσει στην έξοδο το σήμα ΑΝΟΙΚΤΟ θα πρέπει να λάβει υπόψη του μια χρονική ακολουθία από σήματα εισόδου τα οποία θα αντιστοιχούν στο σωστό συνδυασμό αριθμών που απαιτούνται για να ξεκλειδώσει ο μηχανισμός.

Ορισμοί και Ιδιότητες

Αντίστοιχη είναι και η λειτουργία ακολουθιακών κυκλωμάτων όπως για παράδειγμα ένα SR (ή RS) flip-flop η λειτουργία του οποίου εξαρτάται από τις αρχικές τιμές στις εισόδους αλλά και από την ανατροφοδότηση των εξόδων στις κατάλληλες εισόδους.

Για την ψηφιακή σχεδίαση αυτό που μας ενδιαφέρει δεν είναι η ηλεκτρονική των κυκλωμάτων αυτών αλλά η λειτουργία όπως περιγράφεται από τον πίνακα αληθείας (truth table) και αποδίδεται σχηματικά για τη ροή της πληροφορίας. Άρα ένα ακολουθιακό κύκλωμα είναι ένας μηχανισμός επεξεργασίας της πληροφορίας η λειτουργία του οποίου εξαρτάται από το παρελθόν των σημάτων εισόδου, δηλαδή, από την ακολουθία των εισόδων.

Ορισμοί και Ιδιότητες

Κατά συνέπεια, σε αντίθεση με την περίπτωση της λάμπας ή μιας πύλης AND ο μηχανισμός κλειδώματος του χρηματοκιβωτίου ή αντίστοιχα ένα flip-flop θεωρούνται ως μηχανές (ή μηχανισμοί) επεξεργασίας πληροφοριών που διαθέτουν μνήμη γιατί απαιτείται να «θυμούνται» τι έχει συμβεί στο παρελθόν.

Ένα ακόμη τυπικό παράδειγμα μηχανής με μνήμη αφορά τις μηχανές αυτόματης πώλησης οι οποίες δέχονται κέρματα και όταν συμπληρωθεί το ποσό που αξίζει ένα προϊόν τότε η μηχανή παραδίδει το προϊόν (καφέ, αναψυκτικό, νερό κ.λπ). Σαν σήμα εισόδου εδώ είναι δυνατόν να θεωρηθεί το ποσό του κάθε κέρματος που εισάγεται στη μηχανή και σαν σήμα εξόδου το OK για την παράδοση του προϊόντος.

Ορισμοί και Ιδιότητες

Η μνήμη που διαθέτει η εν λόγω μηχανή είναι προφανές ότι δε λαμβάνει υπόψη της ούτε το πλήθος των κερμάτων ούτε τη σειρά με την οποία της εδόθηκαν τα κέρματα. Εκείνο που «θυμάται» η μηχανή είναι ανά πάσα στιγμή το συνολικό ποσό (αθροιστικά) που της έχει δοθεί και όταν το ποσό αυτό είναι εκείνο που απαιτείται τότε παράγεται το σήμα εξόδου, δηλαδή η παράδοση στον πελάτη του προϊόντος που έχει επιλέξει.

Έτσι, οι μηχανισμοί αυτοί επεξεργασίας της πληροφορίας που διαθέτουν μνήμη απομνημονεύουν καταστάσεις και όχι αριθμούς ή άλλες πληροφορίες.

Αν το πλήθος των καταστάσεων είναι πεπερασμένο τότε η μηχανή ονομάζεται μηχανή πεπαρασμένων καταστάσεων αλλιώς ονομάζεται μηχανή απείρων καταστάσεων.

Ορισμοί και Ιδιότητες

Ορισμός

Ένα αυτόματο, ή μηχανή, πεπερασμένων καταστάσεων $\mathbf{M}$ (Finite State Automaton - Machine) είναι μια εξάδα στοιχείων $\mathbf{M=(S, s_0, I, O, f, g)}$ **:**

- $\mathbf{S=\{s_0, s_1, s_2, \cdots, s_k\}}$ **:** ένα πεπερασμένο σύνολο στοιχείων που αποτελούν τις **καταστάσεις του αυτομάτου**.
- $\mathbf{s_0 \in S}$ **: αρχική κατάσταση**.
- $\mathbf{I=\{i_1, i_2, \cdots, i_n\}}$ **:** ένα πεπερασμένο σύνολο στοιχείων που αποτελούν τα **σήματα εισόδου**.
- $\mathbf{O=\{o_1, o_2, \cdots, o_m\}}$ **:** ένα πεπερασμένο σύνολο στοιχείων που αποτελούν τα **σήματα εξόδου**.
- $\mathbf{f : S \times I \to S}$ **:** μια συνάρτηση που αποτελεί τη **συνάρτηση μετάβασης** (στην επόμενη κατάσταση).
- $\mathbf{g : S \to O}$ **:** μια συνάρτηση που αποτελεί τη **συνάρτηση εξόδου**.

Ορισμοί και Ιδιότητες

Περιγραφή λειτουργίας του αυτομάτου

1. Αρχικά το αυτόματο βρίσκεται στην κατάσταση $\mathbf{s_0}$ (αρχική κατάσταση).
2. Με τη λήψη ενός σήματος εισόδου $\mathbf{i}$ το αυτόματο αλλάζει κατάσταση μεταβαίνοντας στην κατάσταση $\mathbf{s_j = f(s_0, i)}$ σύμφωνα με τη συνάρτηση μετάβασης
3. Στη νέα αυτή κατάσταση το αυτόματο παράγει ένα σήμα εξόδου σύμφωνα με τη συνάρτηση εξόδου και επιστρέφει στο βήμα 2.
4. Το αυτόματο σταματά να λειτουργεί όταν εξαντληθούν τα σήματα εισόδου ή όταν για κάποια κατάσταση και κάποια συγκεκριμένη είσοδο δεν έχει προβλεφθεί μετάβαση σε άλλη κατάσταση. Στην περίπτωση αυτή το αυτόματο «μπλοκάρει».

Ορισμοί και Ιδιότητες

Παράδειγμα:

Έστω μια μηχανή αυτόματης πώλησης σνακ η οποία δέχεται κέρματα των 10, 20 και 50 λεπτών, δεν επιστρέφει ρέστα και δίνει ένα σνακ κάθε φορά που συμπληρώνεται το ποσό των 80 λεπτών.

Η μηχανή επεξεργασίας της πληροφορίας που περιγράφει τη λειτουργία του αυτόματου πωλητή είναι ένα αυτόματο πεπερασμένων καταστασεων που αποτελείται από τα ακόλουθα στοιχεία :

- $\mathbf{S=\{s_0, s_1, s_2, \cdots, s_8\}}$ **:** όπου η κατάσταση $\mathbf{s_j, 1 \leq j \leq 8}$ έχει την έννοια ότι $\mathbf{j*10}$ λεπτά έχουν εισαχθεί στη μηχανή.
- $\mathbf{s_0}$ **:** είναι η αρχική κατάσταση.
- $\mathbf{I=\{i_1,\ i_2, i_3\}}$ **:** είναι τα σήματα εισόδου που αντιστοιχούν στην εισαγωγή 10, 20 και 50 λεπτών αντίστοιχα

Ορισμοί και Ιδιότητες

- $\mathbf{O=\{1, 0\}:}$ είναι τα σήματα εξόδου που αντιστοιχούν στη συμπλήρωση του ποσού που κοστίζει ένα σνακ και στη μη συμπλήρωση αντίστοιχα.
- $\mathbf{f:}$ η συνάρτηση μετάβασης περιγράφεται από ένα πίνακα διπλής εισόδου που δίνεται στη συνέχεια.
- $\mathbf{g:}$ η συνάρτηση εξόδου δίνει σε κάθε κατάσταση το σήμα $\mathbf{0}$ αν στην κατάσταση δεν έχει συμπληρωθεί το ποσό που κοστίζει ένα σνακ και $\mathbf{1}$ αν το ποσό έχει συμπληρωθεί. Έτσι, για τη συνάρτηση $\mathbf{g}$ έχουμε $\mathbf{g(s_j) = 0}$ για $\mathbf{0 \leq j \leq 7}$ και $\mathbf{g(s_8) = 1}$.

Ορισμοί και Ιδιότητες

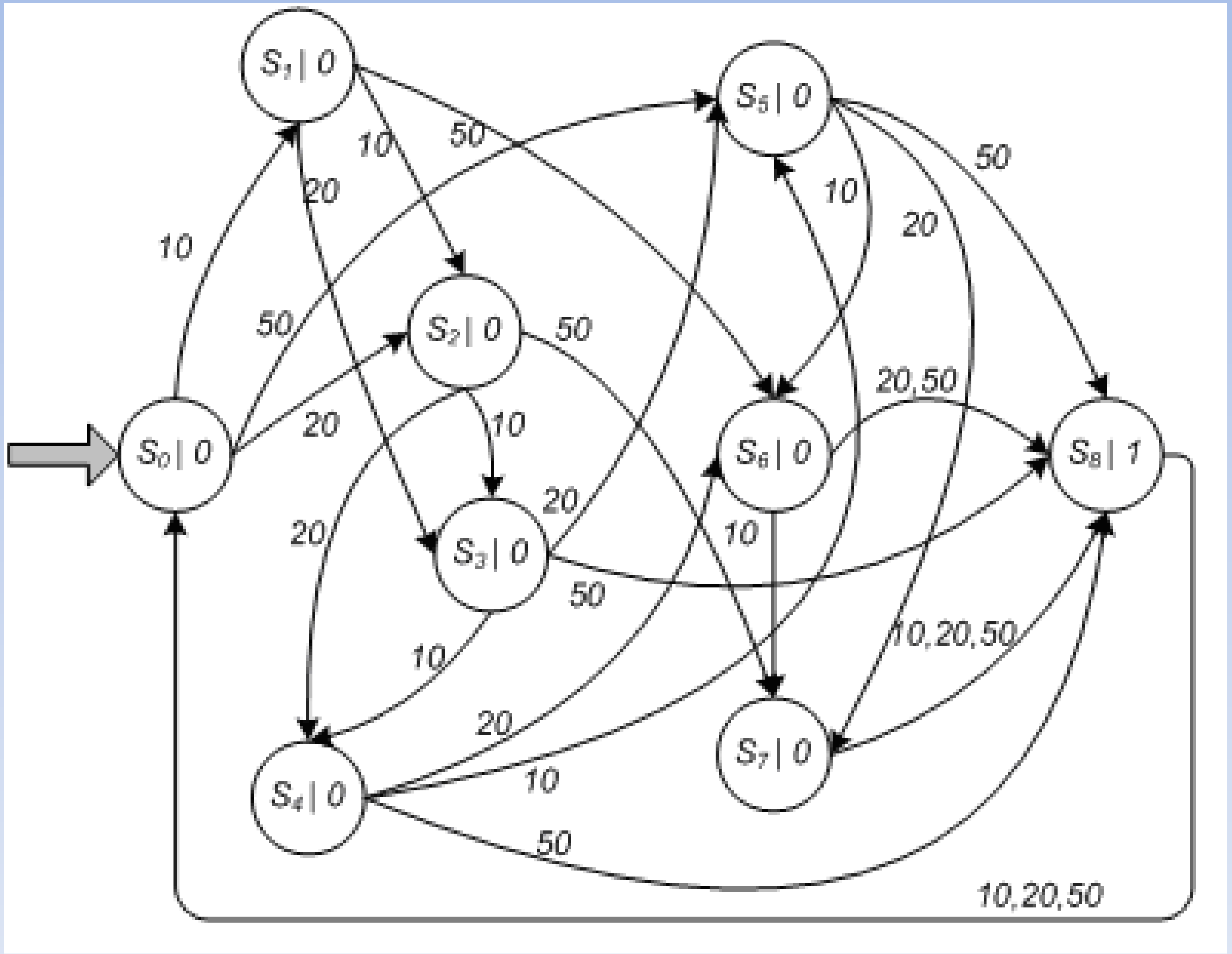
Πίνακας διπλής εισόδου για τις συναρτήσεις **f** και **g**

f	i_1	i_2	i_3
s_0	s_1	s_2	s_5
s_1	s_2	s_3	s_6
s_2	s_3	s_4	s_7
s_3	s_4	s_5	s_8
s_4	s_5	s_6	s_8
s_5	s_6	s_7	s_8
s_6	s_7	s_8	s_8
s_7	s_8	s_8	s_8
s_8	s_0		

g	O
s_0	0
s_1	0
s_2	0
s_3	0
s_4	0
s_5	0
s_6	0
s_7	0
s_8	1

Ορισμοί και Ιδιότητες

Διάγραμμα μετάβασης της συνάρτησης **f**



Ορισμοί και Ιδιότητες

Τα αυτόματα πεπερασμένων καταστάσεων είναι δυνατό να χρησιμοποιηθούν για τη μοντελοποίηση-αναπαράσταση φυσικών ή τεχνολογικών συστημάτων. Εκτός από το προηγούμενο παράδειγμα ένα άλλο παράδειγμα χρήσης είναι αυτό που προτείνεται στο βιβλίο του C. L. Liu, Στοιχεία Διακριτών Μαθηματικών, σελ. 288, για την αναπαράσταση ενός κυκλώματος που συγκρίνει δύο ακολουθίες από δυαδικά ψηφία.

Η αναπαράσταση ενός φυσικού συστήματος από ένα αυτόματο πεπερασμένων καταστάσεων δεν είναι μοναδική αλλά είναι δυνατόν να σχεδιασθούν διάφορα αυτόματα τα οποία αποδίδουν επακριβώς τη συμπεριφορά του συστήματος όταν δέχονται τα ίδια σήματα εισόδου. Για τα αυτόματα αυτά δίνεται ο ακόλουθος ορισμός.

Ορισμοί και Ιδιότητες

Ορισμός

Δύο αυτόματα πεπερασμένων καταστάσεων ονομάζονται ισοδύναμα όταν παράγουν την ίδια ακολουθία σημάτων εξόδου κάθε φορά που ξεκινούν από την ίδια αρχική κατάσταση και δέχονται την ίδια ακολουθία σημάτων εισόδου.

Στα αυτόματα πεπερασμένων καταστάσεων, όπως ορίστηκαν, η συνάρτηση μετάβασης $\mathbf{f}: \mathbf{S} \times \mathbf{I} \rightarrow \mathbf{S}$ για κάθε ζεύγος κατάστασης και σήματος εισόδου ορίζει μία και μόνο μία κατάσταση. Τα αυτόματα αυτά ονομάζονται **αιτιοκρατικά (deterministic) αυτόματα**, σε αντίθεση με τα ονομαζόμενα μη-αιτιοκρατικά **αυτόματα (non deterministic automata)**.

Ορισμοί και Ιδιότητες

Ορισμός

Ένα **μη-ατιοκρατικό (non deterministic) αυτόματο πεπερασμένων καταστάσεων** ορίζεται με τον ίδιο τρόπο όπως ένα αιτιοκρατικό αυτόματο με τη διαφορά ότι η συνάρτηση μετάβασης $\mathbf{f}$ ορίζεται ως εξής: $\mathbf{f: S \times I \to \mathcal{P}(S)}$, όπου $\mathcal{P}(S)$ συμβολίζει το δυναμοσύνολο του S (power set of S).

Αυτό σημαίνει πως όταν το αυτόματο βρίσκεται σε κάποια κατάσταση $\mathbf{s}$ και δεχθεί ένα σήμα εισόδου $\mathbf{i}$ τότε η νέα κατάσταση στην οποία θα βρεθεί ενδέχεται να μην υπάρχει ή ακόμη και να μην ορίζεται μονοσήμαντα, δηλαδή, ενδέχεται να υπάρχουν περισσότερες από μία νέες καταστάσεις δεδομένου ότι $\mathbf{f(s,i)} \in \mathcal{P}(S)$ και άρα $\mathbf{f(s,i)}$ είναι ένα σύνολο υποσύνολο του S το οποίο είναι ίσως και το κενό.

Ορισμοί και Ιδιότητες

Σημείωση

Τα μη αιτιοκρατικά πεπερασμένα αυτόματα δεν έχουν πρακτικές εφαρμογές αλλά είναι ισχυρές θεωρητικές μηχανές για εύκολη μοντελοποίηση συστημάτων.

Επί πλέον στη Θεωρία Υπολογισμού (Theory of Computation) αποδεικνύεται ότι για κάθε μη αιτιοκρατικό πεπερασμένο αυτόματο υπάρχει ένα ισοδύναμο αιτιοκρατικό. Για τη μετατροπή ενός μη αιτιοκρατικού σε ένα αιτιοκρατικό πεπερασμένο αυτόματο υπάρχει συγκεκριμένος αλγόριθμος μετατροπής.

Στην πράξη και στα πλαίσια αυτού του μαθήματος θα ασχοληθούμε αποκλειστικά με τα αιτιοκρατικά αυτόματα, δηλαδή τα αυτόματα στα οποία για κάθε κατάσταση και για κάθε σήμα εισόδου υπάρχει μια και μόνο μία κατάσταση που ορίζεται από τη συνάρτηση μετάβασης.

Αυτόματα Πεπερασμένων Καταστάσεων και Τυπικές Γλώσσες

Ορισμοί από τις Τυπικές Γλώσσες

- Έστω $\boldsymbol{\Sigma}$ είναι ένα μη κενό πεπερασμένο σύνολο, το οποίο ονομάζεται αλφάβητο και τα στοιχεία του σύμβολα. Με τον όρο συμβολοσειρά ή λέξη ονομάζουμε μια πεπερασμένη ακολουθία στοιχείων του $\boldsymbol{\Sigma}$.

  Για παράδειγμα η ακολουθία $\boldsymbol{\sigma_1 \sigma_2 \sigma_3 \cdots \sigma_n}$ όπου $\boldsymbol{\sigma_j \in \Sigma}$ για $\mathbf{1 \leq j \leq n}$ είναι μια συμβολοσειρά επί του $\boldsymbol{\Sigma}$. Στην ακολουθία αυτή όλα τα σύμβολα δεν είναι υποχρεωτικά διαφορετικά και κάποια μπορούν να επαναλαμβάνονται.
- Το σύνολο όλων των συμβολοσειρών επί του $\boldsymbol{\Sigma}$ συμβολίζεται με $\boldsymbol{\Sigma^*}$.
- Μια συμβολοσειρά επί του $\boldsymbol{\Sigma}$ μπορεί να ονομαστεί, όπως για παράδειγμα, $\mathbf{w} = \boldsymbol{\sigma_1 \sigma_2 \sigma_3 \cdots \sigma_n}$
- Το πλήθος των συμβόλων μιας συμβολοσειράς ονομάζεται μήκος της συμβολοσειράς και συμβολίζεται με $|\mathbf{w}| = \mathbf{n}$

Αυτόματα Πεπερασμένων Καταστάσεων και Τυπικές Γλώσσες

Ορισμοί από τις Τυπικές Γλώσσες

- Ορίζουμε ως κενή ή μηδενική συμβολοσειρά (null string) τη μοναδική συμβολοσειρά χωρίς στοιχεία και άρα μήκους $\mathbf{0}$ την οποία συμβολίζουμε με **e** ή **ε** ή **λ**.
- Αν $\mathbf{w= a_1\ a_2\ a_3 \cdots a_n}$ και $\mathbf{v= b_1\ b_2\ b_3 \cdots b_m}$ είναι δύο συμβολοσειρές του $\mathbf{\Sigma^*}$, τότε ονομάζουμε γινόμενο ή παράθεση του $\mathbf{w}$ και του $\mathbf{v}$ και συμβολίζουμε με $\mathbf{w \circ v}$ τη συμβολοσειρά που προκύπτει από τη συνένωση της $\mathbf{w}$ με τη $\mathbf{v}$, δηλαδή, $\mathbf{w \circ v = a_1\ a_2\ a_3 \cdots a_n\ b_1\ b_2\ b_3 \cdots b_m}$.
- Η πράξη της παράθεσης συμβολοσειρών δεν είναι αντιμεταθετική πράξη δηλαδή, $\mathbf{w \circ v \neq v \circ w}$.
- Η πράξη της παράθεσης συμβολοσειρών έχει την προσεταιριστική ιδιότητα , δηλαδή, $\mathbf{w \circ (v \circ u) = (w \circ v) \circ u}$.

Αυτόματα Πεπερασμένων Καταστάσεων και Τυπικές Γλώσσες

Ορισμοί από τις Τυπικές Γλώσσες

- Η κενή συμβολοσειρά (null string) είναι το ουδέτερο στοιχείο της πράξης της παράθεσης συμβολοσειρών, δηλαδή, $\mathbf{w \circ e = e \circ w = w}$.

  → Στην πράξη το σύμβολο ∘ παραλείπεται

- Αν $\mathbf{u}$ και $\mathbf{w}$ είναι δύο συμβολοσειρές του $\mathbf{\Sigma^*}$, τότε η $\mathbf{u}$ ονομάζεται υπο-συμβολοσειρά της $\mathbf{w}$ αν η $\mathbf{u}$ είναι τμήμα της $\mathbf{w}$ δηλαδή αν υπάρχουν συμβολοσειρές $\mathbf{s}$ και $\mathbf{t}$, ενδεχόμενα κενές, τέτοιες ώστε το $\mathbf{w}$ να γράφεται με τη μορφή $\mathbf{w = s\,u\,t}$ $\mathbf{(=s \circ u \circ t)}$.

- Για κάθε φυσικό αριθμό $\mathbf{n \in \mathbb{N}}$ ορίζουμε ως $\mathbf{\Sigma^n}$ το σύνολο των συμβολοσειρών μήκους $\mathbf{n}$ με στοιχεία του $\mathbf{\Sigma}$, δηλαδή, το σύνολο $\mathbf{\Sigma^n = \{w \in \Sigma^* : |w| = n\}}$.

  Έτσι ορίζονται τα σύνολα, $\mathbf{\Sigma^0 = \{e\}}$, $\mathbf{\Sigma^1 = \{w \in \Sigma^* : w = \sigma \in \Sigma\}}$.

- Να σημειωθεί ότι $\mathbf{\Sigma^* = \cup\, \Sigma^n}$, για όλα τα $\mathbf{n = 0, \cdots, \infty}$

Αυτόματα Πεπερασμένων Καταστάσεων και Τυπικές Γλώσσες

Ορισμοί από τις Τυπικές Γλώσσες

- Μια γλώσσα $\mathbf{L}$ επί του $\mathbf{\Sigma}$ είναι ένα υποσύνολο, $\mathbf{L} \subseteq \mathbf{\Sigma^*}$

  Με άλλα λόγια, κάθε σύνολο συμβολοσειρών του $\mathbf{\Sigma^*}$ είναι μια γλώσσα επί του $\mathbf{\Sigma}$.
- Έστω, $\mathbf{L} \subseteq \mathbf{\Sigma^*}$ μια γλώσσα επί του $\mathbf{\Sigma}$. Τότε, ορίζουμε τις ακόλουθες γλώσσες:
- $\mathbf{L^0 = \{e\}}$
- $\mathbf{L^1 = L}$
- $\mathbf{L^n = LL\cdots L}$ ($\mathbf{n}$ φορές) $=\{\mathbf{w_1\, w_2\, w_3 \cdots w_n : w_i \in L}\}$ για οποιοδήποτε φυσικό αριθμό $\mathbf{n \in \mathbb{N}}$
- $\mathbf{L^+ = \cup\, L^n,\ n=1,\cdots,\infty}$
- $\mathbf{L^* = L^0 \cup L^+}$

Αυτόματα Πεπερασμένων Καταστάσεων και Τυπικές Γλώσσες

Ορισμοί από τις Τυπικές Γλώσσες

- Αν $\mathbf{u}, \mathbf{v}$ είναι συμβολοσειρές της γλώσσας $\mathbf{L}$ τότε με $\mathbf{w} = \mathbf{u} + \mathbf{v}$ ή $\mathbf{w} = \mathbf{u} \mid \mathbf{v}$ συμβολίζεται η συμβολοσειρά $\mathbf{w}$ για την οποία ισχύει είτε $\mathbf{w} = \mathbf{u}$ είτε $\mathbf{w} = \mathbf{v}$. Η πράξη αυτή ονομάζεται διάζευξη ή εναλλακτική επιλογή.

  $\rightarrow$ Αρκετοί συγγραφείς την πράξη αυτή της διάζευξης τη συμβολίζουν με $\mathbf{w} = \mathbf{u} \cup \mathbf{v}$.
- Αν $\mathbf{w}$ είναι μια συμβολοσειρά της γλώσσας $\mathbf{L}$ τότε $\mathbf{w}^*$ (Kleene star) είναι η συμβολοσειρά της γλώσσας $\mathbf{L}^*$ που αντιστοιχεί στη συμβολοσειρά που παράγεται με μηδέν ή οποιονδήποτε αριθμό επαναλήψεων της $\mathbf{w}$.
- Όμοια μπορούμε να ορίσουμε τη συμβολοσειρά $\mathbf{w}^+$ που αντιστοιχεί στη συμβολοσειρά που παράγεται με ένα ή οποιονδήποτε αριθμό επαναλήψεων της $\mathbf{w}$.

Αυτόματα Πεπερασμένων Καταστάσεων και Τυπικές Γλώσσες

Ορισμοί από τις Τυπικές Γλώσσες

- Αν $\mathbf{L_1, L_2 \subseteq \Sigma^*}$ είναι δύο γλώσσες επί του $\mathbf{\Sigma}$, τότε δεδομένου ότι οι γλώσσες ορίσθηκαν ως σύνολα συμβολοσειρών, όλες οι πράξεις των συνόλων μπορούν να εφαρμοσθούν στις γλώσσες. Έτσι ορίζουμε και συμβολίζουμε με:
  - $\mathbf{L_1 \cap L_2 = \{ w \in \Sigma^* : w \in L_1}$ και $\mathbf{w \in L_2 \}}$, η συνολοθεωρητική τομή των δύο γλωσσών
  - $\mathbf{L_1 \cup L_2 = \{ w \in \Sigma^* : w \in L_1}$ ή $\mathbf{w \in L_2 \}}$, η συνολοθεωρητική ένωση των δύο γλωσσών
- Βεβαίως ορίζονται οι εξής γλώσσες:
  - $\mathbf{L_1 L_2 = \{ w \in \Sigma^* : w = w_1 w_2,\ w_1 \in L_1}$ και $\mathbf{w_2 \in L_2 \}}$,
  - $\mathbf{L_1 \mid L_2 = \{ w \in \Sigma^* : w = w_1 \mid w_2,\ w_1 \in L_1}$ και $\mathbf{w_2 \in L_2 \}}$,

Αυτόματα Πεπερασμένων Καταστάσεων και Τυπικές Γλώσσες

Παραδείγματα:

- Αν $\boldsymbol{\Sigma=\{0,1\}}$, τότε $\boldsymbol{\Sigma^*=\{0,1\}^*}$
- Αν $\boldsymbol{\Sigma=\{0,1\}}$, τότε $\boldsymbol{\Sigma^2=\{00,01,10,11\}}$
- Αν $\boldsymbol{\Sigma=\{1,2,3\}}$, τότε $\boldsymbol{\Sigma^2=\{11,\ 22,\ 33,12,13,21,23,31,32\}}$
- Αν $\boldsymbol{\Sigma=\{1,2,3\}}$, και $\mathbf{L=\{1,22,333\}}$ τότε οι συμβολοσειρές $\mathbf{3331,}$ $\mathbf{122,2222,22333\ \in L^2}$
- Αν $\boldsymbol{\Sigma=\{0,1\}}$ και $\mathbf{L=\{01,10\}}$ τ ότε η συμβολοσειρά **e** $\mathbf{\in L^*}$

Αυτόματα Πεπερασμένων Καταστάσεων και Τυπικές Γλώσσες

Πρόταση

Αν $\mathbf{L_1, L_2 \subseteq \Sigma^*}$ είναι δύο γλώσσες επί του $\mathbf{\Sigma}$. Τότε ισχύουν οι ακόλουθες σχέσεις:

- $\mathbf{(L_1 \cap L_2)^* \subseteq L_1^* \cap L_2^*}$
- $\mathbf{L_1^* \mid L_2^* \subseteq (L_1 \mid L_2)^*}$ ή αλλιώς, $\mathbf{L_1^* + L_2^* \subseteq (L_1 + L_2)^*}$
- $\mathbf{L\,L^* = L^*\,L = L^+}$
- $\mathbf{L^* = L\,L^* \cup \{e\}}$

Απόδειξη: (παραλείπεται)

Αυτόματα Πεπερασμένων Καταστάσεων και Τυπικές Γλώσσες

Ασκήσεις

- Αν $\boldsymbol{\Sigma}=\{\mathbf{0}, \mathbf{1}\}$ και $\mathbf{L} = \{\mathbf{00}, \mathbf{11}, \mathbf{01}, \mathbf{10}\}$ μια γλώσσα επί του $\boldsymbol{\Sigma}$ τότε ορίστε τη γλώσσα των οκτάδων (octets).
- Αν $\boldsymbol{\Sigma}=\{\mathbf{0}, \mathbf{1}\}$ και $\mathbf{L} = \{\mathbf{0000}\}$ μια γλώσσα επί του $\boldsymbol{\Sigma}$. Εξηγείστε τι συμβολίζει η γλώσσα $\mathbf{L}\boldsymbol{\Sigma}^{4}$ όπου $\boldsymbol{\Sigma}^{4} = \{\mathbf{w} \in \boldsymbol{\Sigma}^{*} : |\mathbf{w}| = \mathbf{4}\}$.
- Αν $\boldsymbol{\Sigma}=\{\mathbf{0}, \mathbf{1}\}$, δώστε παραδείγματα συμβολοσειρών της γλώσσας $\mathbf{L}$ όπου, $\mathbf{L} = \{\mathbf{w} \in \boldsymbol{\Sigma}^{*} : \mathbf{w} = \mathbf{01}^{*}\}$
- Αν $\boldsymbol{\Sigma}=\{\mathbf{1}, \mathbf{2}, \mathbf{3}\}$ και $\mathbf{L} = \{\mathbf{3}, \mathbf{12}, \mathbf{123}, \mathbf{2222}\}$ μια γλώσσα επί του $\boldsymbol{\Sigma}$. Δείξτε ότι $\mathbf{3123222212} \in \mathbf{L}^{4} \cap \mathbf{L}^{5}$.
- Αν $\mathbf{L}$ είναι μια γλώσσα επί ενός αλφαβήτου $\boldsymbol{\Sigma}$, τι θα πρέπει να ισχύει για το αλφάβητο και για τη γλώσσα ώστε οποιαδήποτε για συμβολοσειρά $\mathbf{w} \in \mathbf{L}^{4}$ να ισχύει $|\mathbf{w}| = \mathbf{4}$;

Κανονικές Εκφράσεις και Κανονικές Γλώσσες

Ορισμός

Έστω **Σ** ένα αλφάβητο**.** Μια Κανονική Έκφραση (ΚΕ) είναι μια έκφραση που συντίθεται από τα σύμβολα του αλφαβήτου, από τα σύμβολα των πράξεων επί συμβολοσειρών **(∘, | , *)** : παράθεση, διάζευξη, επανάληψη (Kleene star) για τις αντίστοιχες πράξεις και τέλος από παρενθέσεις **(,) .**

Η χρήση των παρενθέσεων επιβάλλεται για να τονιστεί η προτεραιότητα εφαρμογής των τελεστών των πράξεων όπου αυτή αμφισβητείται. Η προτεραιότητα των πράξεων είναι :

→ Kleene star, παράθεση και τέλος διάξευξη (υψηλότερη προς χαμηλότερη).

Έτσι, οι παρενθέσεις μπορούν να παραλείπονται:

(a)|((b)*(c)) = a| b*c.

Κανονικές Εκφράσεις και Κανονικές Γλώσσες

Ορισμός (συνέχεια)

Μια ΚΕ παράγεται αναδρομικά με χρήση των κανόνων :

1. Η κενή συμβολοσειρά $\mathbf{e}$ όπως και κάθε σύμβολο του αλφαβήτου είναι ΚΕ
2. Αν $\mathbf{r}$ είναι ΚΕ τότε $(\mathbf{r})$ είναι ΚΕ
3. Αν $\mathbf{r}$ και $\mathbf{s}$ είναι ΚΕ τότε $\mathbf{r} \circ \mathbf{s}$, δηλαδή η $\mathbf{r\,s}$, είναι ΚΕ
4. Αν $\mathbf{r}$ και $\mathbf{s}$ είναι ΚΕ τότε $\mathbf{r} \mid \mathbf{s}$ $(\mathbf{r} + \mathbf{s})$ είναι ΚΕ
5. Αν $\mathbf{r}$ είναι ΚΕ τότε $\mathbf{r}^*$ είναι ΚΕ

Κανονικές Εκφράσεις και Κανονικές Γλώσσες

Παρατηρήσεις:

- ΚΕ είναι οποιαδήποτε έκφραση παράγεται αποκλειστικά με τη χρήση των προηγούμενων κανόνων
- Δεν υπάρχει ΚΕ μήκους $\mathbf{0}$
- Οι ΚΕ μήκους $\mathbf{1}$ είναι η **e** και τα σύμβολα του αλφαβήτου Σ
- Η ΚΕ μήκους $\mathbf{2}$ είναι οι συμβολοσειρές του $\boldsymbol{\Sigma}^2$ καθώς και οι συμβολοσειρές της μορφής $\boldsymbol{\sigma}$*****, όπου $\boldsymbol{\sigma} \in \boldsymbol{\Sigma}$

  $\rightarrow$ Ορισμένοι συγγραφείς θεωρούν ότι η μορφή **( )** είναι ΚΕ μήκους 2
- Οι ΚΕ λειτουργούν ως μηχανισμοί παραγωγής συμβολοσειρών

Κανονικές Εκφράσεις και Κανονικές Γλώσσες

Παραδείγματα : Έστω ότι **Σ = {a, b}.** Οι επόμενες μορφές είναι κανονικές εκφράσεις:

- **a | ba,**
- **a*(bb)***
- **(b | a)*bbb**
- **(a | b)*(aab)(a | b)***

Οι επόμενες συμβολοσειρές παράγονται από την ΚΕ **a*(bb)***
e, abb, aaaaa, bb, aaabbbb, aaaaaaa, aabb, …

ενώ δε συμβαίνει το ίδιο με τις επόμενες συμβολοσειρές
ab, bbb, abbb, bba, ababa, baa, aabba, …

Η δόμηση σε blocks ενός προγράμματος σε μια block-structured programming language ακολουθεί τους κανόνες δόμησης μιας ΚΕ.

Κανονικές Εκφράσεις και Κανονικές Γλώσσες

Ερωτήματα : Έστω ότι $\boldsymbol{\Sigma = \{a, b\}}$.

- Για μια συγκεκριμένη ΚΕ είναι δυνατό να διατυπώσουμε λεκτικά τη μορφή των συμβολοσειρών που παράγει;
- Μια λεκτική διατύπωση όπως, «οι συμβολοσειρές που δεν περιέχουν ταυτόχρονα τις υποσυμβολοσειρές $\mathbf{aa}$ και $\mathbf{bbb}$» μπορεί να περιγραφτεί με μια ΚΕ;
- Για μια δεδομένη συμβολοσειρά είναι δυνατό να απαντήσουμε ότι παράγεται από συγκεκριμένη ΚΕ;
- Αντίστροφα, για μια δεδομένη συμβολοσειρά είναι δυνατό να απαντήσουμε ότι δεν παράγεται από συγκεκριμένη ΚΕ;

Κανονικές Εκφράσεις και Κανονικές Γλώσσες

Ορισμός

Έστω ότι $\Sigma = \{a, b\}$. Μια γλώσσα $L \subseteq \Sigma^*$ ονομάζεται κανονική αν όλες οι συμβολοσειρές της γλώσσας παράγονται από μια ΚΕ. Αν με r συμβολίσουμε μια ΚΕ τότε η γλώσσα που αντιστοιχεί (παράγεται) από την r συμβολίζεται με $L(r)$.

Παραδείγματα:

- Αν **r=a | b*a** τότε οι επόμενες συμβολοσειρές ανήκουν στην κανονική γλώσσα $L(r)$: **a, ba, bba, bbba, bbbba, …**

  Οποιαδήποτε άλλη συμβολοσειρά δεν ανήκει στην $L(r)$
- Αν **r= (ab | aa)*** τότε οι επόμενες συμβολοσειρές ανήκουν στην κανονική γλώσσα $L(r)$: **e, ab, abaa, aaab, ababab, ababaa, …**

  Αντίθετα συμβολοσειρές όπως οι, **bbb, baaab, bbabba, …** δεν ανήκουν στην $L(r)$

Κανονικές Εκφράσεις και Κανονικές Γλώσσες

Ασκήσεις

1. Περιγράψτε σε φυσική γλώσσα τη μορφή των συμβολοσειρών που παράγει η ΚΕ $\mathbf{r = a\,b^*a}$.
2. Περιγράψτε σε φυσική γλώσσα τη μορφή των συμβολοσειρών της γλώσσας που παράγεται από την ΚΕ $\mathbf{r = a\,bbb\,(a \mid b)^*}$.
3. Έστω το αλφάβητο $\mathbf{\Sigma=\{a, b, c\}}$.
   - Η έκφραση $\mathbf{r = (ab \mid aa)^*}$ είναι μια ΚΕ που παράγει μια γλώσσα $\mathbf{L(r) \subseteq \Sigma^*}$;
   - Η έκφραση $\mathbf{r = abc \mid e}$ είναι μια ΚΕ που παράγει μια γλώσσα $\mathbf{L(r) \subseteq \Sigma^*}$;
   - Η έκφραση $\mathbf{r = (abc)^* \mid (cd)^*}$ είναι μια ΚΕ που παράγει μια γλώσσα $\mathbf{L(r) \subseteq \Sigma^*}$;

Κανονικές Εκφράσεις και Κανονικές Γλώσσες

Ασκήσεις

4. Έστω το αλφάβητο $\boldsymbol{\Sigma=\{0,1\}}$. Περιγράψτε την ΚΕ που παράγει τις συμβολοσειρές που περιέχουν ένα μη μηδενικό και άρτιο αριθμό από **1**.
5. Έστω το αλφάβητο $\boldsymbol{\Sigma=\{0,1\}}$. Περιγράψτε την ΚΕ που παράγει τις συμβολοσειρές που περιέχουν τη $\mathbf{0011}$ ως υποσυμβολοσειρά.
6. Έστω το αλφάβητο $\boldsymbol{\Sigma=\{a,b\}}$. Ποιές από τις επόμενες συμβολοσειρές ανήκουν στη γλώσσα $\mathbf{L(r)}$ που παράγεται από την ΚΕ $\mathbf{r = (ba \mid aa)^*}$;

$\rightarrow$ $\mathbf{ba, bbaba, babaaa, aaba, aabba, e,}$

Κανονικές Γλώσσες και Αναγνώριση

Έστω ότι $\mathbf{\Sigma = \{a, b\}}$ και μια κανονική γλώσσα $\mathbf{L \subseteq \Sigma^*}$. Ένα από τα βασικά υπολογιστικά προβλήματα είναι η εύρεση ενός αλγορίθμου ο οποίος αν δεχτεί στην είσοδό του μια συμβολοσειρά θα αποφασίζει αν αυτή η συμβολοσειρά ανήκει ή όχι στη γλώσσα $\mathbf{L}$.

Αν με $\mathbf{r}$ συμβολίσουμε την ΚΕ που παράγει τις συμβολοσειρές της γλώσσας, τότε θα έχουμε $\mathbf{L{=}L(r)}$.

Από τη Θεωρία Υπολογισμού είναι γνωστό ότι, οι κανονικές γλώσσες είναι εκείνες οι γλώσσες για τις οποίες είναι δυνατό να γραφεί ένας αλγόριθμος ο οποίος θα αποφασίζει αν αν μια συμβολοσειρά ανήκει ή όχι σε μια κανονική γλώσσα.

Για τον αλγόριθμο αυτό είναι φυσιολογικό να τεθεί το ερώτημα ποιά είναι τα υπολογιστικά χαρκτηριστικά του από την άποψη της μνήμης και του χρόνου που χρειάζεται για την αναγνώριση ή μη μιας συμβολοσειράς.

Κανονικές Γλώσσες και Αναγνώριση

Έτσι είναι γνωστό ότι, αν $\mathbf{w} \in \boldsymbol{\Sigma}^*$ είναι μια συμβολοσειρά τότε υπάρχει αλγόριθμος ο οποίος μπορεί να αποφασίσει αν η συμβολοσειρά ανήκει στην κανονική γλώσσα $\mathbf{L}$ σε χρόνο $\mathcal{O}(|\mathbf{w}|)$ ενώ η μνήμη που απαιτείται για ένα τέτοιο αλγόριθμο είναι ανεξάρτητη από το μήκος της συμβολοσειράς .

Εκείνο που ουσιαστικά ενδιαφέρει τον αλγόριθμο αυτό είναι ένας αριθμός από καταστάσεις και όχι ο αριθμός των συμβόλων αυτός καθαυτός.

Παράδειγμα: Οι δυαδικοί αριθμοί που έχουν τη μορφή $\mathbf{1(1|0)^*}$ ή $\mathbf{1(1+0)^*}$.

Όλοι οι αριθμοί αυτής της μορφής ξεκινούν με $\mathbf{1}$ το οποίο ακολουθείται από μια συμβολοσειρά που αποτελείται από ένα οποιοδήποτε αριθμό $\mathbf{0}$ ή $\mathbf{1}$ ακόμη και μηδενικό.

Κανονικές Γλώσσες και Αναγνώριση

Παράδειγμα: Δυαδικές συμβολοσειρές που περιέχουν τη λέξη $\mathbf{1011}$ είναι δυνατό να αναπαρασταθούν με τη μορφή $\mathbf{(0|1)^*1011(0|1)^*}$ ή $\mathbf{(0+1)^*1011(0+1)^*}$.

Σε οποιαδήποτε συμβολοσειρά αυτής της μορφής αυτό που ενδιαφέρει είναι η συμβολοσειρά $\mathbf{1011}$ ανεξάρτητα από το πλήθος των $\mathbf{0}$ ή $\mathbf{1}$ που προηγούνται ή έπονται.

Κανονικές Γλώσσες και Αναγνώριση

Από τα ανωτέρω μπορεί κάποιος να αντιληφθεί ότι τα αυτόματα πεπερασμένων καταστάσεων, τα οποία όπως έχουν ορισθεί διαθέτουν πεπερασμένη μνήμη, αποτελούν το κατάλληλο υπολογιστικό μοντέλο για την αναγνώριση μιας κανονικής γλώσσας και των λέξεων της.

Είναι γνωστό από τη Θεωρία Υπολογισμού ότι μια κανονική γλώσσα ορίζεται, επίσης, ως η γλώσσα που αναγνωρίζεται από ένα αυτόματο πεπερασμένων καταστάσεων και για το λόγο αυτό ονομάζεται και γλώσσα πεπερασμένων καταστάσεων.

Άρα, τίθεται το ζήτημα για το πως ένα αυτόματο πεπερασμένων καταστάσεων μπορεί να οριστεί ώστε να λειτουργήσει ως αναγνωριστής κανονικών γλωσσών ή κανονικών εκφράσεων.

Κανονικές Γλώσσες και Αναγνώριση

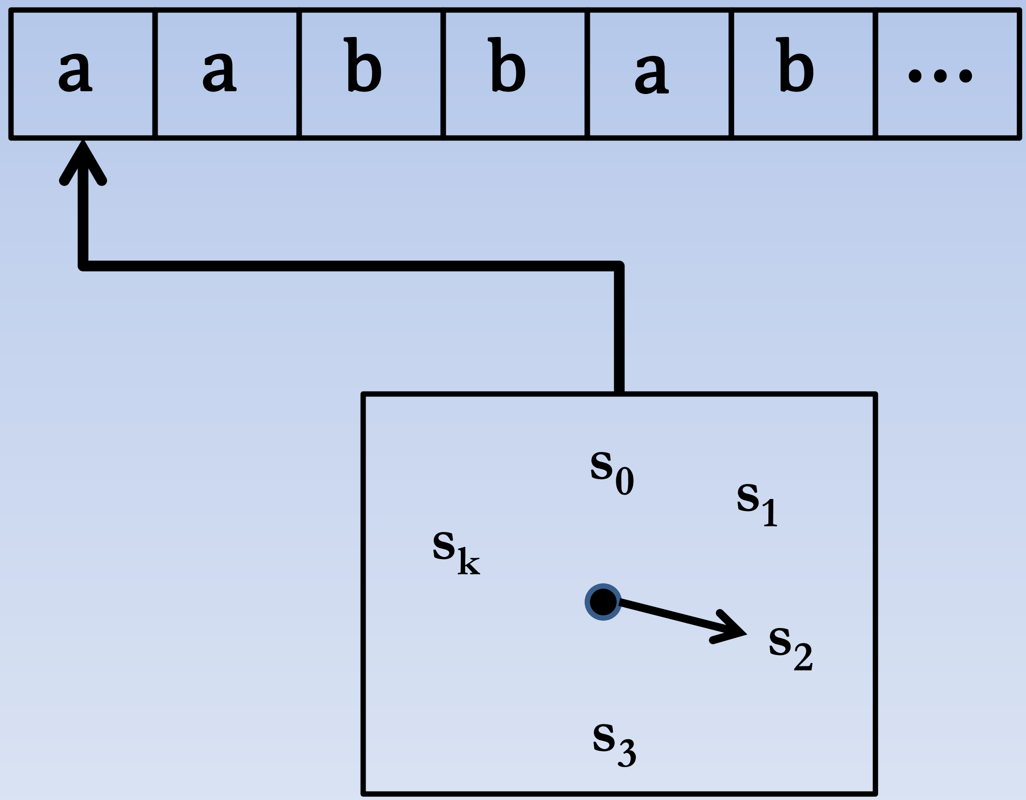
Για να χρησιμοποιηθεί ένα αυτόματο πεπερασμένων καταστάσεων $\mathbf{M=(S, s_0, I, O, f, g)}$ ως αναγνωριστής γλώσσας εισάγουμε τις ακόλουθες τροποποιήσεις:

- Το σύνολο των σημάτων εξόδου $\mathbf{O=\{o_1, o_2, \cdots, o_m\}}$ καταργείται.
- Η συνάρτηση εξόδου $\mathbf{g : S \to O}$ καταργείται κατά συνέπεια.
- Ορίζεται ένα υποσύνολο $\mathbf{F}$ του συνόλου των καταστάσεων $\mathbf{F \subseteq S}$ το οποίο περιλαμβάνει τις λεγόμενες τελικές ή αλλίως τερματικές καταστάσεις ή καταστάσεις αποδοχής.
- Ορίζεται μια κατάσταση η οποία ονομάζεται κατάσταση παγίδα, ή αλλιώς καταβόθρα, και είναι δυνατόν να συμβολισθεί με $\mathbf{s_\nabla}$ ή $\mathbf{s_\times}$.

Έτσι το αυτόματο πεπερασμένων καταστάσεων θα συμβολίζεται ως $\mathbf{M=(S, s_0, I, F, f)}$ ή ακόμη $\mathbf{M=(S, s_0, I, F, \delta)}$.

Κανονικές Γλώσσες και Αναγνώριση

Σχηματικά ένα αυτόματο $\mathbf{M=(S, s_0, I, F, \delta)}$ μπορεί να περιγραφεί ως εξής:



Κανονικές Γλώσσες και Αναγνώριση

Κατά τη λειτουργία του το αυτόματο ξεκινά πάντα από την αρχική κατάσταση. Τα σύμβολα της συμβολοσειράς διαβάζονται από αριστερά προς τα δεξιά. Κάθε φορά που ένα σύμβολο διαβάζεται αυτό καταναλώνεται, δηλαδή διαγράφεται από τη συμβολοσειρά.

Με κάθε νέο σύμβολο που διαβάζεται το αυτόματο μετακινείται σε μία νέα κατάσταση όπως περιγράφεται από τη συνάρτηση μετάβασης.

Η λειτουργία του αυτομάτου τερματίζεται όταν έχουν διαβαστεί όλα τα σύμβολα της συμβολοσειράς η οποία είναι αποδεκτή αν το αυτόματο βρεθεί σε μια κατάσταση αποδοχής.

Αν το αυτόματο με την ολοκλήρωση της συμβολοσειράς βρεθεί σε οποιαδήποτε άλλη κατάσταση τότε μπλοκάρει και η συμβολοσειρά απορρίπτεται.

Κανονικές Γλώσσες και Αναγνώριση

Παράδειγμα: Έστω το αλφάβητο $\mathbf{\Sigma=\{a, b\}}$. Ζητείται να περιγραφεί το αυτόματο πεπερασμένων καταστάσεων που αναγνωρίζει τη γλώσσα οι συμβολοσειρές της οποίας περιέχουν ένα αριθμό από **b** ο οποίος είναι πολλαπλάσιος του **3**. Τα σύμβολα **b** μπορούν να εμφανίζονται σε οποιοδήποτε σημείο και με οποιοδήποτε αριθμό σε μια συμβολοσειρά όπως στη συνέχεια:

bbb, abbbaa, baabba, aaababaaaba, …

Η ΚΕ που περιγράφει τις συμβολοσειρές της γλώσσας είναι:

a*ba*ba*ba*

Ποιά είναι η ΚΕ αν δεχτούμε ότι οι συμβολοσειρές μπορούν να μην περιέχουν **b** δεδομένου ότι μηδενικός αριθμός προκύπτει ως $\mathbf{3\times 0}$;

Κανονικές Γλώσσες και Αναγνώριση

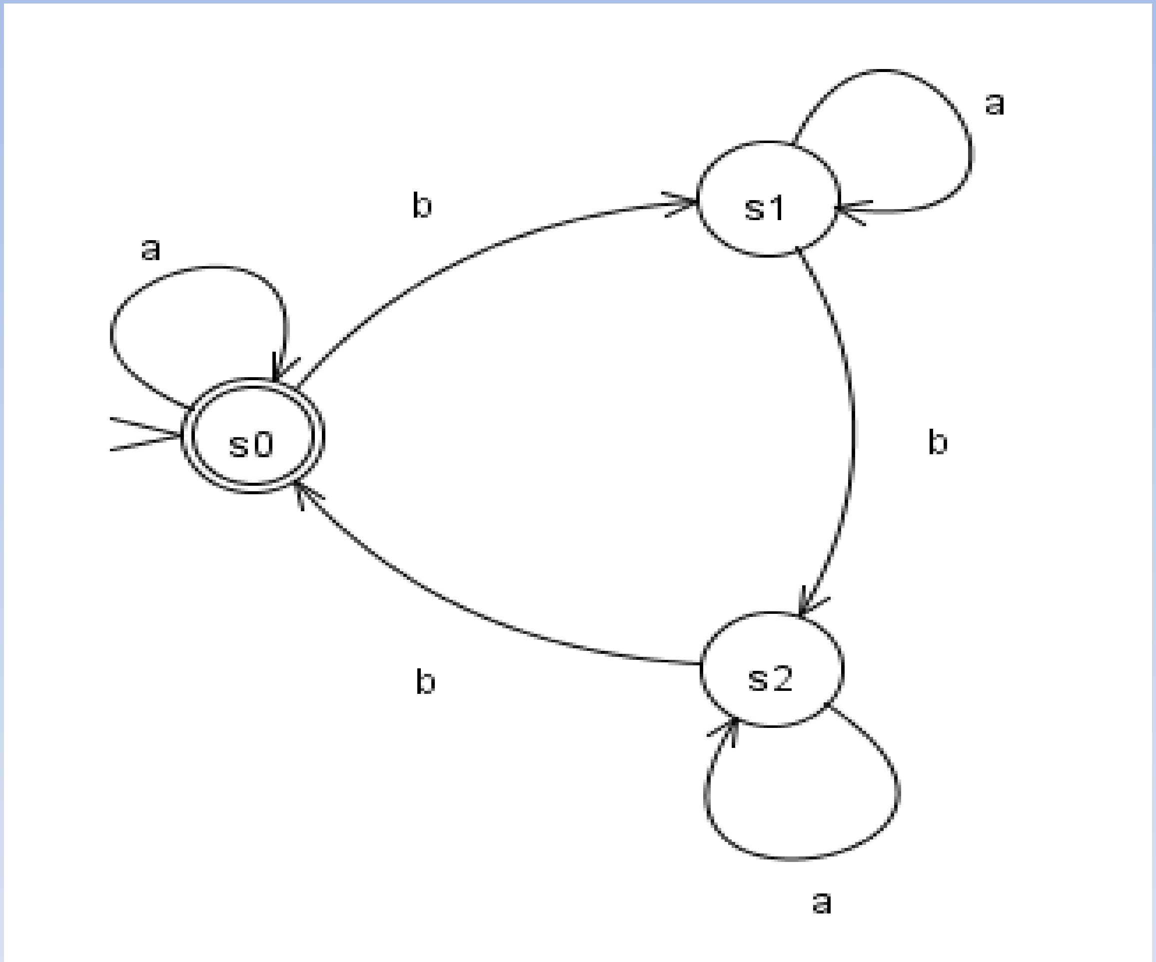
Ένα αυτόματο πεπερασμένων καταστάσεων που αναγνωρίζει τις συμβολοσειρές αυτής της γλώσσας πρέπει:

1. Να είναι αιτιοκρατικό
2. Να αναγνωρίζει όλες τις συμβολοσειρές που ανήκουν στη γλώσσα και μόνον αυτές
3. Να απορρίπτει όλες τις συμβολοσειρές που δεν ανήκουν στη γλώσσα.

Σημειώνεται ότι συνήθως η συνάρτηση μεταβασης περιγράφεται με τη μορφή διαγραμμάτων μετάβασης όπως στη συνέχεια.

Έτσι, το αυτόματο για το παράδειγμα αυτό είναι το ακόλουθο:

Κανονικές Γλώσσες και Αναγνώριση



Κανονικές Γλώσσες και Αναγνώριση

Ερωτήματα επί του παραδείγματος:

1. Είναι αιτιοκρατικό;
2. Αναγνωρίζει όλες τις συμβολοσειρές που ανήκουν στη γλώσσα και μόνον αυτές;
3. Απορρίπτει όλες τις συμβολοσειρές που δεν ανήκουν στη γλώσσα;
4. Είναι πλήρες ως προς τη σχεδίαση;