Θεωρία Υπολογισμού

Βιβλιογραφία:

Α. Δημητρίου, Θεμελιώσεις της Επιστήμης των Η/Υ. Τόμος γ' , Αυτόματα και Τυπικές Γλώσσες, Εκδ. ΕΑΠ.

Κ. Χαλάτση, Θεμελιώσεις της Επιστήμης των Η/Υ. Τόμος β' , Θεωρία Υπολογισμού, Εκδ. ΕΑΠ.

M. Sipser, Introduction to the theory of computation, PWS Publishing Co.

H. Lewis, C. Papadimitriou: Στοιχεία θεωρίας υπολογισμού, 1η έκδοση Εκδ. Τ.Ε.Ε. 2η έκδοση εκδ. Κριτική.

John E. Hopcroft, Rajeev Motwani, Jeffrey D. Ullman Introduction to Automata Theory, Languages, and Computation (2nd Edition).

J. Hopcroft, , J. Ullman - Introduction to Automata Theory, Languages, and Computation (1st Edition).

Σχέσεις, σύνολα, αλφάβητα, γλώσσες.

Τυπική Λογική

Προτάσεις: Εκφράσεις οι οποίες έχουν σαφώς καθορισμένη και μόνο μία τιμή αλήθειας (Boolean values), δηλωτικές εκφράσεις, δεν υπάρχουν μεταβλητές.

Τα μάτια του Γιάννη είναι πράσινα.

Ποιο είναι το όνομά σου; (δεν είναι πρόταση).

Ενδιαφερόμαστε για μαθηματικές εκφράσεις: Πχ $3x = 7$. Η αληθοτιμή τους εξαρτάται από το σύνολο που αναφερόμαστε, στους φυσικούς και τους ακέραιους είναι ψευδής, στους ρητούς (κλάσματα) είναι αληθής.

Παρατήρηση: μεγαλώνουμε τον κόσμο μας για να δώσουμε νόημα σε κάποιες εκφράσεις, φυσικοί αριθμοί, ακέραιοι αριθμοί, ρητοί αριθμοί, άρρητοι αριθμοί (αλγεβρικοί $\sqrt{3}$, υπερβατικοί π, e).

Ατομικές προτάσεις. Δεν αναλύονται σε απλούστερες, Το σακάκι που φοράει ο Νίκος είναι κόκκινο.

Σύνθετες προτάσεις: Σύζευξη, διάζευξη, άρνηση, συνεπαγωγή. Λογικοί σύνδεσμοι.

Αν βρέξει αύριο θα πάμε στο βουνό (συνεπαγωγή). Ο αριθμός $\sqrt{3}$ είναι ρίζα της εξίσωσης $3x^2 - 5 = 0$, είτε ο αριθμός $\sqrt{2}$ είναι ρίζα της εξίσωσης $3x - 7 = 0$.

Αν q συμβολίζει μια πρόταση, τότε με $\neg q$ (όχι q) συμβολίζουμε την άρνησή της.

Αν p, q συμβολίζουν προτάσεις, τότε με $p \wedge q$ (p και q) συμβολίζουμε την σύζευξή τους.

Αν p, q συμβολίζουν προτάσεις, τότε με $p \vee q$ (p είτε q) συμβολίζουμε την διάζευξη τους, δεν αποκλείουμε να ισχύουν και οι δύο προτάσεις, εγκλειστική διάζευξη.

Αν p, q συμβολίζουν προτάσεις, με $p \to q$, $p \Rightarrow q$ (αν p τότε q) συμβολίζουμε την συνεπαγωγή τους.

Αν p, q συμβολίζουν προτάσεις, τότε με $p \leftrightarrow q$, $p \Leftrightarrow q$ (p αν και μόνο αν q, p ισοδύναμο με q , p ανν q) συμβολίζουμε την ισοδυναμία των p , q.

Οι λογικοί σύνδεσμοι αποκτούν τιμές αλήθειας από του πίνακες που ακολουθούν όπου T, F συμβολίζουν τις έννοιες αλήθεια, ψεύδος:

p	q	$p \vee q$	$p \wedge q$	$p \to q$	$p \leftrightarrow q$
T	T	T	T	T	T
T	F	T	F	F	F
F	T	T	F	T	F
F	F	F	F	T	T

Σύνολα: Σύνολο είναι μία σαφώς καθορισμένη συλλογή από αντικείμενα. Πχ τα σύμβολα 1, 2, 3, 4. Συμβολισμός $\{1, 2, 3, 4\}$ (αναγραφή, απαρίθμηση, πρώτο στοιχείο, δεύτερο, κλπ). Μπορούμε να ορίσουμε σύνολα με περιγραφή της ιδιότητας των στοιχείων του συνόλου. Πχ Το σύνολο των διαιρετών του αριθμού 12345678910.

Αν έχουμε άπειρα στοιχεία περιγράφουμε ένα σύνολο με την χαρακτηριστική του ιδιότητα (αυτή η ιδιότητα είναι μοναδική), πχ το σύνολο των αρτίων αριθμών, $\{2, 4, 6, 8, \ldots\} = \{2k | k \in \mathbb{N}\}$.

Το σύμβολο $\in$ δηλώνει την σχέση μέλους ενός συνόλου. Η άρνησή του συμβολίζεται με $\notin$. Σε σύνολα δεν έχουμε επανάληψη των στοιχείων.

Υποσύνολο ενός συνόλου Α είναι ένα άλλο σύνολο Β που περιέχει κάποια από τα στοιχεία του Α. Αν $A = \{1, 2, 3, 4\}$, τότε το σύνολο $\{1, 3, 4\}$ είναι υποσύνολο του Α και το συμβολίζουμε με $\{1, 3, 4\} \subseteq A$. Αν δεν έχουμε ισότητα τότε ο συμβολισμός είναι $\{1, 3, 4\} \subset A$. Ισχύει για

κάθε σύνολο A, $A \subseteq A$. Το σύνολο που δεν περιέχει κανένα στοιχείο ονομάζεται κενό σύνολο και συμβολίζεται με$\emptyset$. Ισχύει $\emptyset \subseteq A$ για κάθε σύνολο A. Για δύο σύνολα A, B ισχύει ότι $A = B$ ακριβώς όταν (τότε και μόνο τότε, ανν) $A \subseteq B$ και $B \subseteq A$.

Το σύνολο όλων των υποσυνόλων ενός συνόλου A ονομάζεται δυναμοσύνολο του A και συμβολίζεται με 2^A είτε $\mathcal{P}(A)$. Οι λογικές πράξεις, (λογικοί σύνδεσμοι) μπορούν να ορίσουν τις πράξεις σε σύνολα (είτε αντίστροφα), η άρνηση ορίζει το συμπλήρωμα ενός συνόλου A ως προς ένα άλλο B, η διάζευξη την ένωση συνόλων, η σύζευξη την τομή, η συνεπαγωγή το υποσύνολο, η ισοδυναμία την ισότητα. Με σύμβολα έχουμε $A \cup B = \{x : x \in A \vee x \in B\}$, $A \cap B = \{x : x \in A \wedge x \in B\}$. Αν $A \subseteq B$, το συμπλήρωμα του A ως προς το B ορίζεται σαν $B - A = \{x : x \in B \wedge x \notin A\}$.

Ιδιότητες των πράξεων των συνόλων

1. Μεταθετική ιδιότητα της τομής και της ένωσης $A \cup B = B \cup A$, $A \cap B = B \cap A$.

2. Οι πράξεις της τομής και ης ένωσης είναι αδύναμες $A \cup A = A$, $A \cap A = A$.

3. Προσεταιριστική (δεν έχει σημασία η σειρά με την οποία εκτελούμε τις πράξεις) $(A \cup B) \cup C = A \cup (B \cup C)$, $(A \cap B) \cap C = A \cap (B \cap C)$.

4. Επιμεριστική της τομής ως προς την ένωση $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$, $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$.

5. Νόμος απορρόφησης$(A \cup B) \cap A = A$, $(A \cap B) \cup A = A$.

6. Νόμοι του De Morgan $X - (A \cup B) = (X - A) \cap (X - B)$, $X - (A \cap B) = (X - A) \cup (X - B)$.

Οι ιδιότητες των πράξεων των συνόλων είναι οι γνωστές ιδιότητες των Αλγεβρών Boole. Είναι ανάλογες με τις ιδιότητες των λογικών συνδέσμων.

Σχέσεις: Θέλουμε να εκφράσουμε ένα γράφημα (την σχέση ότι δύο κορυφές συνδέονται), ένα δέντρο (ότι ένας κόμβος είναι αμέσως μετά από κάποιο άλλο κόμβο), κάποιες οντότητες σχετίζονται με κάποιο τρόπο (το σημείο Α βρίσκεται ανάμεσα στο σημεία Β και Γ), διάταξη σε ένα σύνολο.

Διαταγμένο ζευγάρι των a, b (το συμβολίζουμε με (a, b)) είναι το σύνολο $\{\{a\}, \{a, b\}\}$. Με βάση τον ορισμό αυτό μπορούμε να έχουμε ότι $(a,\ b) \neq (b,\ a)$ για $a \neq b$ και $(a,\ b) = (c,\ d)$ αν και μόνο αν $a = c$ και $b = d$.

Για δύο σύνολα A, B το Καρτεσιανό γινόμενο των A, B είναι σύνολο από όλα τα διαταγμένα ζευγάρια $(a,\ b)$ με $a \in A$, $b \in B$. Το Καρτεσιανό γινόμενο των συνόλων A, B το συμβολίζουμε με $A \times B$ Παράδειγμα για $A = \{1,\ 2,\ 3,\ 4\}, B = \{a, e\}, A \times B = \{(1, a)\ ,\ (1, e)\ ,\ (2, a)\ ,\ (3, e)\ ,\ (4, a)\ ,\ (4, e)\}$.

Τα σύνολα εκφράζουν ιδιότητες μίας μεταβλητής, χρειαζόμαστε να εκφράσουμε ιδιότητες που περιέχουν πολλές μεταβλητές, η σχετική έννοια είναι η έννοια της σχέσης.

Σχέση διμελής ανάμεσα στα σύνολα A, B ονομάζεται κάθε υποσύνολο του Καρτεσιανού γινομένου των A, B. Συνήθως μια σχέση την συμβολίζουμε με R (relation). Παρ. στο σύνολο $A = \{1,\ 2,\ 3,\ 4\}$ ορίζουμε την σχέση μικρότερος (η σχέση αυτή είναι υποσύνολο του $A \times A = A^2$). Στο ίδιο σύνολο η σχέση "ο x διαιρεί τον y" είναι μία άλλη διμελής σχέση. Αν $R \subseteq A \times B = \{(x, y) : xRy\}$ είναι μία διμελής σχέση τότε το σύνολο $\{(y, x) : xRy\}$ ονομάζεται αντίστροφη σχέση της R και συμβολίζεται με R^{-1}.

Όπως ορίσαμε διαταγμένο ζευγάρι, όμοια μπορούμε να ορίσουμε διαταγμένη τριάδα, τετράδα κλπ. και να ορίσουμε το Καρτεσιανό γινόμενο τριών, τεσσάρων κλπ συνόλων. Παρ. Η πρόσθεση αριθμών μπορεί να ιδωθεί και σαν σχέση τριών αριθμών δηλ. $\{(x, y, z) : x, y, x \in,\ x + y = z\}$.

Συναρτήσεις: εκφράζουμε την εξάρτηση μίας ποσότητας – μεταβλητής από κάποια άλλη, $y = 2x + \log_2 \left(x^2 - 7\right)$.

Στην πληροφορική μας ενδιαφέρει την εξάρτηση αυτή να την εκφράσουμε με αλγόριθμο, είτε να αποδείξουμε ότι δεν είναι δυνατόν αυτή η εξάρτηση να εκφραστεί με αλγόριθμο.

Συνάρτηση με πεδίο ορισμού το σύνολο και πεδίο τιμών το σύνολο B είναι μία σχέση f από το A στο B, $f \subseteq A \times B$, ώστε για κάθε a από το A υπάρχει ακριβώς ένα b από το B ώστε $(a, b) \in f$, συμβολικά το εκφράζουμε $\forall a \in A \ \exists' b \in B : (a, b) \in f$, δηλ. η εξίσωση $f(a) = x$ έχει το πολύ μία λύση $\forall a \in A$. Ο συμβολισμός για μία συνάρτηση f από το A στο B είναι $f : A \to B$. Αν θέλουμε να δώσουμε και τον τύπο της συνάρτησης (όταν υπάρχει) χρησιμοποιούμε τον συμβολισμό $f : x \mapsto f(x)$, πχ $f : x \mapsto f(x) = y = 2x + \log_2\left(x^2 - 7\right)$.

Αν για μία συνάρτηση f ισχύει $x \neq x' \to f(x) \neq f(x')$ τότε έχουμε μία ένα προς ένα συνάρτηση. Αν για μία συνάρτηση $f : A \to B$ ισχύει $\forall b \in B \ \exists a \in A : f(a) = b$ έχουμε μια επί συνάρτηση. Μία συνάρτηση f είναι επί αν η εξίσωση $f(x) = b$ έχει λύση για κάθε $b \in B$. Μία συνάρτηση f είναι ένα προς ένα αν η εξίσωση $f(x) = b$ έχει το πολύ μία λύση για κάθε $b \in B$. Αν μια συνάρτηση είναι ένα προς ένα και επί τότε και η αντίστροφη σχέση f^{-1} είναι συνάρτηση.

Ακολουθίες.

Μια συνάρτηση με πεδίο ορισμού τους φυσικούς αριθμούς ονομάζεται ακολουθία. Στις ακολουθίες χρησιμοποιούμε τον συμβολισμό a_n, $\{a_n : n \in \mathbb{N}\}$. Παράδειγμα ακολουθίας: $a_n = 2^n$.

Αξιοσημείωτες σχέσεις: Μία σχέση $R \subseteq A^2$ ονομάζεται ανακλαστική (αυτοπαθής) αν $\forall a \in A\,(a, a) \in R$. Μία σχέση $R \subseteq A \times A$ ονομάζεται συμμετρική αν $(a, b) \in R \to (b, a) \in R$. Μία σχέση $R \subseteq A \times A$ ονομάζεται αντισυμμετρική αν $(a, b) \in R, a \neq b \to (b, a) \notin R$. Μία σχέση $R \subseteq A \times A$ ονομάζεται μεταβατική αν $(a, b) \in R, (b, c) \in R \to (a, c) \in R$.

Μία σχέση στο σύνολο A ονομάζεται σχέση ισοδυναμίας αν είναι ανακλαστική, συμμετρική και μεταβατική (ισότητα σε σχέση με το ύψος).

Μία σχέση στο σύνολο A ονομάζεται σχέση διάταξης (διάταξη) αν είναι ανακλαστική, αντισυμμετρική και μεταβατική. Αν

$$\forall a, b \in A, (a, b) \in R \vee (b, a) \in R$$

έχουμε σχέση ολικής διάταξης στο A.

Αλφάβητα, συμβολοσειρές.

Αλφάβητο ονομάζουμε ένα σύνολο από σύμβολα. Συμβολοσειρά (λέξη, string) από το αλφάβητο Σ ονομάζουμε μια πεπερασμένη ακολουθία από σύμβολα στοιχεία του Σ. Το μήκος της ακολουθίας είναι και το μήκος της συμβολοσειράς. Η συμβολοσειρά με μήκος 0 ονομάζεται κενή συμβολοσειρά και συμβολίζεται με e. Το μήκος μιας συμβολοσειράς w συμβολίζεται με $|w|$ είτε με length(w).

Στο σύνολο των συμβολοσειρών από ένα αλφάβητο ορίζουμε πράξη την παράθεση (concatenation) συμβολοσειρών. Η παράθεση συμβολοσειρών συμβολίζεται με ο. Παρ. αη ο γιαννης = αηγιαννης.

Ο μαθηματικός ορισμός της παράθεσης είναι: αν $w(1)$, $w(2)$, …, $w(n)$, $v(1)$, $v(2)$, …, $v(m)$ είναι δύο συμβολοσειρές w, v με $w(1)$, $w(2)$, …, $w(n)$, $v(1)$, $v(2)$, …, $v(m)$ τα επιμέρους ατομικά σύμβολά τους τότε $z = w \circ v$ είναι η

συμβολοσειρά για την οποία ισχύει $z(1) = w(1)$, $z(2) = w(2)$, …, $z(n) = w(n)$, $z(n+1) = v(1)$, $z(n+2) = v(2)$, …, $z(n+m) = v(m)$. Για την παράθεση συμβολοσειρών ισχύει $w \circ e = e \circ w = w$ (e είναι η κενή συμβολοσειρά). Η παράθεση των συμβολοσειρών ικανοποιεί την προσεταιριστική ιδιότητα $(w \circ v) \circ t = w \circ (v \circ t)$. Η προσεταιριστική ιδιότητα σημαίνει ότι δεν έχει σημασία η σειρά που εκτελούμε τις πράξεις και μπορούμε να παραλείπουμε τις παρενθέσεις (δεν ισχύει στην αφαίρεση αριθμών).

Ορίζουμε δυνάμεις μίας συμβολοσειράς με επαγωγή όπως τον πολλαπλασιασμό είτε την δύναμη αριθμών, δηλ

$$w^0 = e, \qquad w^{(n+1)} = w^n \circ w \text{ για } n > 0.$$

Για λόγους ευκολίας παραλείπουμε το σύμβολο $\circ$ στην παράθεση συμβολοσειρών.

Η συμβολοσειρά u είναι πρόθεμα (prefix) της συμβολοσειράς v αν $\exists t v = u \circ t$. Η συμβολοσειρά u είναι κατάληξη (suffix)

της συμβολοσειράς v αν $\exists t v = t u$. Η συμβολοσειρά u είναι υποσυμβολοσειρά (υπολέξη – substring) της συμβολοσειράς v αν $\exists t, \exists w \;\; ; v = wut$.

Αντίστροφη (ανάστροφη, reverse) συμβολοσειρά. Για μία συμβολοσειρά w με μήκος m ορίζουμε την αντίστροφη της w και την συμβολίζουμε με w^r , είτε w^R την εξής ακολουθία συμβόλων: $w^r(k) = w(m + 1 - k)$, $0 < k < m$.

Για ένα αλφάβητο Σ συμβολίζουμε με Σ^* το σύνολο όλων των συμβολοσειρών με σύμβολα από το Σ.

Μπορούμε να ορίσουμε την πράξη της αναστροφής μίας συμβολοσειρα με αναδρομή:

$a^r = a$ για $a \in \mathbf{\Sigma}$, $\qquad (wa)^r = aw^r$ για $a \in \mathbf{\Sigma}, w \in \mathbf{\Sigma}^*$

Ασκηση: Αν w είναι μία λέξη, w_1 είναι η ανάστροφή της σύμφωνα με τον πρώτο ορισμό, w_2 είναι η ανάστροφή της σύμφωνα με τον δεύτερο ορισμό, $w_1 = w_2$. Επαγωγή στο μήκος της w.

Γλώσσες

Για ένα σύνολο συμβόλων Σ ονομάζουμε τυπική γλώσσα (για συντομία γλώσσα) ένα υποσύνολο του Σ^*. Η ονομασία τυπική εννοεί ότι δεν ενδιαφερόμαστε για την ερμηνεία (σημασία) των συμβολοσειρών αλλά για την μορφή (τύπο) της γλώσσας μόνο.

Μας ενδιαφέρουν γλώσσες με άπειρα στοιχεία. Οι γλώσσες με πεπερασμένα στοιχεία περιγράφονται με απαρίθμηση των στοιχείων τους πχ $\Sigma = \{0, 1\}$, $\Gamma = \{000, 01, 10010, 1000\}$ Από τις γλώσσες με άπειρα στοιχεία μας ενδιαφέρουν οι γλώσσες που έχουν απλή περιγραφή, πχ $\Sigma = \{0, 1\}$, γλώσσες είναι τα εξής σύνολα: οι συμβολοσειρές με άρτιο αριθμό από 0, οι συμβολοσειρές που λήγουν σε 01001, οι συμβολοσειρές οι οποίες περιέχουν τον ίδιο αριθμό από 0 και 1. Οι γλώσσες μπορούν να κωδικοποιήσουν προβλήματα πχ. Τα γραφήματα (γράφοι) που έχουν χρωματικό αριθμό ίσο με k.

Κλειστότητα Kleene: Αν w είναι μία λέξη στο αλφάβητο με w^* συμβολίζουμε το σύνολο $\{e, w, w^2, w^3, \ldots\} = \{w^n \mid$

$n \in \mathbb{N}\}$. Αν L είναι μία γλώσσα με L^* συμβολίζουμε το σύνολο $\{w \mid w = w_1 \circ w_2 \circ \ldots w_n \circ,\ w_1, w_2, \ldots w_n, \in L\}$. Το σύνολο L^* ονομάζεται κλειστότητα Kleene της γλώσσας L.

Κανονικές εκφράσεις: Αν $\mathbf{\Sigma}$ είναι ένα αλφάβητο τότε μια κανονική έκφραση (ΚΕ) στο αλφάβητο αυτό είναι το σύνολο που ικανοποιεί τα εξής:

1. $\{a\}$, $\forall a \in \mathbf{\Sigma}$ είναι κανονική έκφραση,

2. $\{e\}$, $\emptyset$, είναι κανονική έκφραση,

3. $R_1 \cup R_2$, είναι κανονική έκφραση, όπου R_1, R_2 είναι κανονικές εκφράσεις,

4. $R_1 \circ R_2$, είναι κανονική έκφραση, όπου R_1, R_2 είναι κανονικές εκφράσεις,

5. R_1^* είναι κανονική έκφραση, όπου R_1, είναι κανονική έκφραση,

Τίποτα άλλο δεν είναι κανονική έκφραση.

Ο ορισμός των ΚΕ είναι αναδρομικός από απλές εκφράσεις δημιουργούμε πιο πολύπλοκες.

Καταχρηστικά θα παραλείπουμε τα άγκιστρα που συμβολίζουν το σύνολο για τα μονοσύνολα.

Προσοχή η κανονική έκφραση e δεν είναι η ίδια με την κανονική έκφραση $\emptyset$.

Οι κανονικές εκφράσεις είναι περιγραφές γλωσσών. Θα δούμε ισοδύναμο χαρακτηρισμό των κανονικών εκφράσεων με υπολογιστικ σχήματα τα πεπερασμένα αυτόματα.

Παρ.

0^*10^* η ΚΕ περιγράφει την γλώσσα που περιέχει συμβολοσειρές από το αλφάβητο $\{0, 1\}$ που περιέχουν ακριβώς ένα 1.

$\Sigma^*1\Sigma^*$ η ΚΕ περιγράφει την γλώσσα που περιέχει συμβολοσειρές από το αλφάβητο $\{0, 1\}$ που περιέχουν τουλάχιστον ένα 1.

$(\Sigma\Sigma)^*$ η ΚΕ περιγράφει την γλώσσα που περιέχει συμβολοσειρές από το αλφάβητο $\{0, 1\}$ που περιέχουν άρτιο αριθμό σύμβόλων. Το ίδιο σύνολο περιγράφεται με την κανονική έκφραση $(01 \cup 10 \cup$

$1^*\emptyset = \emptyset$. Παραθέτοντας το κενό σύνολο με οποιαδήποτε γλώσσα το αποτέλεσμα είναι το κενό σύνολο.

$\emptyset^* = e$. Το αποτέλεσμα της κλειστότητας Kleene είναι σύνολο $\{e, w, w^2, w^3, \ldots\}$δηλ. η κενή συμβολοσειρά.

$(0 \cup 1)^*\, 11\, (0 \cup 1)$ συμβολίζει τις συμβολοσειρές που περιέχουν δύο συνεχόμενα 1 στην προτελευταία και αντι προτελευταία θέση.

Αντί για το σύμβολο $\cup$ χρησιμοποιούμε το σύμβολο $+$ στις ΚΕ .

Οι ΚΕ μπορούν να περιγράψουν την λεκτική δομή ενός προγράμματος

Παρ. η έκφραση $\{+, -, e\}\ (DD^* \cup DD^*.D^* \cup D^*.DD^*)$ όπου $D = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ περιγράφει τις εκφράσεις που αντιστοιχούν στις δεκαδικές παραστάσεις αριθμών. Η λεκτική ανάλυση γλωσσών προγραμματισμού βασίζεται σε κανονικές εκφράσεις αφού οι ΚΕ περιγράφουν το λεκτικό μέρος μίας γλώσσας προγραμματισμού (αναγνωριστικά, identifiers).

Πεπερασμένα αυτόματα

Ένα πεπερασμένο αυτόματο (ΠΑ, finite automaton) είναι μία πεντάδα $M = (\Sigma, K, \delta, s, F)$, από τα εξής στοιχεία:

1. Σ ένα αλφάβητο,
2. K ένα σύνολο καταστάσεων (states),
3. s μια διακεκριμένη αρχική κατάσταση (start state),
4. $F \subset K$ τις τελικές καταστάσεις (final states) ,
5. δ μια πεπερασμένη συνάρτηση από το $K \times \Sigma$ στο K, την συνάρτηση μετάβασης (transition fuction).

Ένα ΠΑ έχει μια ταινία εισόδου όπου βρίσκεται μία συμβολοσειρά. Το αυτόματο διαβάζει το αριστερότερο σύμβολο της συμβολοσειράς

με μια κεφαλή ανάγνωσης που διαβάζει ένα ακριβώς σύμβολο. Διαβάζει το τρέχον σύμβολο από την ταινία εισόδου και αλλάζει κατάσταση σύμφωνα με την συνάρτηση μετάβασης και διαβάζει το επόμενο σύμβολο στην ταινία εισόδου. Η κατεύθυνση της κίνησης στην ταινία εισόδου είναι μία από αριστερά προς τα δεξιά.

Ένα πεπερασμένο αυτόματο ονομάζεται και μηχανή πεπερασμένων καταστάσεων (finite state machine).

Παρ. $K = \{q_0, q_1\}$, $\Sigma = \{0, 1\}$, $s = \{q_0\}$, $F = \{q_0\}$,

$$\delta = \begin{array}{ccc} \hline q & \sigma & \delta(q, s) \\ \hline q_0 & a & q_0 \\ q_0 & b & q_1 \\ q_1 & a & q_1 \\ q_1 & b & q_0 \end{array}$$

Εναλλακτικά μπορούμε να περιγράψουμε ένα ΠΑ με πίνακα

	a	b
q_1	q_1	q_2
q_2	q_1	q_2

Τελική κατάσταση είναι η q_2, το αυτόματο δέχεται τις συμβολοσειρές που λήγουν σε b. Το παραπάνω αυτόματο μπορούμε να τα εκφράσουμε γραφικά με το εξής σχήμα.

Για ένα ΠΑ ορίζουμε συνολική κατάσταση (ΣΚ) του αυτόματου (configuration) το διαταγμένο ζευγάρι $(q, w) \in K \times \Sigma^*$ όπου το ΠΑ βρίσκεται στην κατάσταση q και απομένει να διαβάσει την συμβολοσειρά w.

Παρ. $(q_1, 0101001)$. Η ερμηνεία είναι ότι το αυτόματο έχει διαβάσει κάποιο αρχικό τμήμα μιας συμβολοσειράς και απομένει να διαβάσει την συμβολοσειρά 0101001 με την κεφαλή ανάγνωσης η οποία διαβάζει το αριστερότερο σύμβολο της 0101001 και το αυτόματο βρίσκεται στην κατάσταση q_1.

Η συνολική κατάσταση του ΠΑ μας εκφράζει σε ποια κατάσταση βρίσκεται το ΠΑ και πόσο ακόμα απομένει να εξετασθεί από την συμβολοσειρά εισόδου. Το σύνολο των συνολικών καταστάσεων ενός ΠΑ το συμβολίζουμε με Config (M).

Για ένα ΠΑ ορίζουμε την διμελή σχέση παράγει σε ένα βήμα (και την συμβολίζουμε με $\vdash_M$ σαν το υποσύνολο του Config $(M) \times$ Config (M) ώστε να ισχύει:

$$(q, aw) \vdash_M \left(q', w\right)$$

όπου $\delta(q,a) = q'$. Σε αυτή την περίπτωση λέμε ότι το αυτόματο από την συνολική κατάσταση (q, aw) παράγει την (q', w) σε ένα βήμα. Η σχέση παράγει σε ένα βήμα είναι συνάρτηση. Το ΠΑ τερματίζει την λειτουργία του όταν εξαντλήσει την συμβολοσειρά από την ταινία εισόδου.

Γλώσσα που αντιστοιχεί σε ένα πεπερασμένο αυτόματο.

Μία συμβολοσειρά w είναι δεκτή από ένα πεπερασμένο αυτόματο A αν υπάρχει τελική κατάσταση $q \in F$ ώστε $(s, w) \vdash_A^* (q, e)$, δηλ. στο τέλος της συμβολοσειράς καταλήγουμε σε τελική κατάσταση. Αν δεν συμβεί το προηγούμενο τότε το αυτόματα δεν δέχεται την συμβολοσειρά (την απορρίπτει). Το σύνολο των συμβολοσειρών που καταλήγουν σε τελική κατάσταση ονομάζεται η γλώσσα που αναγνωρίζει το αυτόματο A και συμβολίζεται με $L(A)$.

Παρ. ένα αυτόματο που δέχεται την γλώσσα που τελειώνει σε ένα b.

	a	b
q_1	q_1	q_2
q_2	q_1	q_2

Το παραπάνω αυτόματο μπορούμε να τα εκφράσουμε γραφικά με το εξής σχήμα.

Για μία διμελή σχέση $R \subset A \times A$ ορίζουμε την μεταβατική κλειστότητα (transitive closure) της σχέσης R και την συμβολίζουμε με R^+, το εξής σύνολο:

1. Αν $(a, b) \in R$ τότε $(a, b) \in R^+$,

2. Αν $(a, b) \in R^+$, $(b, c) \in R$ τότε $(a, c) \in R^+$.

Τίποτα άλλο δεν ανήκει στο σύνολο R^+. Παρ. για ένα δένδρο η μεταβατική κλειστότητα της σχέσης αμέσως επόμενος κόμβος περιγράφει τα μονοπάτια από την ρίζα στα φύλλα του δένδρου.

Η ανακλαστική μεταβατική κλειστότητα μίας διμελούς σχέσης $R \subset A \times A$ είναι το σύνολο $R^+ \cup \{(a, a) \mid a \in A\}$ και συμβολίζεται με R^* (όμοια με την κανονική έκφραση *).

Για την σχέση $\vdash_M$ ορίζουμε την $\vdash_M^*$. Για ένα ΠΑ η σχέση $\vdash_M^*$, δηλ. $(q, w) \vdash_M^* (q', w')$ εκφράζει ότι το αυτόματο από

την κατάσταση q διαβάζοντας την συμβολοσειρά w μετά από κάποιο αριθμό βημάτων (ίσως και κανένα βήμα) βρίσκεται στην κατάσταση q', και διαβάζει την συμβολοσειρά w'. Η συμβολοσειρά w' είναι κατάληξη της w για πεπερασμένα αυτόματα.

Ένα μη αιτιοκρατικό (μη ντετερμινιστικό, non deterministic) πεπερασμένο αυτόματο (ΜΠΑ, non deterministic finite automaton) είναι μία πεντάδα $M = (\Sigma, K, \delta, s, F)$, από τα εξής στοιχεία:

1. Σ ένα αλφάβητο,

2. K ένα σύνολο καταστάσεων (states),

3. s μια διακεκριμένη αρχική κατάσταση (start state),

4. $F \subset K$ τις τελικές καταστάσεις (final state) ,

5. Δ μια πεπερασμένη σχέση στο σύνολο $K \times \boldsymbol{\Sigma}^* \times K$, την σχέση μετάβασης (transition relation).

Μπορούμε δηλ. από μία κατάσταση ενώ διαβάζουμε μια συμβολοσειρά να πάμε σε επόμενη κατάσταση διαβάζοντας κανένα, ένα είτε περισσότερα σύμβολα. Όπως θα δούμε τα ΜΠΑ αντιστοιχούν σε ΚΕ (εύκολη απόδειξη), και είναι ισοδύναμα με ΠΑ με την έννοια ότι υπάρχει αλγόριθμος μετατροπής ΜΠΑ σε ΠΑ (πίο δύσκολη απόδειξη). Τελικά τα ΜΠΑ και τα ΠΑ αναγνωρίζουν τις ίδιες γλώσσες. Το πλεονέκτημα των ΜΠΑ είναι ότι περιγράφουμε εύκολα τις κανονικές εκφράσεις με ΜΠΑ χωρίς ουσιαστικά να έχουμε ισχυρότερο μηχανισμό αναγνώρισης γλωσσών Με τα αυτόματα δηλ έχουμε απλό αλγοριθμικό μηχανισμό αναγνώρισης γλωσσών. Ο χαρακτηρισμός απλός σημαίνει ότι έχουμε πεπερασμένη μνήμη τις καταστάσεις του αυτομάτου.

Στο στοιχείο 5 μπορεί να έχουμε την κενή συμβολοσειρά, στην περίπτωση αυτή έχουμε e μετάβαση. Ένα ΠΑ είναι και ΜΠΑ, δεν ισχύει το αντίστροφο.

Μία συνολική κατάσταση (configuration) είναι ένα στοιχείο του $K \times \Sigma^*$. Το σύνολο των συνολικών καταστάσεων ενός αυτομάτου M συμβολίζεται με $\mathrm{Config}\,(M)$.

Η σχέση το αυτόματο παράγει σε ένα βήμα είναι ένα υποσύνολο του $\mathrm{Config}\,(M) \times \mathrm{Config}\,(M)$, και ορίζεται σαν

$$(q, w) \vdash_M \left(q', w'\right)$$

ανν και μόνο αν υπάρχει συμβολοσειρά u ώστε $w = uw'$, και $(q, u, q') \in \Delta$. Αν ισχύει ότι $(q, w) \vdash_M (q', w')$ τότε η (q, w) παράγει σε ένα βήμα την (q', w'). Για ένα μη αιτιοκρατικό αυτόματο η σχέση $\vdash_M$ δεν είναι πάντα συνάρτηση.

Η σχέση το αυτόματο παράγει σε κάποιο αριθμό βημάτων (πιθανόν και κανένα αριθμό βημάτων) είναι η ανακλαστική μεταβατική κλειστότητα της σχέσης $\vdash_M$ και συμβολίζεται με $\vdash_M^*$.

Μια συμβολοσειρά w γίνεται δεκτή από το ΜΠΑ M ανν υπάρχει τελική κατάσταση $q \in F$ ώστε $(s, w) \vdash_M^* (q, e)$. Το σύνολο των συμβολοσειρών που γίνονται δεκτές από ένα

ΜΠΑ ονομάζεται η γλώσσα που αναγνωρίζει το ΜΠΑ M και συμβολίζεται με $L(M)$. Σε ένα ΜΠΑ M είναι δυνατόν να έχουμε $(q,w) \vdash_M^* (q_1,e)$, $(q,w) \vdash_M^* (q_2,e)$ και $q_1 \notin F$, $q_2 \in F$, στην περίπτωση αυτή αρκεί μία από τις καταστάσεις να είναι τελική ώστε το αυτόματο να δεχθεί την συμβολοσειρά.

Δύο αυτόματα M_1, M_2 ονομάζονται ισοδύναμα ανν $L(M_1) = L(M_2)$. Ένα από τα προβλήματα που θα δούμε είναι η αναζήτηση αλγόριθμου ώστε να αποφασίζουμε αν δύο πεπερασμένα αυτόματα είναι ισοδύναμα δηλ. αναγνωρίζουν την ίδια γλώσσα.