Κανονικές Γλώσσες

Κανονικές Εκφράσεις,
Πεπερασμένα Αυτόματα,
Κανονικές Γλώσσες,
Γλώσσες που δεν είναι Κανονικές

Κανονικές Εκφράσεις - Κανονικές Γλώσσες

Πεπερασμένη Αναπαράσταση Γλωσσών

αποδεικνύεται ότι **δεν είναι δυνατόν** να αναπαραστήσουμε με **πεπερασμένο τρόπο** **όλες** τις γλώσσες γιατί

όλες οι πεπερασμένες αναπαραστάσεις αποτελούν **αριθμήσιμο σύνολο**, ενώ

το σύνολο όλων των γλωσσών είναι **μη αριθμήσιμο σύνολο**.

Κανονικές εκφράσεις (***regular expressions***)

Κανονικές Εκφράσεις - Κανονικές Γλώσσες

Κανονικές Εκφράσεις

Οι κανονικές εκφράσεις στο **αλφάβητο** Σ είναι **όλες** οι ακολουθίες συμβόλων από το $\Sigma \cup \{$ **(,), *,** $\cup$**,** $\varnothing$ $\}$ όπου:

1a. Η $\varnothing$ είναι Κ.Ε. που αντιστοιχεί στην **κενή γλώσσα** $\varnothing$.

1b. $\forall$ **α** $\in \Sigma$, η **α** είναι Κ.Ε. που αντιστοιχεί στη **γλώσσα** $\{$ **α** $\}$.

Έστω ***r***, ***s*** Κ.Ε. που αντιστοιχούν στις γλώσσες L_r, L_s.

2. Η $(r \cup s)$ είναι Κ.Ε. που αντιστοιχεί στη γλώσσα $L_r \cup L_s$.
3. Η $(r\, s)$ είναι Κ.Ε. που αντιστοιχεί στη γλώσσα $L_r\, L_s$.
4. Η r^* είναι Κ.Ε. που αντιστοιχεί στη γλώσσα $L_r{}^*$.
5. Μόνον ό,τι προκύπτει από τους κανόνες 1-4 είναι Κ.Ε.

Σημείωση*:** η ***e περιγράφεται από $(\varnothing^*)$

Σε κάποια συγγράμματα χρησιμοποιείται το **+** (αντί για $\cup$) καθώς και το ***e*** .

Κανονικές Εκφράσεις - Κανονικές Γλώσσες

Σημείωση

r* **:** **οσεσδήποτε** **παραθέσεις** συμβολοσειρών που **περιγράφονται από την κανονική έκφραση** ***r***

Παραδείγματα (Σ = {0, 1})

0 περιγράφει τη **γλώσσα { 0 }**

(0 ∪ 1) περιγράφει τη **γλώσσα { 0, 1 }**

(01) περιγράφει τη **γλώσσα { 01 }**

(((0 ∪ 1) 0) (0 ∪ 1)) περιγράφει **όλες** τις συμβολοσειρές **μήκους 3** με **μεσαίο σύμβολο το 0**

0* περιγράφει **όλες** τις συμβολοσειρές **από 0 και** την ***e***

Κανονικές Εκφράσεις - Κανονικές Γλώσσες

Παραδείγματα (***συνέχεια***)

$(0 \cup 1)^*$ **όλες** οι συμβολοσειρές **από 0, 1 και** η ***e***

$(0^* \cup 1^*)$ **όλες** οι συμβολοσειρές με **μόνο 0 ή μόνο 1 και** η ***e***

$(0^* 1^*)$ **όλες** οι συμβολοσειρές με **μηδέν, ένα,… 0 ακολουθούμενα** από **μηδέν, ένα, … 1**

$((0 \cup 1)^* 1)$ **όλες** οι συμβολοσειρές που **τελειώνουν σε 1**

$(((0 \cup 1)^* (01)) (0 \cup 1)^*)$ **όλες** οι συμβολοσειρές που **περιέχουν 01**

Περιγράφει η Κ.Ε. $(1 \cup (10))^*$ τη συμβολοσειρά **11010110011**;

11010110~~0~~11 ΟΧΙ

Κανονικές Εκφράσεις - Κανονικές Γλώσσες

Σημειώστε:

$(r\,\varnothing) = (\varnothing\, r) = \varnothing$ $\quad$ $(r \cup \varnothing) = (\varnothing \cup r) = r$

$(r\,\varnothing^*) = (\varnothing^*\, r) = r$

Όμως, δεν ισχύει πάντα ότι $(r \cup \varnothing^*) = r$ $\quad$ $(r \cup \varnothing^*) = (\varnothing^* \cup r)$

Προτεραιότητα τελεστών

$*$ $>$ **παράθεση** $>$ $\cup$

π.χ., $0 \cup 10 = 0 \cup (10)$, $\quad 0 \cup 10^* = 0 \cup (1(0^*))$

Επίσης, για την **παράθεση** ισχύει

$((r_1\, r_2)\, r_3) = r_1\, r_2\, r_3 = (r_1\, (r_2\, r_3))$

π.χ., $(((0 \cup 1)^*\, (01))\, (0 \cup 1)^*) = (0 \cup 1)^*\, 01\, (0 \cup 1)^*$

Κανονικές Εκφράσεις - Κανονικές Γλώσσες

Μια **γλώσσα** σε ένα αλφάβητο **Σ** είναι **κανονική** εάν υπάρχει **κανονική έκφραση στο Σ** που την **περιγράφει.**

Με τις **κανονικές εκφράσεις** μπορούμε να **παραγάγουμε** τις κανονικές γλώσσες.

Πρόταση

Οι **κανονικές γλώσσες** είναι **κλειστές** ως προς τις πράξεις της **ένωσης**, **παράθεσης** και **Kleene star**.

Δηλαδή, για **οποιεσδήποτε** κανονικές γλώσσες L_1, L_2:

$L_1 \cup L_2$ είναι κανονική γλώσσα

$L_1 L_2$ είναι κανονική γλώσσα

$L_1{}^*$ είναι κανονική γλώσσα.

Κανονικές Εκφράσεις - Κανονικές Γλώσσες

Παρατηρήσεις

1. **Κάθε πεπερασμένη** γλώσσα L είναι **κανονική γλώσσα**.
   Μια Κ.Ε. σχηματίζεται από τις Κ.Ε. των συμβολοσειρών της L βάζοντας ανάμεσά τους το $\cup$.

2. Για να δείξουμε ότι μια γλώσσα L περιγράφεται από μια Κ.Ε. r πρέπει να δείξουμε:
   - **α.** ότι **κάθε** συμβολοσειρά της L **περιγράφεται** από την r **και**
   - **β.** ότι **κάθε** συμβολοσειρά που περιγράφεται από την r **ανήκει** στην L.

   π.χ., $L = \{ w \in \{0,1\}^* \mid$ πλήθος 1 στην w είναι άρτιος αριθμός $\}$

   $r = 0^* (1\ 0^*\ 1\ 0^*)^*$

Τι λέτε για την $r' = (0^*\ 1\ 0^*\ 1\ 0^*)^*$;

Κανονικές Εκφράσεις - Κανονικές Γλώσσες

Εύρεση κανονικής έκφρασης για δοθείσα γλώσσα

Θεωρούμε μια **τυπική συμβολοσειρά** της γλώσσας,
βλέπουμε πως αυτή **δομείται** λόγω της περιγραφής της γλώσσας,
βρίσκουμε **επιμέρους κανονικές εκφράσεις** και
τις **συνθέτουμε**.

π.χ., $\boldsymbol{L} = \{ w \in \{0,1\}^* \mid$ η w **αρχίζει με 0** $\} \Rightarrow \boldsymbol{r} = \mathbf{0\,(0 \cup 1)^*}$

***Προσοχή*:** Εάν η γλώσσα περιγράφεται από μια **αρνητική** πρόταση, πρέπει να βρούμε μια **ισοδύναμη καταφατική περιγραφή**.

π.χ., $\boldsymbol{L} = \{ w \in \{0,1\}^* \mid$ η w **δεν αρχίζει με 1** $\}$

Ισοδύναμα: $\boldsymbol{L} = \{ w \in \{0,1\}^* \mid$ η w **αρχίζει με 0** $\} \cup \{ \boldsymbol{e} \}$

$\Rightarrow \boldsymbol{r} = \mathbf{0\,(0 \cup 1)^* \cup \emptyset^*}$

Πεπερασμένα Αυτόματα

Αναγνώριση γλώσσας *L* από μηχανή *M*

Εάν μια **συμβολοσειρά *x*** δοθεί ως **είσοδος** στη μηχανή *M*, η *M* διαβάζει τα σύμβολα της ***x***, **ένα-ένα** από **αριστερά προς τα δεξιά** και στο τέλος **αποφασίζει** αν $x \in L$.

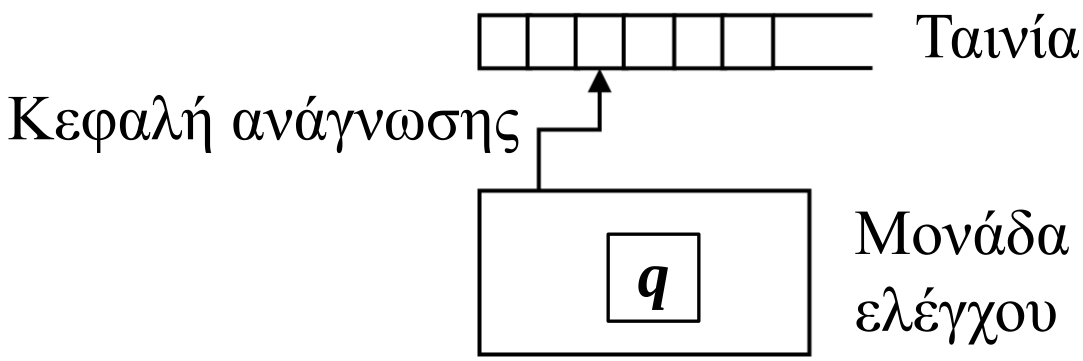
Πεπερασμένα αυτόματα ή **μηχανές πεπερασμένων καταστάσεων** (***finite automata***, ***finite state machines***)

οι **απλούστερες** μηχανές

αναγνωρίζουν τις **κανονικές γλώσσες**

***Εφαρμογές*:** **αυτόματος πωλητής** αναψυκτικών
αναζήτηση λέξης σε κείμενο
αναγνώριση λέξης-κλειδιού σε πρόγραμμα

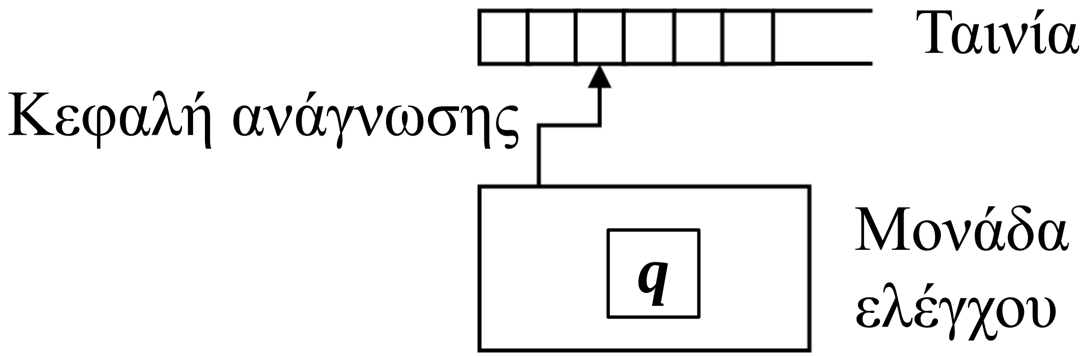
Πεπερασμένα Αυτόματα



Μέρη πεπερασμένου αυτομάτου

- Η **ταινία** εκτείνεται στο **άπειρο προς τα δεξιά** και είναι χωρισμένη σε τετράγωνα με **1 σύμβολο σε κάθε τετράγωνο.**
- Η **κεφαλή ανάγνωσης** διαβάζει **ένα σύμβολο τη φορά** από την ταινία και κινείται **μόνο προς τα δεξιά.**
- Η **μονάδα ελέγχου** αποθηκεύει την **τρέχουσα κατάσταση** του αυτόματου και **από την τρέχουσα κατάσταση και το σύμβολο που διαβάζεται ορίζει τη νέα κατάσταση** του αυτομάτου.

Πεπερασμένα Αυτόματα



Αρχικά:

- Η **ταινία** αποθηκεύει την **είσοδο *w* στα |*w*| αριστερότερα τετράγωνά** της.
- Η **κεφαλή ανάγνωσης** διαβάζει το **σύμβολο στο αριστερότερο τετράγωνο** της ταινίας.
- Η **μονάδα ελέγχου** δηλώνει ότι το αυτόματο βρίσκεται στην **αρχική του κατάσταση**.

Αιτιοκρατικά Πεπερασμένα Αυτόματα

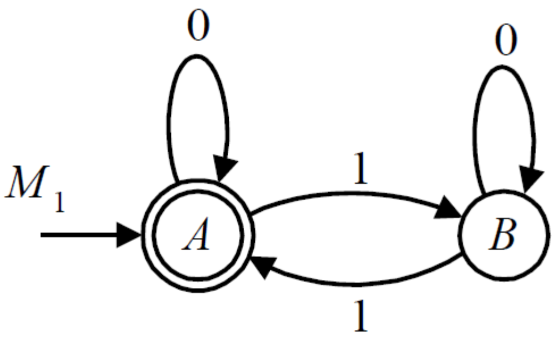
Αιτιοκρατικό (Ντετερμινιστικό) Πεπερασμένο Αυτόματο

μια 5-άδα **(Q, Σ, q_0, δ, F)** όπου

- Q : ένα **πεπερασμένο** **σύνολο καταστάσεων**,
- Σ : το **αλφάβητο** των συμβόλων **εισόδου**,
- $q_0 \in Q$: η **αρχική κατάσταση**,
- $\delta : Q \times \Sigma \rightarrow Q$: η **συνάρτηση μετάβασης** και
- $F \subseteq Q$: το σύνολο των **τελικών καταστάσεων**.

Παράδειγμα: $M_1 = (Q, \Sigma, q_0, \delta, F)$

$Q = \{ A, B \}$	$\delta(A,0) = A$
$\Sigma = \{ 0, 1 \}$	$\delta(A,1) = B$
$q_0 = A$	$\delta(B,0) = B$
$F = \{ A \}$	$\delta(B,1) = A$



Διάγραμμα μεταβάσεων

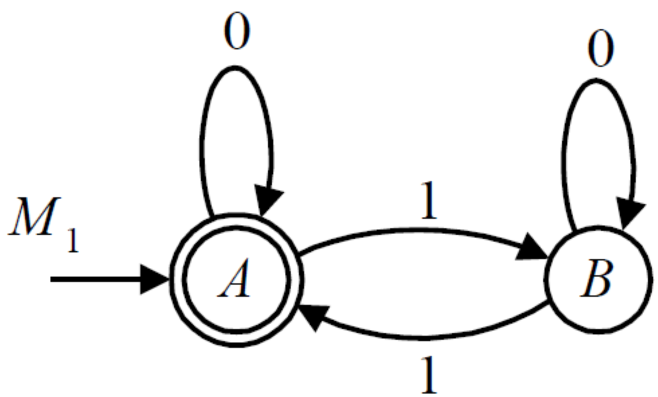
Αναγνώριση/Αποδοχή Γλώσσας

Έστω Αιτ. Π. Α. $\boldsymbol{M = (Q, \Sigma, q_0, \delta, F)}$.

Μια **συνολική κατάσταση** είναι ένα στοιχείο $(\boldsymbol{q}, \boldsymbol{w})$ του $\boldsymbol{Q \times \Sigma^*}$
όπου $\boldsymbol{q}$ είναι η **τρέχουσα κατάσταση** του αυτόματου M και
$\boldsymbol{w}$ είναι το τμήμα της **εισόδου που δεν έχει διαβαστεί**.

Μια συνολική κατάσταση $\boldsymbol{(q, ax)}$ όπου $\boldsymbol{q \in Q}$, $\boldsymbol{a \in \Sigma}$, $\boldsymbol{x \in \Sigma^*}$ **παράγει** (**δίνει**) **σε ένα βήμα** τη συνολική κατάσταση $\boldsymbol{(r, x)}$ αν $\boldsymbol{\delta(q, a) = r}$. Γράφουμε: $\boldsymbol{(q, ax) \vdash (r, x)}$

Αναγνώριση/Αποδοχή Γλώσσας



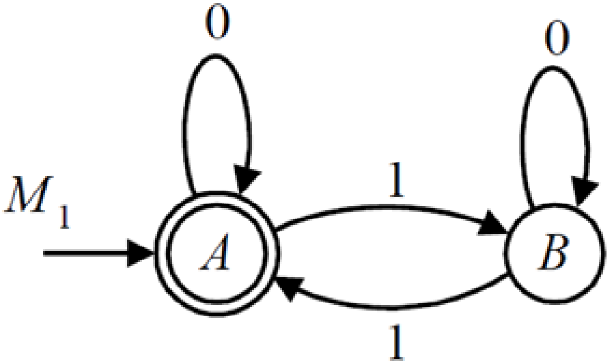
$$
\begin{aligned}
(A, 00101) &\vdash (A, 0101)\\
&\vdash (A, 101)\\
&\vdash (B, 01)\\
&\vdash (B, 1)\\
&\vdash (A, e)
\end{aligned}
$$

Αναγνώριση/Αποδοχή Γλώσσας

Έστω $\vdash^*$ η **ανακλαστική & μεταβατική κλειστότητα** της $\vdash$.
Γράφουμε $(\boldsymbol{q}, \boldsymbol{w}) \vdash^* (\boldsymbol{s}, \boldsymbol{w'})$ όταν η συνολική κατάσταση $(\boldsymbol{q}, \boldsymbol{w})$ **παράγει** (***σε 0,1,… βήματα***) την $(\boldsymbol{s}, \boldsymbol{w'})$.

Ένα Αιτ. Π. Α. $\boldsymbol{M = (Q, \Sigma, q_0, \delta, F)}$ **αναγνωρίζει** ή **αποδέχεται** μια συμβολοσειρά $\boldsymbol{x} \in \boldsymbol{\Sigma^*}$ αν $(\boldsymbol{q_0}, \boldsymbol{x}) \vdash^* (\boldsymbol{q}, \boldsymbol{e})$ **και** $\boldsymbol{q} \in \boldsymbol{F}$.

$$
\begin{aligned}
(A, \mathbf{00101}) &\vdash (A, \mathbf{0101}) \\
&\vdash (A, \mathbf{101}) \\
&\vdash (B, \mathbf{01}) \\
&\vdash (B, \mathbf{1}) \\
&\vdash (A, e)
\end{aligned}
$$



Το Αιτ.Π.Α. M_1 **αποδέχεται** τη συμβολοσειρά **00101**.

Αναγνώριση/Αποδοχή Γλώσσας

Έστω $\vdash^*$ η **ανακλαστική & μεταβατική κλειστότητα** της $\vdash$. Γράφουμε $(q, w) \vdash^* (s, w')$ όταν η συνολική κατάσταση (q, w) **παράγει** (***σε 0,1,… βήματα***) την (s, w').

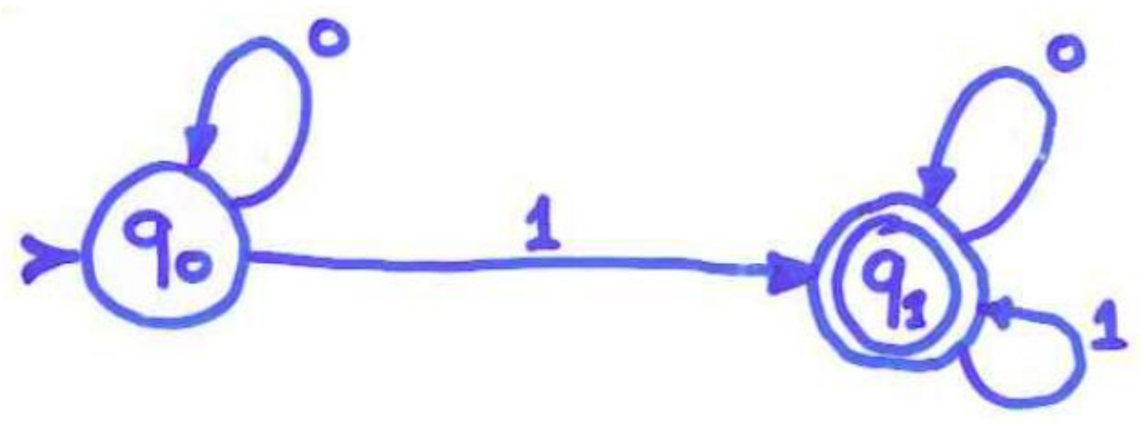
Ένα Αιτ. Π. Α. $M = (Q, \Sigma, q_0, \delta, F)$ **αναγνωρίζει** ή **αποδέχεται** μια συμβολοσειρά $x \in \Sigma^*$ αν $(q_0, x) \vdash^* (q, e)$ **και** $q \in F$.

Γλώσσα $L(M)$ που **αναγνωρίζει** ή **αποδέχεται** το M
= το σύνολο των συμβολοσειρών που **αποδέχεται** το M

Προσοχή*:** ένα αυτόματο που **αναγνωρίζει μια γλώσσα *L πρέπει να **αποδέχεται όλες** τις συμβολοσειρές που **ανήκουν στην *L*** **και** να **μην αποδέχεται καμία από τις υπόλοιπες**.

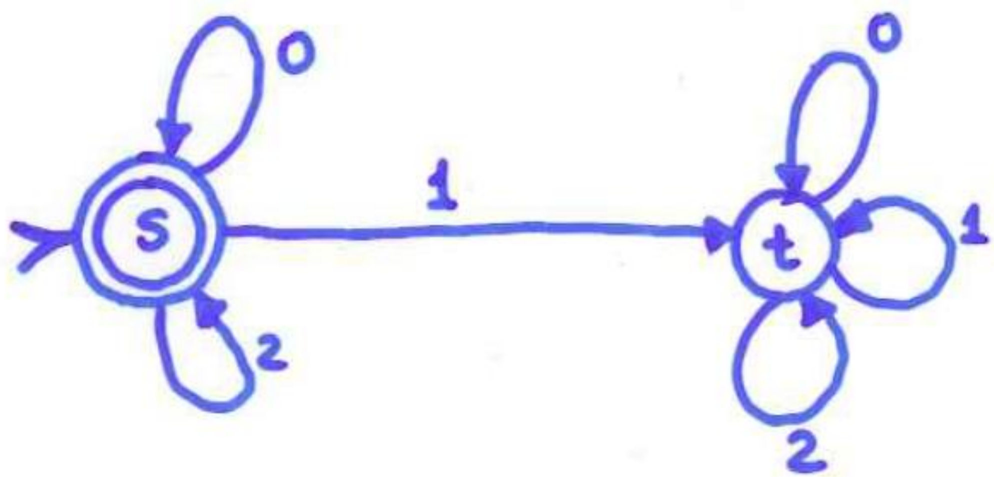
Αναγνώριση/Αποδοχή Γλώσσας

Αιτιοκρατικό Πεπερασμένο Αυτόματο που αναγνωρίζει τη γλώσσα των συμβολοσειρών από το αλφάβητο {0,1} που **περιέχουν (τουλάχιστον ένα) 1**.



Αναγνώριση/Αποδοχή Γλώσσας

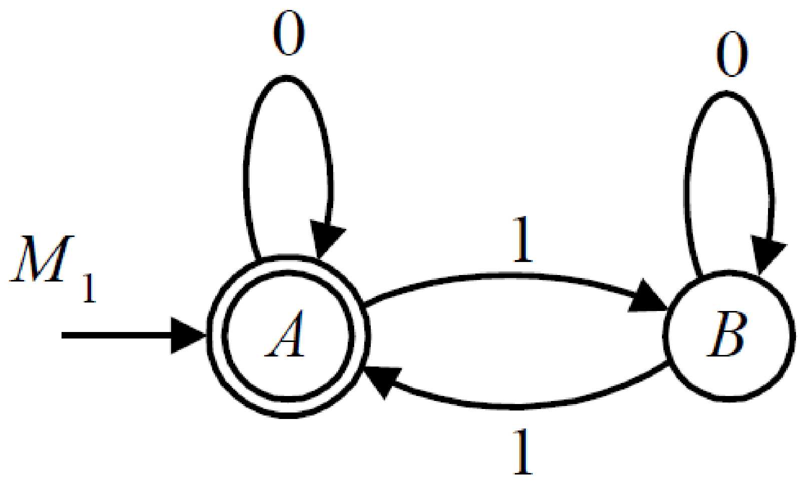
Αιτιοκρατικό Πεπερασμένο Αυτόματο που αναγνωρίζει τη γλώσσα των συμβολοσειρών από το αλφάβητο {0,1,2} που **δεν περιέχουν 1**.



t : κατάσταση **καταβόθρα** ή κατάσταση **αδιεξόδου** (*εάν το αυτόματο βρεθεί σε αυτήν την κατάσταση, μένει σε αυτήν* ***για πάντα***)

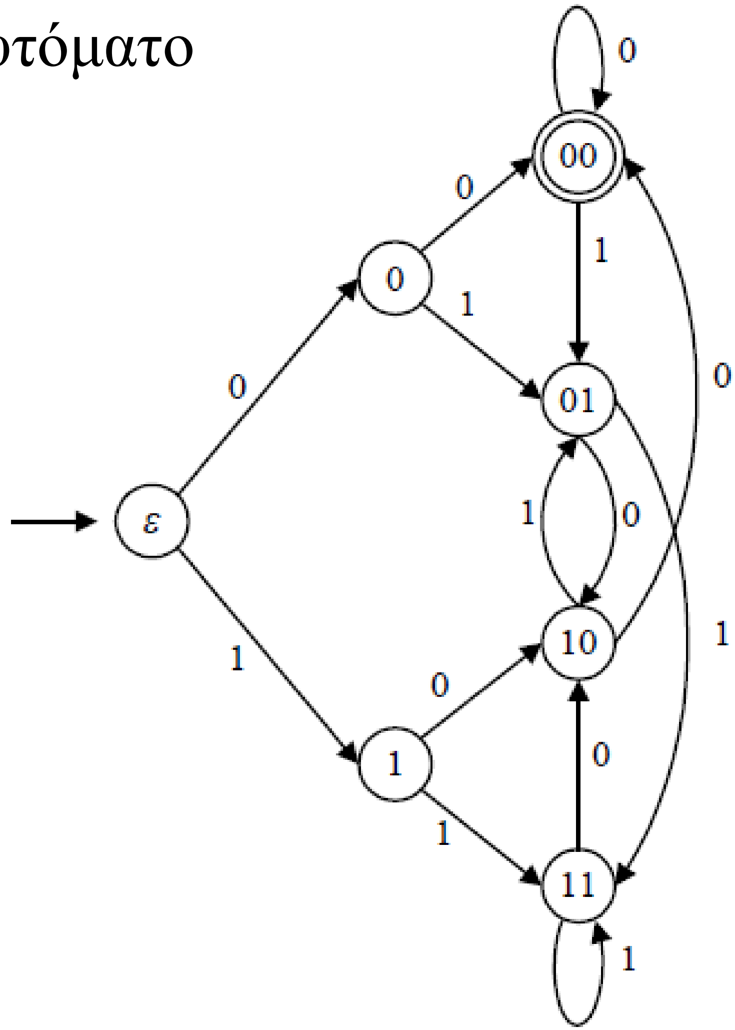
Αναγνώριση/Αποδοχή Γλώσσας

Αιτιοκρατικό Πεπερασμένο Αυτόματο που αναγνωρίζει τη γλώσσα των συμβολοσειρών από το αλφάβητο {0,1} που **περιέχουν άρτιο πλήθος 1**.

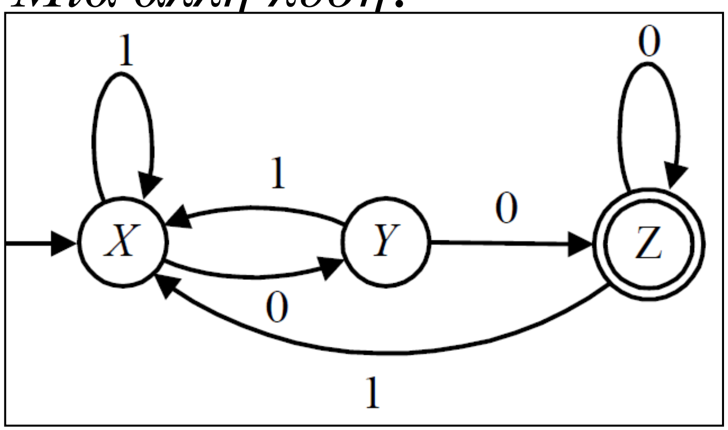


Αναγνώριση/Αποδοχή Γλώσσας

Αιτιοκρατικό Πεπερασμένο Αυτόματο
που αναγνωρίζει τη γλώσσα
των συμβολοσειρών
από το αλφάβητο {0,1}
που **τελειώνουν σε 00**.



Μια άλλη λύση:

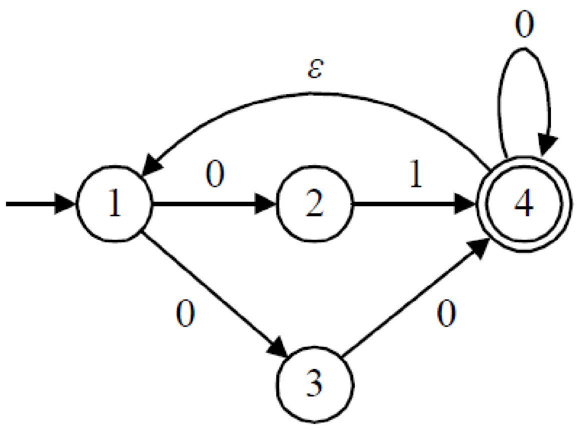


Μη Αιτιοκρατικά Πεπερασμένα Αυτόματα

Στα **αιτιοκρατικά** Π.Α. έχουμε **ακριβώς 1** μετάβαση από **κάθε** κατάσταση για **κάθε** σύμβολο του αλφαβήτου.

Στα **μη αιτιοκρατικά** Π.Α. επιτρέπεται να έχουμε:

- **καμία, μία ή περισσότερες** μεταβάσεις για κάποιο σύμβολο από κάποια κατάσταση
- ***μεταβάσεις για e*** (= μεταβάσεις **χωρίς ανάγνωση συμβόλου**)

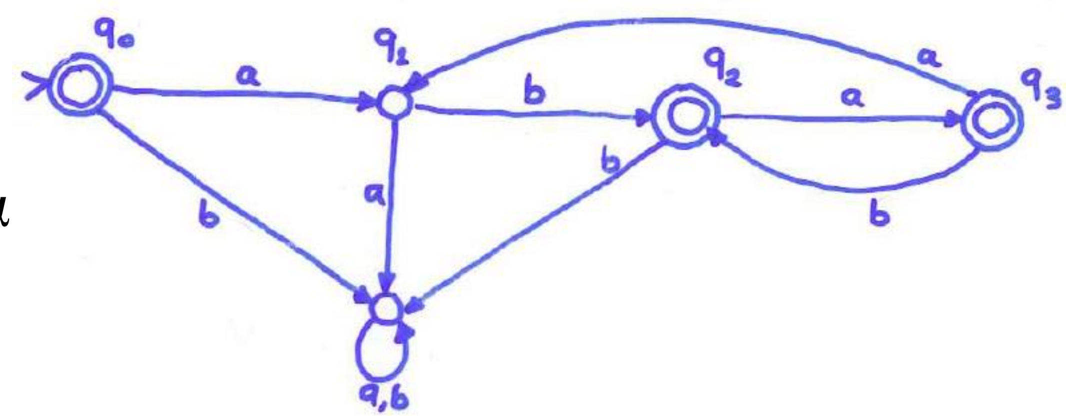


Μη Αιτιοκρατικά Πεπερασμένα Αυτόματα

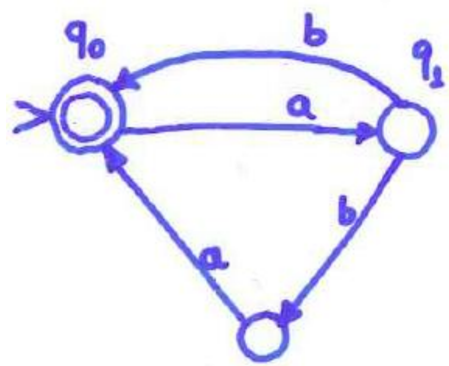
Γιατί;

- Θεωρήματα, κατασκευές
- Απλή περιγραφή

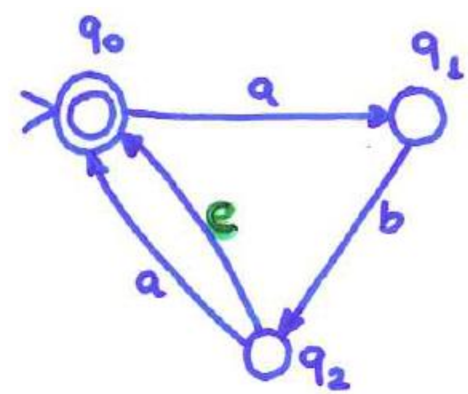
***Παράδειγμα*:**
Αιτ.Π.Α. για τη γλώσσα
με Κ.Ε. **(ab ∪ aba)***



Μη Αιτ.Π.Α.
για τη γλώσσα
με Κ.Ε. **(ab ∪ aba)***



ή



Μη Αιτιοκρατικά Πεπερασμένα Αυτόματα

Μη Αιτιοκρατικό Πεπερασμένο Αυτόματο

μια 5-άδα **($Q, \Sigma, q_0, \Delta, F$)** όπου

Q : ένα **πεπερασμένο σύνολο καταστάσεων**,

Σ : το **αλφάβητο** των συμβόλων **εισόδου**,

$q_0 \in Q$: η **αρχική κατάσταση**,

$\Delta \subseteq \boxed{Q \times \{\Sigma \cup \{e\}\} \times Q}$: η **σχέση μετάβασης** και

$F \subseteq Q$: το σύνολο των **τελικών καταστάσεων**.

Αιτιοκρατικό (Ντετερμινιστικό) Πεπερασμένο Αυτόματο

μια 5-άδα ($Q, \Sigma, q_0, \delta, F$) όπου

Q : ένα **πεπερασμένο σύνολο καταστάσεων**,

Σ : το **αλφάβητο** των συμβόλων **εισόδου**,

$q_0 \in Q$: η **αρχική κατάσταση**,

$\delta : Q \times \Sigma \rightarrow Q$: η **συνάρτηση μετάβασης** και

$F \subseteq Q$: το σύνολο των **τελικών καταστάσεων**.

Μη Αιτιοκρατικά Πεπερασμένα Αυτόματα

Μη Αιτιοκρατικό Πεπερασμένο Αυτόματο

μια 5-άδα $(Q, \Sigma, q_0, \Delta, F)$ όπου

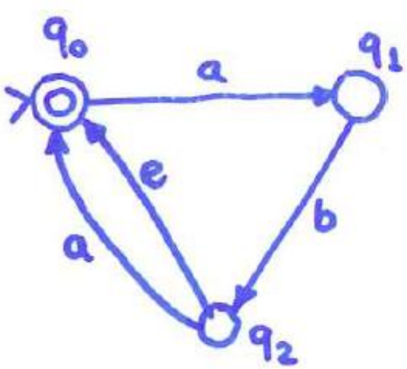
Q : ένα **πεπερασμένο** **σύνολο καταστάσεων**,

Σ : το **αλφάβητο** των συμβόλων **εισόδου**,

$q_0 \in Q$: η **αρχική κατάσταση**,

$\Delta \subseteq \boxed{Q \times \{\Sigma \cup \{e\}\} \times Q}$: η **σχέση μετάβασης** και

$F \subseteq Q$: το σύνολο των **τελικών καταστάσεων**.



$K = \{q_0, q_1, q_2\}$

$\Sigma = \{a, b\}$

$s = q_0$

$F = \{q_0\}$

$\Delta = \{ (q_0, a, q_1), (q_1, b, q_2), (q_2, a, q_0), (q_2, e, q_0) \}$

Μη Αιτιοκρατικά Πεπερασμένα Αυτόματα

Εναλλακτικός Ορισμός

Μη Αιτιοκρατικό Πεπερασμένο Αυτόματο

μια 5-άδα $(Q, \Sigma, q_0, \delta, F)$ όπου

Q : ένα **πεπερασμένο σύνολο καταστάσεων**,

Σ : το **αλφάβητο** των συμβόλων **εισόδου**,

$q_0 \in Q$: η **αρχική κατάσταση**,

$\delta : \boxed{Q \times \{\Sigma \cup \{e\}\} \rightarrow 2^Q}$: η **συνάρτηση μετάβασης** και

$F \subseteq Q$: το σύνολο των **τελικών καταστάσεων**.

Αιτιοκρατικό (Ντετερμινιστικό) Πεπερασμένο Αυτόματο

μια 5-άδα $(Q, \Sigma, q_0, \delta, F)$ όπου

Q : ένα **πεπερασμένο σύνολο καταστάσεων**,

Σ : το **αλφάβητο** των συμβόλων **εισόδου**,

$q_0 \in Q$: η **αρχική κατάσταση**,

$\delta : Q \times \Sigma \rightarrow Q$: η **συνάρτηση μετάβασης** και

$F \subseteq Q$: το σύνολο των **τελικών καταστάσεων**.

Αναγνώριση/Αποδοχή Γλώσσας

Έστω μη Αιτ. Π. Α. $\boldsymbol{M = (Q, \Sigma, q_0, \Delta, F)}$.

Μια **συνολική κατάσταση** είναι ένα στοιχείο $(\boldsymbol{q}, \boldsymbol{w})$ του $\boldsymbol{Q \times \Sigma^*}$
όπου $\boldsymbol{q}$ είναι η **τρέχουσα κατάσταση** του αυτόματου M και
$\boldsymbol{w}$ είναι το τμήμα της **εισόδου που δεν έχει διαβαστεί**.

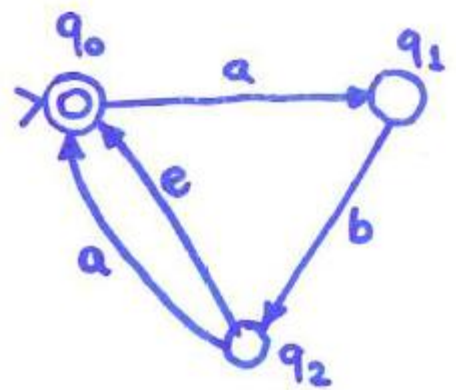
Μια συνολική κατάσταση $(\boldsymbol{q}, \boldsymbol{w})$ **παράγει** (**δίνει**) **σε ένα βήμα** τη συνολική κατάσταση $(\boldsymbol{r}, \boldsymbol{w'})$ αν υπάρχει $\boldsymbol{u} \in \boldsymbol{\Sigma} \cup \{\boldsymbol{e}\}$: $\boldsymbol{w} = \boldsymbol{uw'}$ και $(\boldsymbol{q}, \boldsymbol{u}, \boldsymbol{r}) \in \boldsymbol{\Delta}$. Γράφουμε: $\boldsymbol{(q, w) \vdash (r, w')}$

Έστω $\vdash^*$ η **ανακλαστική & μεταβατική κλειστότητα** της $\vdash$.
Γράφουμε $(\boldsymbol{q}, \boldsymbol{w}) \vdash^* (\boldsymbol{s}, \boldsymbol{w'})$ όταν η συνολική κατάσταση $(\boldsymbol{q}, \boldsymbol{w})$ **παράγει** (***σε 0,1,… βήματα***) την $(\boldsymbol{s}, \boldsymbol{w'})$.

Αναγνώριση/Αποδοχή Γλώσσας

Ένα μη Αιτ. Π. Α. $\boldsymbol{M} = (\boldsymbol{Q}, \boldsymbol{\Sigma}, \boldsymbol{q_0}, \boldsymbol{\Delta}, \boldsymbol{F})$ **αποδέχεται** ή **αναγνωρίζει** μια συμβολοσειρά $\boldsymbol{x} \in \boldsymbol{\Sigma}^*$ αν $(\boldsymbol{q_0}, \boldsymbol{x}) \vdash^* (\boldsymbol{q}, \boldsymbol{e})$ **και** $\boldsymbol{q} \in \boldsymbol{F}$. Δηλαδή, **αρκεί** να υπάρχει **μία** ακολουθία συνολικών καταστάσεων που οδηγεί από την $(\boldsymbol{q_0}, \boldsymbol{x})$ σε $(\boldsymbol{q}, \boldsymbol{e})$ με $\boldsymbol{q} \in \boldsymbol{F}$.

Π.χ., αποδέχεται το μη Αιτ. Π.Α.
του σχήματος τη συμβολοσειρά
abab ;



Γλώσσα $\boldsymbol{L(M)}$ που **αναγνωρίζει** ή **αποδέχεται** το M
= το σύνολο των συμβολοσειρών που **αποδέχεται** το M