ΕΠΛ 211:
Θεωρία Υπολογισμού και Πολυπλοκότητας

Διάλεξη 5: Κανονικές Εκφράσεις

Τι θα κάνουμε σήμερα…

- Κλειστότητα Κανονικών Πράξεων (1.2.3)
- Εισαγωγή στις Κανονικές Εκφράσεις
- Τυπικός ορισμός της κανονικής έκφρασης (1.3.1)
- Ισοδυναμία με πεπερασμένα αυτόματα (1.3.2)

Πράξεις σε Κανονικές Γλώσσες

- Έστω δυο γλώσσες A και B:

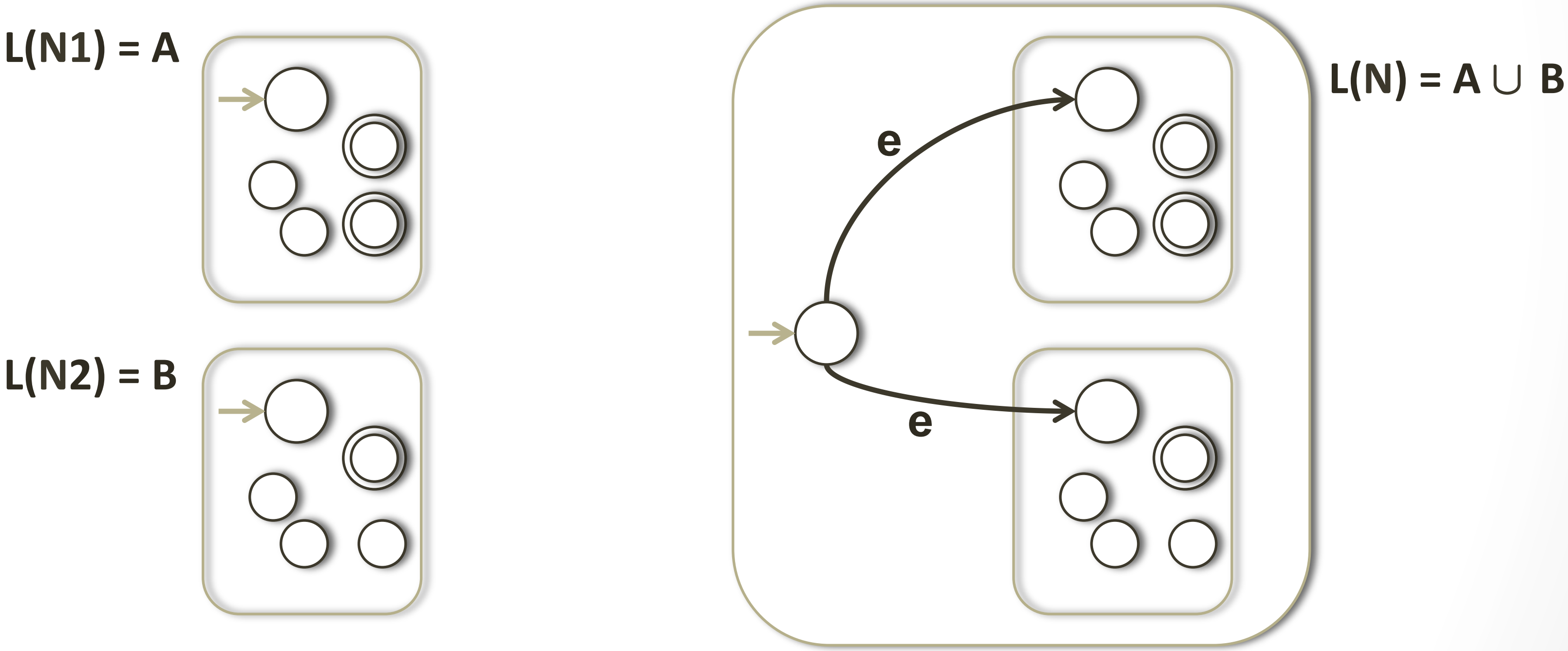
- **Ένωση**: $A \cup B = \{x | x \in A \text{ or } x \in B\}$

- **Συναρμογή (Σύμπτυξη)**: $AB = \{xy | x \in A \text{ and } y \in B\}$

- **Σώρευση**: $A^* = \{x_1 x_2 \ldots x_k | k \geq 0 \text{ and } \forall x_i, x_i \in A\}$

Κλειστότητα ως προς την ένωση

Θεώρημα:
Το σύνολο των κανονικών γλωσσών είναι κλειστό ως προς την ένωση.



- Για αυθαίρετες κανονικές γλώσσες A, B πρέπει να δείξουμε ότι A∪B είναι επίσης κανονική

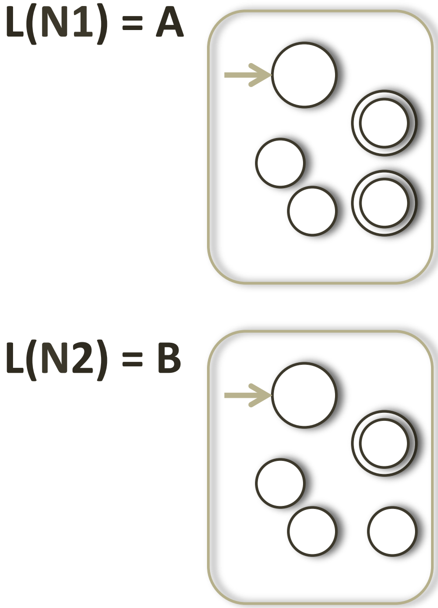
Κλειστότητα ως προς την ένωση

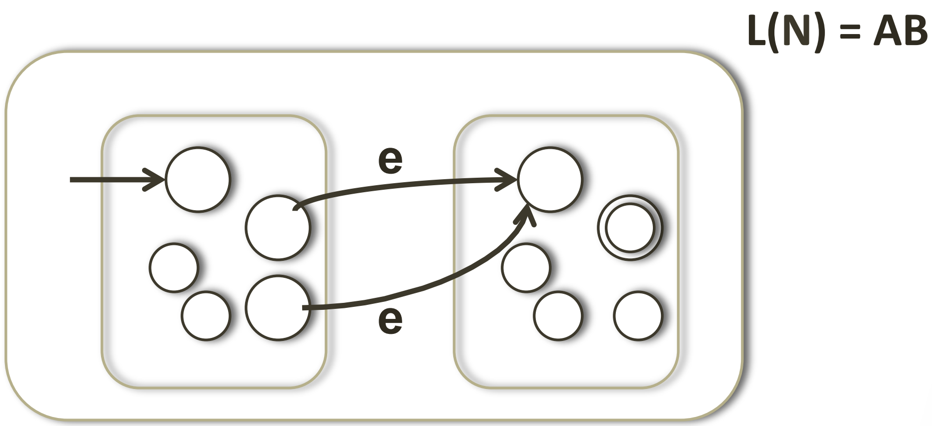
- Έστω ότι έχουμε τα αυτόματα:
  - $N_1 = (Q_1, \Sigma, \delta_1, q_1, F_1), L(N_1) = A$
  - $N_2 = (Q_2, \Sigma, \delta_2, q_2, F_2), L(N_2) = B$
- Ο τυπικός ορισμός του $N = (Q, \Sigma, \delta, q_0, F), L(N) = A \cup B$:
  - Καταστάσεις του N: $Q = \{q_0\} \cup Q_1 \cup Q_2$
  - Εναρκτήρια κατάσταση: q_0
  - Καταστάσεις αποδοχής: $F = F_1 \cup F_2$
  - Μεταβάσεις:

$$\delta(q,a) = \begin{cases} \delta_1(q,a) & \text{if } q \in Q_1 \\ \delta_2(q,a) & \text{if } q \in Q_2 \\ \{q_1, q_2\} & \text{if } q = q_0 \text{ and } a = e \\ \emptyset & \text{if } q = q_0 \text{ and } a \neq e \end{cases}$$

Κλειστότητα ως προς την συναρμογή

Θεώρημα:
Το σύνολο των κανονικών γλωσσών είναι κλειστό ως προς την συναρμογή.





- Για αυθαίρετες κανονικές γλώσσες A, B πρέπει να δείξουμε ότι AB είναι επίσης κανονική

Κλειστότητα ως προς την συναρμογή

- Έστω ότι έχουμε τα αυτόματα:
  - $N_1 = (Q_1, \Sigma, \delta_1, q_1, F_1), L(N_1) = A$
  - $N_2 = (Q_2, \Sigma, \delta_2, q_2, F_2), L(N_2) = B$
- Ο τυπικός ορισμός του $N = (Q, \Sigma, \delta, q_0, F), L(N) = AB$:
  - Καταστάσεις του N: $Q = Q_1 \cup Q_2$
  - Εναρκτήρια κατάσταση: $q_0 = q_1$
  - Καταστάσεις αποδοχής: $F = F_2$
  - Μεταβάσεις:

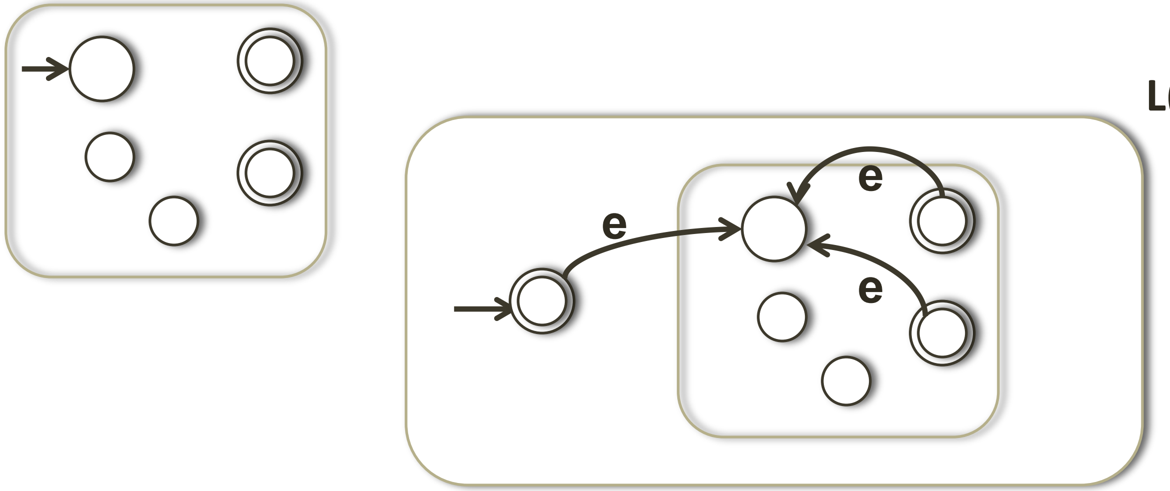
$$\delta(q,a) = \begin{cases} \delta_1(q,a) & \text{if } q \in Q_1 \text{ and } a \notin F_1 \\ \delta_1(q,a) & \text{if } q \in F_1 \text{ and } a \neq e \\ \delta_1(q,a) \cup \{q_2\} & \text{if } q \in F_1 \text{ and } a = e \\ \delta_2(q,a) & \text{if } q \in Q_2 \end{cases}$$

Κλειστότητα ως προς την σώρευση

Θεώρημα:
Το σύνολο των κανονικών γλωσσών είναι κλειστό ως προς την σώρευση.

L(N1) = A

L(N) = A*



- Για αυθαίρετη κανονική γλώσσα Α πρέπει να δείξουμε ότι Α* είναι επίσης κανονική

Κλειστότητα ως προς την συναρμογή

- Έστω ότι έχουμε τα αυτόματα:
  - $N_1 = (Q_1, \Sigma, \delta_1, q_1, F_1), L(N_1) = A$
- Ο τυπικός ορισμός του $N = (Q, \Sigma, \delta, q_0, F), L(N) = A^*$:
  - Καταστάσεις του N: $Q = Q_1 \cup \{q_0\}$
  - Εναρκτήρια κατάσταση: q_0
  - Καταστάσεις αποδοχής: $F = \{q_0\} \cup F_1$
  - Μεταβάσεις:

$$\delta(q,a) = \begin{cases} \delta_1(q,a) & \text{if } q \in Q_1 \text{ and } a \notin F_1 \\ \delta_1(q,a) & \text{if } q \in F_1 \text{ and } a \neq e \\ \delta_1(q,a) \cup \{q_1\} & \text{if } q \in F_1 \text{ and } a = e \\ \{q_1\} & \text{if } q = q_0 \text{ and } a = e \\ \emptyset & \text{if } q = q_0 \text{ and } a \neq e \end{cases}$$

Κανονικές εκφράσεις

- Οι κανονικές πράξεις μας επιτρέπουν να συντάσσουμε εκφράσεις που αναπαριστούν γλώσσες. Οι εκφράσεις αυτές λέγονται **κανονικές**.
  - π.χ. $(0 \cup 1)0^*$

- Ποια γλώσσα παράγεται;
  - Τα σύμβολα 0 και 1 είναι συντομογραφίες των {0} και {1}
  - Άρα $(0 \cup 1) = (\{0\} \cup \{1\})=\{0,1\}=L1$
  - Το 0^* σημαίνει $\{0\}^*=L2$
  - Η έκφραση $(0 \cup 1)0^*$ είναι συντομογραφία της συναρμογής L1L2
  - Έτσι L=L1L2= όλες οι λέξεις που ξεκινούν με 0 ή 1 ακολουθούμενες απο οποιοδήποτε πλήθος από 0.

Κανονικές εκφράσεις

- Παρατηρείστε
  - $(0 \cup 1)^* = \Sigma^*$
- Τι γλώσσα παράγεται από τις εκφράσεις:
  - Σ^*1
  - $(0\Sigma^*) \cup (\Sigma^*1)$
- Σειρά εκτέλεσης (εκτός και αν καθορίζεται από παρενθέσεις)
  - Σώρευση
  - Συναρμογή
  - Ένωση

Τυπικός Ορισμός Κανονικής Έκφρασης

Ορισμός

R είναι μια κανονική έκφραση αν είναι της μορφής

1. ***α, όπου α είναι ένα σύμβολο του αλφαβήτου Σ***
2. ***e,***
3. $\emptyset$
4. $(R_1 \cup R_2)$ ***,όπου R_1 και R_2 δύο κανονικές εκφράσεις***
5. $(R_1 \circ R_2)$ ***,όπου R_1 και R_2 δύο κανονικές εκφράσεις***
6. R_1^* ***,όπου R_1 μια κανονική έκφραση***

Παραδείγματα

- 0*10*
  - {w | η w περιέχει ακριβώς ένα 1}
- (ΣΣ)*
  - {w | η w είναι λέξη άρτιου μήκους}
- (0∪1)*010Σ*
  - {w | η w περιέχει την υπολέξη 010}
- 1*0* ∪ 01*
  - {w | η w αρχίζει με 0 ή περισσότερα 1 και ακολουθείται με 0 η περισσότερα 0, ή ξεκινάει με 0 και ακολουθείται από 0 ή περισσότερα 1}

Ισοδυναμία με πεπερασμένα αυτόματα

Θεώρημα:
Μια γλώσσα είναι κανονική εάν και μόνο εάν υπάρχει κανονική έκφραση που να την περιγράφει.

- Δυο κατευθύνσεις για απόδειξη:
  - Εάν μια γλώσσα περιγράφεται από κάποια κανονική έκφραση, τότε είναι κανονική.
    - Υπάρχει κάποιο NFA που την αναγνωρίζει.
  - Εάν μια γλώσσα είναι κανονική τότε υπάρχει κανονική έκφραση που την περιγράφει.
    - Κάθε DFA μπορεί να μετατραπεί σε κανονική έκφραση

Κανονική Έκφραση => Κανονική Γλώσσα

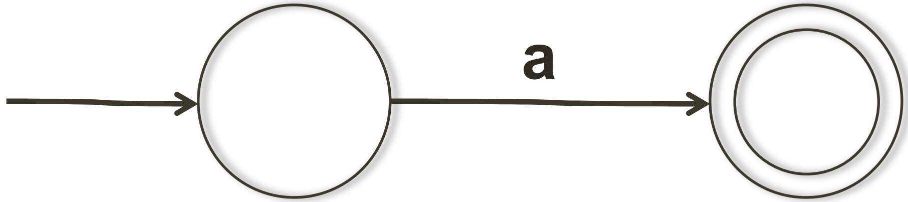
- Στόχος: Να κατασκευάσουμε μηχανισμό για μετατροπή κάθε κανονικής έκφρασης σε NFA.

- Έστω A μια γλώσσα που περιγράφεται από μια έκφραση R

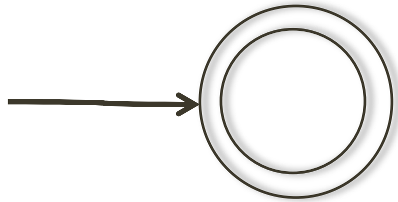
- Εάν μπορούμε να μετατρέψουμε την R σε ένα NFA που να αναγνωρίζει την A, τότε η A είναι κανονική.

Περιπτώσεις Έκφρασης R

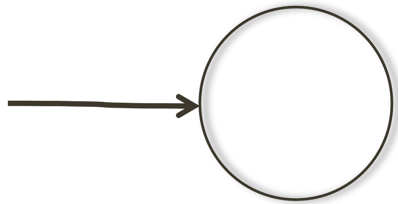
- R = a, για κάποιο α στο αλφάβητο Σ
  - L(R)={a}



- R = e
  - L(R)={e}



- R = $\emptyset$
  - L(R)= $\emptyset$



Περιπτώσεις Έκφρασης R

- Ένωση εκφράσεων: $R = R_1 \cup R_2$
  - Χρησιμοποιούμε την κατασκευή της κλειστότητας ως προς την ένωση
- Συναρμογή εκφράσεων: $R = R_1 \circ R_2$
  - Χρησιμοποιούμε την κατασκευή της κλειστότητας ως προς την συναρμογή
- Σώρευση έκφρασης: $R = R^*$
  - Χρησιμοποιούμε την κατασκευή της κλειστότητας ως προς την σώρευση

Παραδείγματα

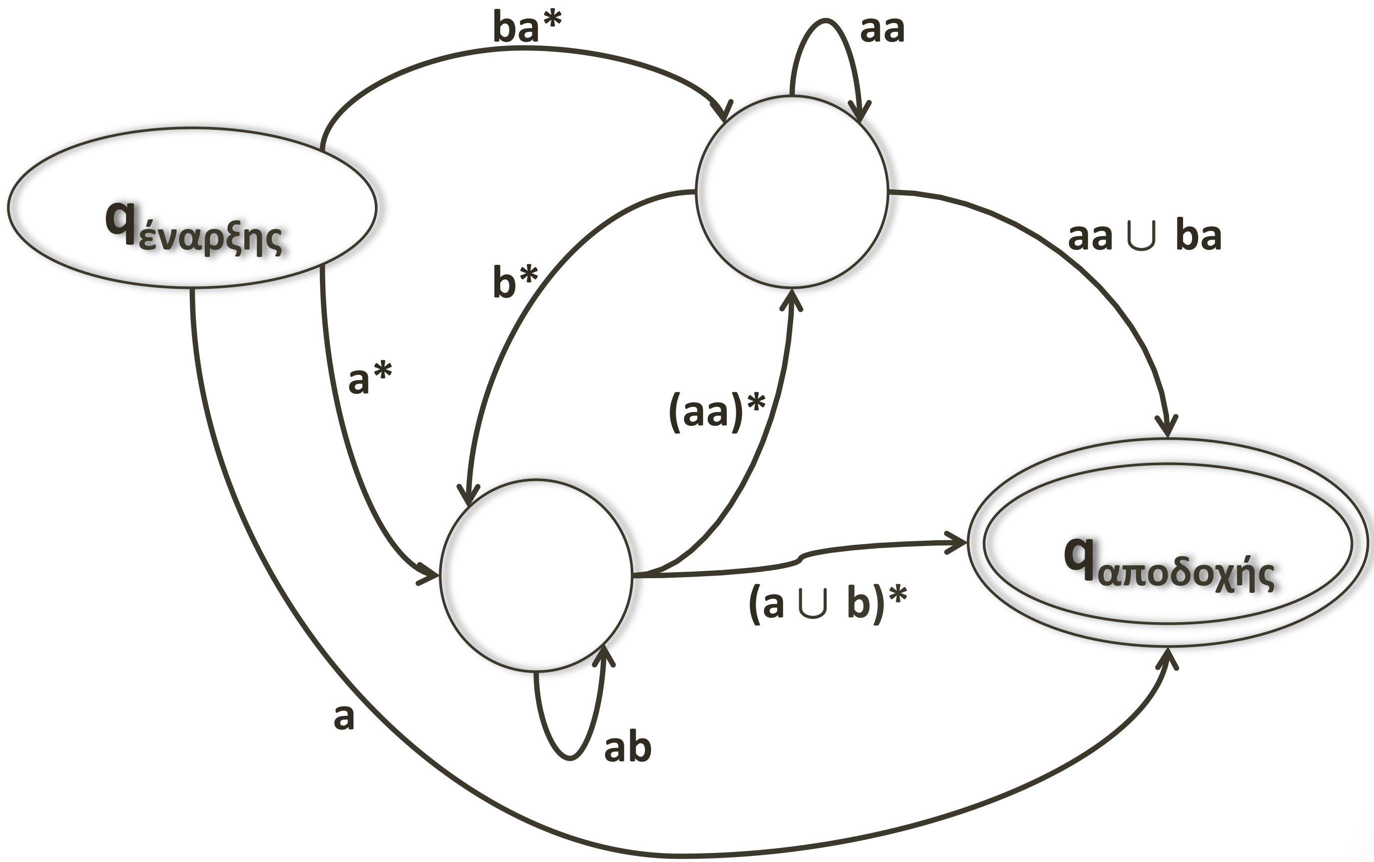
- Μετατροπή κανονικών εκφράσεων σε NFA
  - $(ab \cup a)^*$
  - $(a \cup b)^*aba$

Κανονική Γλώσσα => Κανονική Έκφραση

- Στόχος: Να κατασκευάσουμε μηχανισμό για μετατροπή κάθε DFA σε κανονική έκφραση.
- Αφού η γλώσσα είναι κανονική τότε υπάρχει ένα DFA που την αναγνωρίζει
- Πρέπει να περιγράψουμε μια διαδικασία που μετατρέπει DFAs σε ισοδύναμες κανονικές εκφράσεις
- Διαδικασία: Μετατροπή από DFA σε **γενικευμένο μη ντετερμινιστικό πεπερασμένο αυτόματο** (GNFA), και από GNFA σε κανονική έκφραση.

Γενικευμένο NFA (GNFA)

- Βέλη μετάβασης επιγράφονται με κανονικές εκφράσεις και όχι μόνο με σύμβολα του αλφαβήτου και e.



Ειδική μορφή GNFA

- Η **εναρκτήρια κατάσταση** έχει βέλη προς όλες τις καταστάσεις, αλλά δεν έχει εισερχόμενο βέλος.

- Υπάρχει μόνο **μια κατάσταση αποδοχής**, διαφορετική από την εναρκτήρια, και έχει μόνο εισερχόμενα βέλη από όλες τις υπόλοιπες καταστάσεις.

- Κάθε κατάσταση εκτός από την εναρκτήρια και την τελική, έχει εξερχόμενα βέλη προς όλες τις άλλες καταστάσεις και τον εαυτό της.

DFA σε GNFA

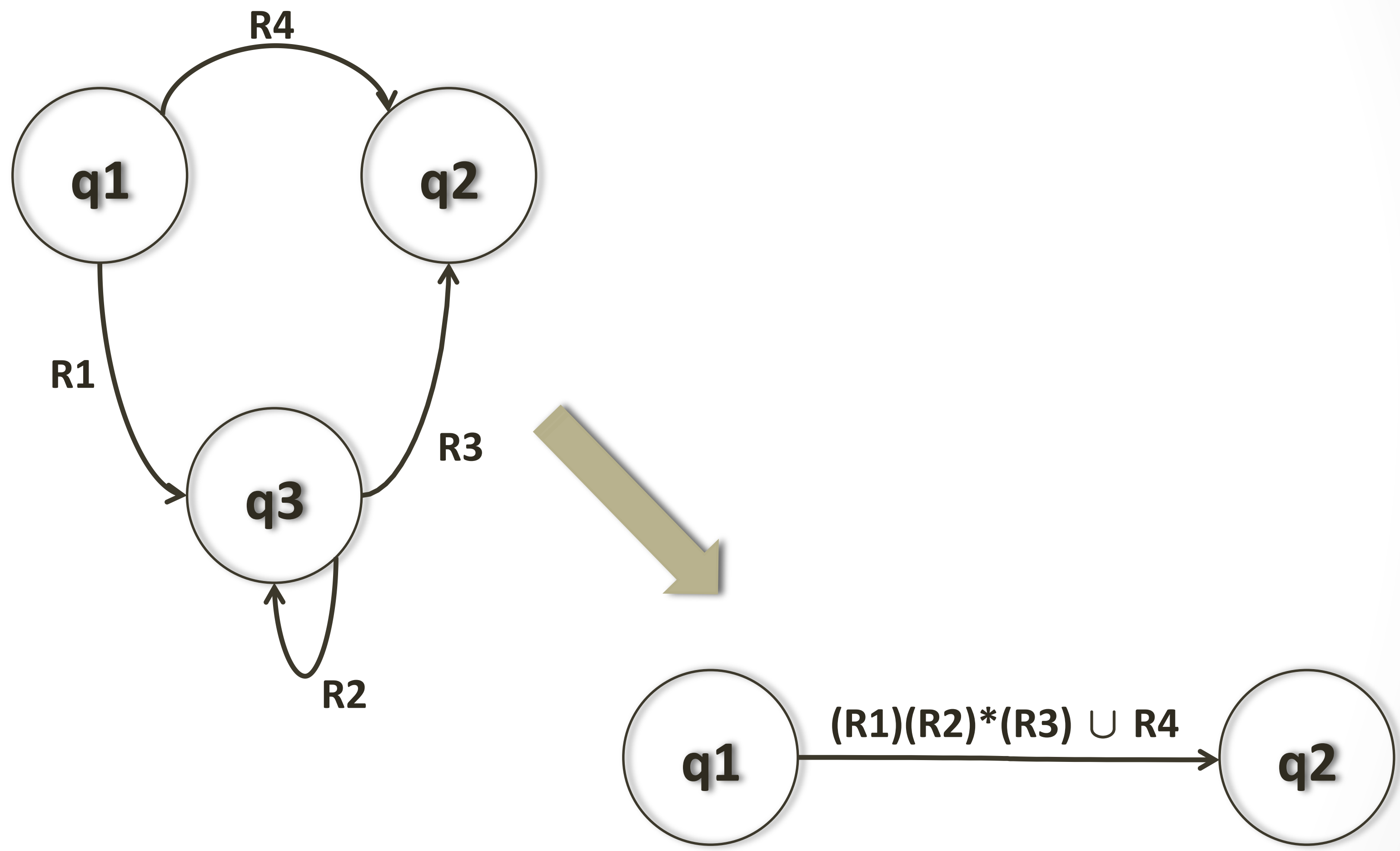
- Βήμα 1:
  - Προσθέτουμε μια νέα εναρκτήρια κατάσταση με βέλος e προς την παλιά
  - Προσθέτουμε μια νέα κατάσταση αποδοχής με εισερχόμενα βέλη από όλες τις παλιές καταστάσεις αποδοχής
- Βήμα 2:
  - Αν ένα βέλος έχει περισσότερα από ένα σύμβολα τότε αντικαθιστούμε με ένα βέλος με επιγραφή την ένωση των συμβόλων
- Βήμα 3:
  - Αν από μια κατάσταση σε κάποια άλλη δεν υπάρχει κανένα βέλος τότε προσθέτουμε ένα με επιγραφή $\emptyset$

GNFA σε Κανονική Έκφραση (Ιδέα)

- Εάν το GNFA έχει 2 καταστάσεις (αρχική και αποδοχης) τότε η κανονική έκφραση είναι η επιγραφή του μοναδικού βέλους από την αρχική στην κατάσταση αποδοχής

- Αν κ>2 καταστάσεις τότε αφαιρούμε μια κατάσταση και αλλάζουμε την επιγραφή των βελών για να πάρουμε ένα ισοδύναμο GNFA κ-1 καταστάσεων. Επαναλαμβάνουμε μέχρι κ=2.

Αφαίρεση κατάστασης



Τυπικός Ορισμός GNFA

Ορισμός

Γενικευμένο Μη Ντετερμινιστικό Πεπερασμένο Αυτόματο ***είναι μια πεντάδα*** $(Q, \Sigma, \delta, q_{start}, q_{end})$

1. ***Q = πεπερασμένο σύνολο καταστάσεων***
2. ***Σ = αλφάβητο***
3. ***Συνάρτηση μετάβασης***
4. $\delta : (Q \setminus \{q_{end}\}) \times (Q \setminus \{q_{start}\}) \rightarrow \mathcal{R}$

   Το βέλος από μια κατάσταση q ∈ Q σε μια κατάσταση q' ∈ Q έχει επιγραφή μια κανονική έκφραση R
5. ***$q_{start} \in Q$ = αρχική κατάσταση***
6. ***$q_{end} \in Q$ = τελική κατάσταση***

Αλγόριθμος μετατροπής GNFA

- ΜΕΤΑΤΡΟΠΗ(G)

1. Έστω k το πλήθος των καταστάσεων
2. **Εάν k=2**
   - Επέστρεψε R: η έκφραση στο βέλος από την αρχική στην τελική κατάσταση
3. **Εάν k>2**
   - Επέλεξε οποιαδήποτε κατάσταση $q^* \in Q \setminus \{q_{start}, q_{end}\}$
   - Κατασκευάζουμε το $G' = (Q', \Sigma, \delta', q_{start}, q_{end})$

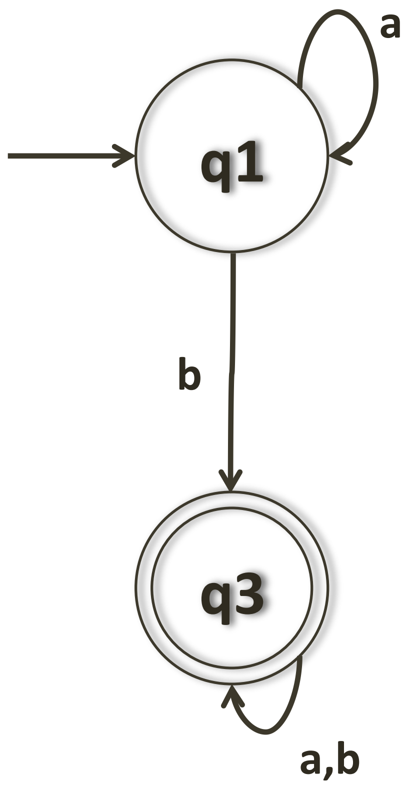
     ① $Q' = Q \setminus \{q^*\}$

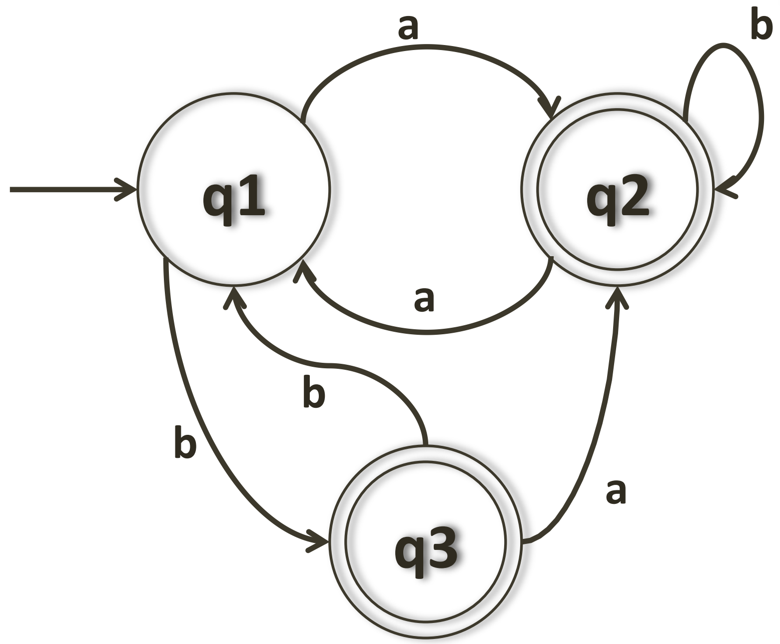
     ② Για κάθε $q_i \in Q' \setminus \{q_{end}\}$ και $q_j \in Q' \setminus \{q_{start}\}$

$$\delta'(q_i, q_j) = (R_1)(R_2)^*(R_3) \cup R_4$$

$$R_1 = \delta'(q_i, q^*), R_2 = \delta'(q^*, q^*), R_3 = \delta'(q^*, q_j), R_4 = \delta'(q_i, q_j)$$

Παραδείγματα





Ερωτήσεις;