

Γλώσσες (Languages)

Κανονικές Γλώσσες (Regular Languages)

Διάλεξη 6: Θεωρία Υπολογισμού · HY-280

Χαράλαμπος Ε. Τσουρακάκης

Γλώσσες

Ορισμός

Μια τυπική γλώσσα (formal language) (ή απλώς γλώσσα) πάνω σε ένα πεπερασμένο αλφάβητο Σ είναι ένα σύνολο συμβολοσειρών (strings) πάνω στο Σ , δηλαδή ένα οποιοδήποτε υποσύνολο του Σ^* .

Συμβολισμός: οι γλώσσες συμβολίζονται με πλάγια κεφαλαία γράμματα (συνήθως L , αλλά και $A, B, C, \dots$).

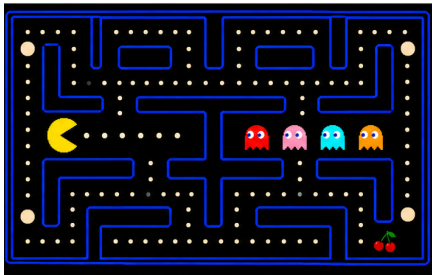
Παραδείγματα γλωσσών:

- Το κενό σύνολο $\emptyset$.
- Το σύνολο $\{\epsilon\}$.
- Το σύνολο $\{0, 1\}^*$.

Καθένα από τα παρακάτω σύνολα είναι γλώσσα:

- Το σύνολο {KATERINA, ANNA, ΔΗΜΗΤΡΙΟΣ, ΕΥΤΥΧΙΟΣ}.
- Το σύνολο όλων των Αγγλικών λέξεων του *Oxford English Dictionary*.
- Οι συμβολοσειρές του $\{0, 1\}^*$ με περιττό πλήθος 1.
- Οι συμβολοσειρές του $\{0, 1\}^*$ που παριστάνουν πρώτο αριθμό στη βάση 13.
- Τα προγράμματα Python που τυπώνουν «Hello World!».

Το Pac-Man ως γλώσσα



Σταθεροποιούμε έναν λαβύρινθο,
μια αρχική θέση s και έναν
στόχο t .

*Μοντέλο λαβυρίνθου: αγνοούμε
τα φαντάσματα και τον χρόνο.*

Αλφάβητο κινήσεων και γλώσσα

$$\Sigma = \{U, D, L, R\}.$$

U = πάνω, D = κάτω,

L = αριστερά, R = δεξιά.

Κάθε σύμβολο ζητά ένα βήμα προς ένα
γειτονικό κελί. Μια κίνηση προς τοίχο
απορρίπτεται.

**Η γλώσσα των διαδρομών από το s
στο t :**

$$L_{s,t} = \{w \in \Sigma^* : s \xrightarrow{w} t\}.$$

Ο κύβος του Rubik ως γλώσσα

Αλφάβητο κινήσεων

$$\Sigma = \{U, D, L, R, F, B\}.$$

U : πάνω, D : κάτω, L : αριστερά,

R : δεξιά, F : μπροστά, B : πίσω.

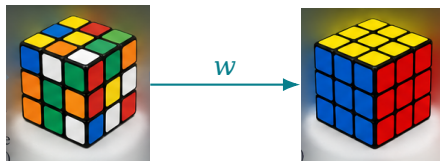
Κάθε σύμβολο: στροφή 90°
δεξιόστροφα, όπως κοιτάμε την έδρα
από έξω.

Η γλώσσα των λύσεων

Σταθεροποιούμε μια αρχική διάταξη q_0 :

$$L_{q_0} = \{w \in \Sigma^* : q_0 \xrightarrow{w} q_{\text{solved}}\}.$$

Μια συμβολοσειρά είναι ένα
«πρόγραμμα κινήσεων».



αρχική q_0

λυμένος

$$w \in L_{q_0} \iff q_0 \xrightarrow{w} q_{\text{solved}}.$$

Για παράδειγμα, $RURFB \in \Sigma^*$. Αν
 $RURFB \in L_{q_0}$ εξαρτάται από το
 q_0 .

Τρία διαφορετικά «κενά» αντικείμενα

Είναι πολύ σημαντικό να διακρίνουμε τα ακόλουθα τρία αντικείμενα:

- $\emptyset$ είναι η **κενή γλώσσα**: ένα σύνολο με μηδέν συμβολοσειρές. Το $\emptyset$ δεν είναι συμβολοσειρά.
- $\{\varepsilon\}$ είναι μια γλώσσα που περιέχει **ακριβώς μία** συμβολοσειρά, μήκους μηδέν. Το $\{\varepsilon\}$ δεν είναι κενό και δεν είναι συμβολοσειρά.
- ε είναι η **κενή συμβολοσειρά**: μια ακολουθία μήκους μηδέν. Το ε δεν είναι γλώσσα.

Προσοχή

$$\emptyset \neq \{\varepsilon\}, \quad \varepsilon \notin \emptyset, \quad \varepsilon \in \{\varepsilon\}.$$

Κατασκευή γλωσσών

Συνολοθεωρητικές πράξεις

Οι γλώσσες είναι σύνολα, άρα συνδυάζονται όπως όλα τα σύνολα. Αν A, B γλώσσες πάνω στο Σ , τότε γλώσσες πάνω στο Σ είναι και:

- η ένωση $A \cup B$ και η τομή $A \cap B$,
- η διαφορά $A \setminus B$,
- η συμμετρική διαφορά $A \oplus B$,
- το συμπλήρωμα $\overline{A} := \Sigma^* \setminus A$.

Υπάρχουν όμως δύο ακόμη πράξεις, ειδικές για σύνολα συμβολοσειρών:

1. συνένωση,
2. αστέρι Kleene.

Ορισμός

Η **συνένωση (concatenation)** δύο γλωσσών A και B , που συμβολίζεται $A \bullet B$ ή απλώς AB , είναι το σύνολο των συμβολοσειρών που προκύπτουν συνενώνοντας μια συμβολοσειρά του A με μια συμβολοσειρά του B : $A \bullet B := \{xy \mid x \in A \text{ και } y \in B\}$.

Παράδειγμα

Αν $A = \{\text{HOCUS}, \text{ABRACA}\}$ και $B = \{\text{POCUS}, \text{DABRA}\}$, τότε
 $A \bullet B = \{\text{HOCUSPOCUS}, \text{ABRACAPOCUS}, \text{HOCUSDABRA}, \text{ABRACADABRA}\}$.

Ουδέτερο και απορροφητικό στοιχείο

Για κάθε γλώσσα A ισχύουν:

$$\emptyset \cdot A = A \cdot \emptyset = \emptyset \quad \text{και} \quad \{\varepsilon\} \cdot A = A \cdot \{\varepsilon\} = A.$$

- Η κενή γλώσσα $\emptyset$ «απορροφά» τη συνένωση (όπως το 0 τον πολλαπλασιασμό).
- Η γλώσσα $\{\varepsilon\}$ είναι ουδέτερο στοιχείο (όπως το 1 για τον πολλαπλασιασμό).
- Προσοχή: η συνένωση δεν είναι αντιμεταθετική! Γενικά $A \cdot B \neq B \cdot A$.

Παράδειγμα

$$A \cdot B = \{\text{HOCUSPOCUS, ABRACAPOCUS, HOCUSDABRA, ABRACADABRA}\} \\ \neq$$

$$B \cdot A = \{\text{POCUSHOCUS, POCUSABRACA, DABRAHOCUS, DABRAABRACA}\}.$$

Το αστέρι Kleene

Ορισμός

Η **κλειστότητα Kleene** ή **αστέρι Kleene** (Kleene star) μιας γλώσσας L , που συμβολίζεται L^* , είναι το σύνολο των συμβολοσειρών που προκύπτουν συνενώνοντας μια ακολουθία μηδέν ή περισσότερων συμβολοσειρών του L .

Παράδειγμα

$\{0, 11\}^* = \{\varepsilon, 0, 00, 11, 000, 011, 110, 0000, 0011, 0110, 1100, 1111, \dots\}$

Τυπικά, το L^* ορίζεται αναδρομικά ως το σύνολο όλων των συμβολοσειρών w ώστε είτε $w = \varepsilon$, είτε $w = xy$ για κάποιες συμβολοσειρές $x \in L$ και $y \in L^*$.

Το αστέρι Kleene — Αναδρομικά

Ορισμός

Το L^* είναι το μικρότερο σύνολο συμβολοσειρών που ικανοποιεί τους εξής κανόνες:

- **Βάση:** $\varepsilon \in L^*$.
- **Αναδρομικό βήμα:** αν $x \in L$ και $y \in L^*$, τότε $xy \in L^*$.

Παράδειγμα

Έστω $L = \{0, 11\}$. Κατασκευάζουμε τη συμβολοσειρά 0110:

$$\varepsilon \in L^* \quad (\text{βάση}),$$

$$0 = 0 \cdot \varepsilon \in L^* \quad (\text{επιλέγουμε } x = 0),$$

$$110 = 11 \cdot 0 \in L^* \quad (\text{επιλέγουμε } x = 11),$$

$$0110 = 0 \cdot 110 \in L^* \quad (\text{επιλέγουμε } x = 0).$$

Ιδιότητες του αστεριού Kleene

- Άμεση συνέπεια του ορισμού:

$$\emptyset^* = \{\varepsilon\}^* = \{\varepsilon\}.$$

- Για κάθε άλλη γλώσσα L , το L^* είναι **άπειρο** και περιέχει οσοδήποτε μεγάλες (αλλά πεπερασμένες!) συμβολοσειρές.
- Ισοδύναμα: το L^* είναι το *μικρότερο* υπερσύνολο του L που περιέχει το ε και είναι κλειστό ως προς τη συνένωση (εξ ου και «κλειστότητα»).
- Το Σ^* είναι ακριβώς η κλειστότητα Kleene του αλφαβήτου Σ
 - Για να δείτε το γιατί, θεωρήστε ένα οποιοδήποτε σύμβολο ως συμβολοσειρά μήκους 1

Το συν του Kleene

Ορισμός

Το **συν του Kleene** (Kleene plus) μιας γλώσσας L ορίζεται ως $L^+ := L \cdot L^*$, δηλαδή το σύνολο των συμβολοσειρών που προκύπτουν συνενώνοντας μία ή περισσότερες συμβολοσειρές του L .

Παράδειγμα

$\{0, 11\}^+ = \{0, 00, 11, 000, 011, 110, \dots\}$, δλδ. όπως το $\{0, 11\}^*$ αλλά χωρίς το ε .

Αλγεβρικές ταυτότητες (Άσκηση + Φροντιστήριο)

Για όλες τις γλώσσες A , B και C ισχύουν:

(a) $A \cup B = B \cup A$

(b) $(A \cup B) \cup C = A \cup (B \cup C)$

(c) $\emptyset \cdot A = A \cdot \emptyset = \emptyset$

(d) $\{\varepsilon\} \cdot A = A \cdot \{\varepsilon\} = A$

(e) $(A \cdot B) \cdot C = A \cdot (B \cdot C)$

(f) $A \cdot (B \cup C) = (A \cdot B) \cup (A \cdot C)$

(g) $(A \cup B) \cdot C = (A \cdot C) \cup (B \cdot C)$

Άσκηση στο set ασκήσεων και στο φροντιστήριο την Παρασκευή

Ταυτότητες

Για κάθε γλώσσα L ισχύουν:

$$\begin{aligned} \text{(a)} \quad L^* &= \{\varepsilon\} \cup L^+ = L^* \cdot L^* = (L \cup \{\varepsilon\})^* = (L \setminus \{\varepsilon\})^* \\ &= \{\varepsilon\} \cup L \cup (L^+ \cdot L^+) \end{aligned}$$

$$\text{(b)} \quad L^+ = L \cdot L^* = L^* \cdot L = L^+ \cdot L^* = L^* \cdot L^+ = L \cup (L^+ \cdot L^+)$$

$$\text{(c)} \quad L^+ = L^* \text{ αν και μόνο αν } \varepsilon \in L.$$

Παρατηρήστε ότι το αστερί είναι ταυτοδύναμο: $(L^*)^* = L^*$ και $L^* \cdot L^* = L^*$.

Δυνάμεις μιας γλώσσας

Έστω $A \subseteq \Sigma^*$ μια γλώσσα.

Ορίζουμε $A^0 = \{\varepsilon\}$, $A^{n+1} = A^n A$.

Δυνάμεις μιας γλώσσας

Έστω $A \subseteq \Sigma^*$ μια γλώσσα.

Ορίζουμε $A^0 = \{\varepsilon\}$, $A^{n+1} = A^n A$.

Από αυτές τις δυνάμεις ορίζουμε:

$$A^* = \bigcup_{n \geq 0} A^n = A^0 \cup A^1 \cup A^2 \cup \dots$$

$$A^+ = \bigcup_{n \geq 1} A^n = A^1 \cup A^2 \cup A^3 \cup \dots$$

Δυνάμεις μιας γλώσσας

Έστω $A \subseteq \Sigma^*$ μια γλώσσα.

Ορίζουμε $A^0 = \{\varepsilon\}$, $A^{n+1} = A^n A$.

Από αυτές τις δυνάμεις ορίζουμε:

$$A^* = \bigcup_{n \geq 0} A^n = A^0 \cup A^1 \cup A^2 \cup \dots$$

$$A^+ = \bigcup_{n \geq 1} A^n = A^1 \cup A^2 \cup A^3 \cup \dots$$

Άρα $A^* = A^0 \cup A^+ = \{\varepsilon\} \cup A^+.$

Άσκηση: πότε ισχύει $A^+ = A^*$;

Θεώρημα

Για κάθε γλώσσα A , $A^+ = A^* \iff \varepsilon \in A$.

Άσκηση: πότε ισχύει $A^+ = A^*$;

Θεώρημα

Για κάθε γλώσσα A , $A^+ = A^* \iff \varepsilon \in A$.

Θα αποδείξουμε τις **δύο κατευθύνσεις** χωριστά:

$$\varepsilon \in A \implies A^+ = A^*$$

και

$$A^+ = A^* \implies \varepsilon \in A.$$

Πριν ξεκινήσουμε

Πάντα ισχύει $A^+ \subseteq A^*$, αφού το A^* επιτρέπει 0, 1, 2, ... συνενώσεις, ενώ το A^+ επιτρέπει 1, 2, ...

Απόδειξη $A^+ = A^*$: η πρώτη κατεύθυνση

Υποθέτουμε ότι $\varepsilon \in A$.

Θέλουμε να δείξουμε $A^* = A^+$.

Απόδειξη $A^+ = A^*$: η πρώτη κατεύθυνση

Υποθέτουμε ότι $\varepsilon \in A$.

Θέλουμε να δείξουμε $A^* = A^+$. Γνωρίζουμε ήδη ότι $A^+ \subseteq A^*$. Άρα αρκεί

να δείξουμε $A^* \subseteq A^+$.

Απόδειξη $A^+ = A^*$: η πρώτη κατεύθυνση

Υποθέτουμε ότι $\varepsilon \in A$.

Θέλουμε να δείξουμε $A^* = A^+$. Γνωρίζουμε ήδη ότι $A^+ \subseteq A^*$. Άρα αρκεί

να δείξουμε $A^* \subseteq A^+$. Αλλά $A^* = A^0 \cup A^+ = \{\varepsilon\} \cup A^+$.

Εφόσον $\varepsilon \in A = A^1 \subseteq A^+$, έχουμε $\{\varepsilon\} \cup A^+ = A^+$.

Επομένως

$$A^* = A^+.$$

Απόδειξη $A^+ = A^*$: η αντίστροφη κατεύθυνση

Τώρα υποθέτουμε $A^+ = A^*$.

Θέλουμε να δείξουμε ότι $\varepsilon \in A$.

Απόδειξη $A^+ = A^*$: η αντίστροφη κατεύθυνση

Τώρα υποθέτουμε $A^+ = A^*$.

Θέλουμε να δείξουμε ότι $\varepsilon \in A$.

Γνωρίζουμε ότι $\varepsilon \in A^0 \subseteq A^*$.

Άρα, από $A^+ = A^*$, $\varepsilon \in A^+$.

Απόδειξη $A^+ = A^*$: η αντίστροφη κατεύθυνση

Τώρα υποθέτουμε $A^+ = A^*$.

Θέλουμε να δείξουμε ότι $\varepsilon \in A$.

Γνωρίζουμε ότι $\varepsilon \in A^0 \subseteq A^*$.

Άρα, από $A^+ = A^*$, $\varepsilon \in A^+$.

Επομένως υπάρχει $n \geq 1$ τέτοιο ώστε $\varepsilon \in A^n$.

Αυτό σημαίνει ότι υπάρχουν $x_1, x_2, \dots, x_n \in A$ τέτοια ώστε $x_1 x_2 \cdots x_n = \varepsilon$. Παίρνοντας μήκη, $0 = |\varepsilon| = |x_1| + |x_2| + \cdots + |x_n|$.

Όλα τα μήκη είναι μη αρνητικά, άρα $|x_1| = \cdots = |x_n| = 0$.

Επομένως $x_1 = \varepsilon$, και αφού $x_1 \in A$,

Έλεγχος με δύο παραδείγματα (sanity check)

Παράδειγμα 1. Έστω $A = \{A, B\}$.

Τότε $A^+ = \{A, B, AA, AB, BA, BB, \dots\}$, ενώ $A^* = A^+ \cup \{\epsilon\}$.

Άρα $A^+ \neq A^*$.

Έλεγχος με δύο παραδείγματα (sanity check)

Παράδειγμα 1. Έστω $A = \{A, B\}$.

Τότε $A^+ = \{A, B, AA, AB, BA, BB, \dots\}$, ενώ $A^* = A^+ \cup \{\varepsilon\}$.

Άρα $A^+ \neq A^*$.

Παράδειγμα 2. Έστω $B = \{\varepsilon, A, B\}$.

Τώρα $\varepsilon \in B$, επομένως $B^+ = B^*$.

Προσέξτε

Το ε είναι **λέξη**. Δεν είναι το κενό σύνολο $\emptyset$.

Κανονικές γλώσσες και κανονικές εκφράσεις

Κανονικές γλώσσες: οι επιτρεπόμενες κατασκευές

Έχουμε ήδη ορίσει πράξεις πάνω σε γλώσσες. Τώρα περιορίζουμε τους τρόπους με τους οποίους επιτρέπεται να τις κατασκευάζουμε.

Ορισμός

Οι **κανονικές γλώσσες** πάνω στο Σ ορίζονται αναδρομικά:

1. Η $\emptyset$ είναι κανονική γλώσσα.
2. Για κάθε $a \in \Sigma$, η $\{a\}$ είναι κανονική γλώσσα.
3. Αν A, B είναι κανονικές, τότε και οι $A \cup B$, AB , A^* είναι κανονικές.

Και τίποτε άλλο

Μια γλώσσα είναι κανονική μόνο αν προκύπτει από **πεπερασμένα** βήματα αυτών των κατασκευών.

Πρώτα παραδείγματα κανονικών γλωσσών

Η γλώσσα $\{\varepsilon\}$. Δεν χρειάζεται ξεχωριστή βασική περίπτωση:

$$\{\varepsilon\} = \emptyset^*.$$

Πρώτα παραδείγματα κανονικών γλωσσών

Η γλώσσα $\{\varepsilon\}$. Δεν χρειάζεται ξεχωριστή βασική περίπτωση:

$$\{\varepsilon\} = \emptyset^*.$$

Η γλώσσα $\{001, 110\}$.

$$\{001, 110\} = \{0\}\{0\}\{1\} \cup \{1\}\{1\}\{0\}.$$

Χρησιμοποιήσαμε μόνο μονοσύνολα συμβόλων, συνένωση και ένωση.

Πρώτα παραδείγματα κανονικών γλωσσών

Η γλώσσα $\{\varepsilon\}$. Δεν χρειάζεται ξεχωριστή βασική περίπτωση:

$$\{\varepsilon\} = \emptyset^*.$$

Η γλώσσα $\{001, 110\}$.

$$\{001, 110\} = \{0\}\{0\}\{1\} \cup \{1\}\{1\}\{0\}.$$

Χρησιμοποιήσαμε μόνο μονοσύνολα συμβόλων, συνένωση και ένωση.

Γενίκευση

Κάθε πεπερασμένη γλώσσα είναι κανονική. Κάθε μη κενή συμβολοσειρά κατασκευάζεται με συνένωση των συμβόλων της· έπειτα παίρνουμε πεπερασμένη ένωση των αντίστοιχων μονοσυνόλων. Οι περιπτώσεις ε και $\emptyset$ καλύπτονται χωριστά όπως παραπάνω.

Αναδρομικός ορισμός: σύνταξη και σημασία

Για $a \in \Sigma$ και κανονικές εκφράσεις r, s :

Κανονική έκφραση	Γλώσσα που περιγράφει
$\emptyset$	$L(\emptyset) = \emptyset$
ε	$L(\varepsilon) = \{\varepsilon\}$
a	$L(a) = \{a\}$
$(r + s)$	$L(r + s) = L(r) \cup L(s)$
(rs)	$L(rs) = L(r)L(s)$
(r^*)	$L(r^*) = L(r)^*$

Δεν υπάρχουν άλλες κανονικές εκφράσεις πέρα από όσες παράγονται με πεπερασμένη εφαρμογή αυτών των κανόνων.

Οι ορισμοί αντιστοιχούν: κάθε κανονική έκφραση περιγράφει κανονική γλώσσα και κάθε κανονική γλώσσα έχει μια τέτοια έκφραση.

Προτεραιότητα πράξεων και παρενθέσεις

αστέρι * > συνένωση > ένωση +

Παραλείπουμε μόνο τις παρενθέσεις που δεν αλλάζουν την ερμηνεία.

Γραφή	Ερμηνεία	Λανθασμένη Ερμηνεία
01*	0(1*)	(01)*
0 + 10	0 + (10)	(0 + 1)0
(0 + 1)*00	((0 + 1)*) (00)	((0 + 1)00)*

Προτεραιότητα πράξεων και παρενθέσεις

αστέρι * > συνένωση > ένωση +

Παραλείπουμε μόνο τις παρενθέσεις που δεν αλλάζουν την ερμηνεία.

Γραφή	Ερμηνεία	Λανθασμένη Ερμηνεία
01^*	$0(1^*)$	$(01)^*$
$0 + 10$	$0 + (10)$	$(0 + 1)0$
$(0 + 1)^*00$	$((0 + 1)^*)(00)$	$((0 + 1)00)^*$

Παράδειγμα ανάγνωσης: $0^*10(01 + 1^*) = (0^*) 1 0 ((01) + (1^*))$.

Μηδέν ή περισσότερα 0, έπειτα 10, και τέλος είτε 01 είτε μηδέν ή περισσότερα 1.

Σύνταξη: κανονική έκφραση

r

$$r = (0 + 1)^*00.$$

Μια πεπερασμένη γραπτή περιγραφή, με σύμβολα, τελεστές και παρενθέσεις.

Σημασία: κανονική γλώσσα

$L(r)$

Όλες οι δυαδικές συμβολοσειρές που τελειώνουν σε **00**:

$$L(r) = \{00, 000, 100, 0100, \dots\}.$$

Για μια συμβολοσειρά w , γράφουμε $w \in L(r)$ όταν η έκφραση r περιγράφει τη w Πχ. **100** $\in L(r)$, **101** $\notin L(r)$, $\varepsilon \notin L(r)$.

Προσοχή

Στις κανονικές εκφράσεις το $+$ δηλώνει **ένωση**, όχι το συν του Kleene. Το $r + s$ και το A^+ είναι διαφορετικοί συμβολισμοί.

Ισοδύναμες κανονικές εκφράσεις

Ορισμός

Δύο κανονικές εκφράσεις r, s είναι **ισοδύναμες** αν

$$r \equiv s \iff L(r) = L(s).$$

Οι ταυτότητες των γλωσσών δίνουν ταυτότητες για τις εκφράσεις:

$$r + s \equiv s + r, \quad r + \emptyset \equiv r,$$

$$r\varepsilon \equiv \varepsilon r \equiv r, \quad r\emptyset \equiv \emptyset r \equiv \emptyset,$$

$$r(s + t) \equiv rs + rt, \quad (r + s)t \equiv rt + st.$$

Ισοδύναμες κανονικές εκφράσεις

Ορισμός

Δύο κανονικές εκφράσεις r, s είναι **ισοδύναμες** αν

$$r \equiv s \iff L(r) = L(s).$$

Οι ταυτότητες των γλωσσών δίνουν ταυτότητες για τις εκφράσεις:

$$r + s \equiv s + r, \quad r + \emptyset \equiv r,$$

$$r\varepsilon \equiv \varepsilon r \equiv r, \quad r\emptyset \equiv \emptyset r \equiv \emptyset,$$

$$r(s + t) \equiv rs + rt, \quad (r + s)t \equiv rt + st.$$

Δεν είναι όλες οι ιδιότητες ίδιες

Η ένωση είναι αντιμεταθετική, η συνένωση γενικά όχι: $01 \neq 10$.

Άσκηση

Θέλουμε να δείξουμε ότι

$$a^*(a+b)^* \equiv (a+ba^*)^*.$$

Ιδέα: Θα δείξουμε ότι και οι δύο εκφράσεις παριστάνουν την ίδια γλώσσα, τη $\{a, b\}^* = L((a+b)^*)$.

Θυμηθείτε ότι οι γλώσσες είναι **σύνολα συμβολοσειρών**. Για να αποδείξουμε ότι δύο γλώσσες A και B είναι ίσες, δείχνουμε ότι για κάθε συμβολοσειρά x ισχύει $x \in A \iff x \in B$.

Η μια συνεπαγωγή $x \in A \iff x \in B$ είναι προφανής εδώ: Ποια; **Λεπτομέρειες στον πίνακα**

Άσκηση

Θέλουμε να δείξουμε ότι

$$a^*(a+b)^* \equiv (a+ba^*)^*.$$

Ιδέα: Θα δείξουμε ότι και οι δύο εκφράσεις παριστάνουν την ίδια γλώσσα, τη $\{a, b\}^* = L((a+b)^*)$.

Θυμηθείτε ότι οι γλώσσες είναι **σύνολα συμβολοσειρών**. Για να αποδείξουμε ότι δύο γλώσσες A και B είναι ίσες, δείχνουμε ότι για κάθε συμβολοσειρά x ισχύει $x \in A \iff x \in B$.

Η μια συνεπαγωγή $x \in A \iff x \in B$ είναι προφανής εδώ: Ποια; **Λεπτομέρειες στον πίνακα**

Αριστερή πλευρά

Αφού $\varepsilon \in a^*$, $(a+b)^* \subseteq a^*(a+b)^*$.

Δεξιά πλευρά

Αφού $b \in ba^*$, τότε $(a+b)^* \subseteq (a+ba^*)^*$.

Σχεδιασμός μιας κανονικής έκφρασης

Σε όλα τα παραδείγματα, $\Sigma = \{0, 1\}$.

Ζητούμενη γλώσσα	Κανονική έκφραση
Συμβολοσειρές που τελειώνουν σε 0	$(0 + 1)^*0$
Συμβολοσειρές που περιέχουν 000	$(0 + 1)^*000(0 + 1)^*$
Συμβολοσειρές άρτιου μήκους	$((0 + 1)(0 + 1))^*$
Συμβολοσειρές άρτιου μήκους	$(00 + 01 + 10 + 11)^*$
Συμβολοσειρές περιττού μήκους	$(0 + 1)((0 + 1)(0 + 1))^*$
Συμβολοσειρές περιττού μήκους	$(00 + 01 + 10 + 11)^*(0 + 1)$

Οι παρενθέσεις αλλάζουν τη γλώσσα

Ο αστέρας εφαρμόζεται στην **αμέσως προηγούμενη έκφραση**. Χωρίς παρενθέσεις, δεν επαναλαμβάνει ό,τι έχει γραφτεί πριν από αυτόν.

Μόνο μηδενικά

0^*

$\{0^n : n \geq 0\}$

$\varepsilon, 0, 00, 000, \dots$

Επιτρέπεται **κάθε μήκος**, και το μηδενικό.

Τουλάχιστον τέσσερα

00000^*

$\{0^n : n \geq 4\}$

$0000, 00000, \dots$

Τέσσερα υποχρεωτικά **0**, και μετά όσα ακόμη θέλουμε.

Μήκος πολλαπλάσιο του 5

$(00000)^*$

$\{0^{5k} : k \geq 0\}$

$\varepsilon, 00000, \dots$

Επαναλαμβάνουμε **ολόκληρη την πεντάδα**.

Έλεγχος: η **0000** ανήκει μόνο στις δύο πρώτες γλώσσες. **ερώτηση: Η ε;**

Παραδείγματα

Ένα σύμβολο κάθε φορά

$$(0 + 1)^*$$

Κάθε επανάληψη προσθέτει ένα 0 ή ένα 1.

$$L((0 + 1)^*) = \Sigma^*.$$

Παραδείγματα: ε , 0, 1, 10, 011.

Δύο σύμβολα κάθε φορά

$$((0 + 1)(0 + 1))^*$$

Κάθε επανάληψη προσθέτει ένα από τα ζεύγη

00, 01, 10, 11.

Άρα παράγονται μόνο λέξεις μήκους $2k$.

Γιατί παράγεται κάθε λέξη άρτιου μήκους;

Χωρίζουμε τη λέξη σε ζεύγη, καθένα από τα οποία ανήκει στο $L((0 + 1)(0 + 1))$.

$$011000 = \begin{array}{ccc} 01 & 10 & 00 \\ | \{z\} & | \{z\} & | \{z\} \end{array} .$$

1ο ζεύγος 2ο ζεύγος 3ο ζεύγος

Εναλλασσόμενα μηδενικά και άσοι

Θέλουμε όλες τις δυαδικές λέξεις χωρίς δύο ίσα γειτονικά σύμβολα, δηλαδή χωρίς τα συνεχόμενα τμήματα **00** και **11**.

Η κατασκευή

$$R = (\varepsilon + \mathbf{1})(\mathbf{01})^*(\varepsilon + \mathbf{0}).$$

Προαιρετικός αρχικός **1**, έπειτα ζεύγη **01**, και προαιρετικό τελικό **0**.

Παραδείγματα

$$\mathbf{01010} = \varepsilon \cdot \mathbf{01} \cdot \mathbf{01} \cdot \mathbf{0},$$

$$\mathbf{10101} = \mathbf{1} \cdot \mathbf{01} \cdot \mathbf{01} \cdot \varepsilon,$$

$$\mathbf{10} = \mathbf{1} \cdot \varepsilon \cdot \mathbf{0}.$$

Περιλαμβάνονται επίσης οι ε , **0**, **1**.

Γιατί δεν χάνουμε καμία λέξη;

Από μια εναλλασσόμενη λέξη αφαιρούμε τον πρώτο **1**, αν ξεκινά με **1**.

Το υπόλοιπο ξεκινά με **0** και χωρίζεται σε ζεύγη **01**, με ένα πιθανό **0** στο τέλος. Μπορεί και να είναι κενό.

Λέξεις που περιέχουν το 0000

Η 0000 πρέπει να εμφανίζεται ως συνεχόμενο τμήμα της λέξης.

Ελεύθερο πρόθεμα — υποχρεωτικό τμήμα — ελεύθερο επίθεμα

$$R = (0 + 1)^* 0000 (0 + 1)^*.$$

$$w \in L(R) \iff w = x 0000 y \text{ για κάποια } x, y \in \Sigma^*.$$

Το x ή το y μπορεί να είναι η ε .

Περιλαμβάνεται

$$101000010 = 101 \cdot 0000 \cdot 10.$$

Περιλαμβάνονται και οι 0000,
0000.

Δεν περιλαμβάνεται

$$0101010.$$

Έχει τέσσερα 0, αλλά όχι
διαδοχικά.

Λέξεις που δεν περιέχουν το 0000: κατασκευή

Ανάμεσα σε δύο διαδοχικούς άσους, πριν από τον πρώτο και μετά τον τελευταίο, επιτρέπονται το πολύ τρία μηδενικά.

Θέτουμε $A = \varepsilon + 0 + 00 + 000$.

Κάθε άσος «κλείνει» την προηγούμενη ομάδα μηδενικών

$$R = (A1)^* A = ((\varepsilon + 0 + 00 + 000)1)^* (\varepsilon + 0 + 00 + 000).$$

Παράδειγμα:

$$0011000100 = 001 \varepsilon 1 0001 00 .$$

$ \{z\} $	$ \{z\} $	$ \{z\} $	$ \{z\} $
$A1$	$A1$	$A1$	A

Γιατί χρειάζεται το τελευταίο A ;

Χωρίς αυτό, κάθε μη κενή παραγόμενη λέξη θα τελείωνε σε 1. Θα χάναμε, για παράδειγμα, τις 00 και 100.

Λέξεις χωρίς 0000: απόδειξη

Θυμίζουμε ότι $A = \varepsilon + 0 + 00 + 000$ και $R = (A1)^* A$.

Κάθε λέξη του $L(R)$ αποφεύγει το 0000

Κάθε επιλογή από το A δίνει έως τρία 0. Διαδοχικές ομάδες μηδενικών χωρίζονται από έναν 1. Επομένως δεν μπορούν να δημιουργηθούν τέσσερα διαδοχικά 0.

Κάθε λέξη που αποφεύγει το 0000 ανήκει στο $L(R)$

Έστω w με k εμφανίσεις του 1. Χωρίζοντάς την στους άσους, γράφουμε

$$w = 0^{a_0} 1 0^{a_1} 1 \dots 0^{a_{k-1}} 1 0^{a_k}.$$

Αφού δεν εμφανίζεται η 0000, ισχύει $a_i \in \{0, 1, 2, 3\}$ για κάθε i . Άρα οι πρώτες k ομάδες περιγράφονται από το $(A1)^*$, και η τελευταία από το A .

Για $k = 0$, έχουμε μόνο $w = 0^{a_0}$ με $0 \leq a_0 \leq 3$. Περιλαμβάνεται και η ε .

Άρτιο πλήθος μηδενικών

Βάζουμε τα μηδενικά σε ζεύγη, όχι κατ' ανάγκη γειτονικά

$$R = 1^*(01^*01^*)^*.$$

Ο αρχικός παράγοντας δίνει μόνο άσους. Κάθε επανάληψη της παρένθεσης δίνει ακριβώς δύο μηδενικά.

Παράδειγμα:

$$1101011001 = 11 \cdot (01011) \cdot (0 \varepsilon 01).$$

Αντίστροφα: σε κάθε λέξη με $2k$ μηδενικά, ξεχωρίζουμε τους αρχικούς άσους και ομαδοποιούμε το 1ο--2ο μηδενικό, το 3ο--4ο, κ.ο.κ., μαζί με τους άσους που ακολουθούν το καθένα.

Δύο διαφορετικές ιδιότητες

Η **1** έχει άρτιο πλήθος μηδενικών (0), αλλά περιττό μήκος. Η **01** έχει άρτιο μήκος, αλλά περιττό πλήθος μηδενικών.

Φροντιστήριο: Γλώσσα $L = \{0, 10\}^*$

Η γλώσσα $\{0, 10\}^*$ περιγράφεται από την έκφραση

$$r = (0 + 10)^*.$$

Συνενώνουμε μηδέν ή περισσότερα κομμάτια 0 και 10:

$$0010010100 = 0 \cdot 0 \cdot 10 \cdot 0 \cdot 10 \cdot 10 \cdot 0.$$

Χαρακτηρισμός

Το $L(r)$ περιέχει ακριβώς την ε και τις μη κενές δυαδικές συμβολοσειρές που **τελειώνουν σε 0** και **δεν περιέχουν 11**.

Να αποδειχθεί ο χαρακτηρισμός: φροντιστήριο την Παρασκευή!

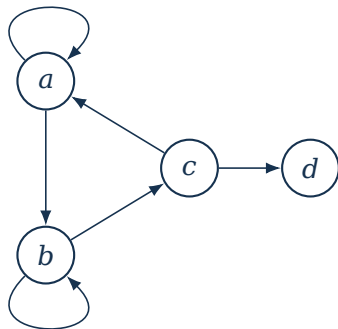
Αναπαράσταση κανονικών εκφράσεων με Γράφους Εισαγωγή στα αυτόματα

Κατευθυνόμενα γραφήματα και διαδρομές

Ένα κατευθυνόμενο
γράφημα/κατευθυνόμενος γράφος
 $G = (V, E)$ έχει κορυφές και
κατευθυνόμενες ακμές. Η ακμή (u, v)
πηγαίνει από το u στο v . Μια διαδρομή
 $(v_0, \dots, v_k)$ ακολουθεί ακμή $v_i \rightarrow v_{i+1}$ σε
κάθε βήμα.

(a, b, c, d) είναι διαδρομή,

ενώ η (a, c, d) δεν είναι: δεν υπάρχει ακμή
 $a \rightarrow c$.



Η σύμβαση που θα χρησιμοποιούμε

Οι διαδρομές επιτρέπεται να **επαναλαμβάνουν κορυφές και ακμές**. Έτσι μπορούμε να κάνουμε έναν κύκλο πολλές φορές.

Πώς μια διαδρομή «γράφει» μια συμβολοσειρά

Σε ένα γράφημα με ετικέτες, κάθε ακμή φέρει σύμβολο του $\Sigma \cup \{\varepsilon\}$. Η συμβολοσειρά μιας διαδρομής προκύπτει με συνένωση των ετικετών της.



$$\begin{array}{c} 0 \varepsilon 1 \\ \hline \{z\} \\ \text{ετικέτες ακμών} \end{array} = 01.$$

Τι σημαίνει μια ε -ακμή;

Την ακολουθούμε **χωρίς να προσθέσουμε σύμβολο** στη συμβολοσειρά. Το ε δεν είναι νέο σύμβολο του αλφαβήτου.

Το βέλος εισόδου δείχνει την αρχική κορυφή και ο διπλός κύκλος την τελική.

Από μια έκφραση σε ένα γράφημα

Για μια κανονική έκφραση r ξεκινάμε με δύο κορυφές και μία ακμή:



Τι σημαίνει προσωρινά μια σύνθετη ετικέτα;

Μια ακμή με ετικέτα f μπορεί να συνεισφέρει οποιαδήποτε συμβολοσειρά του $L(f)$.

Αντικαθιστούμε σταδιακά τις ακμές με ετικέτες

$$f + g, \quad fg, \quad f^*$$

μέχρι να απομείνουν μόνο ετικέτες από το $\Sigma \cup \{\varepsilon\}$. Οι ακμές με ετικέτα $\emptyset$ διαγράφονται.

Κανόνας 1: η ένωση γίνεται επιλογή

$$L(f + g) = L(f) \cup L(g).$$



Ερμηνεία

Για να πάμε από το u στο v , επιλέγουμε είτε μια λέξη του $L(f)$ είτε μια λέξη του $L(g)$.

Οι δύο νέες ακμές έχουν **ίδια άκρα**. Επιτρέπουμε παράλληλες ακμές με διαφορετικές ετικέτες.

Κανόνας 2: η συνένωση γίνεται διαδοχή

$$L(fg) = \{xy : x \in L(f), y \in L(g)\}.$$



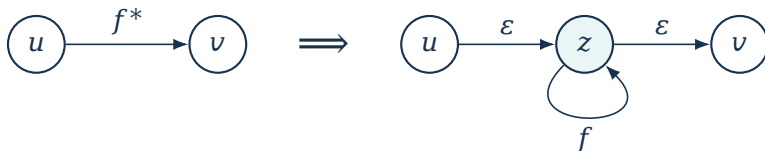
Ερμηνεία

Πρώτα διαβάζουμε μια λέξη $x \in L(f)$ και μετά μια λέξη $y \in L(g)$. Η συνολική ετικέτα είναι xy .

Η z είναι **νέα κορυφή**, χωρίς άλλες συνδέσεις στο υπάρχον γράφημα. Ο κανόνας ισχύει και αν η αρχική ακμή είναι βρόχος ($u = v$).

Κανόνας 3: το αστέρι γίνεται επανάληψη

$$L(f^*) = L(f)^*.$$



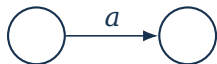
Η z είναι νέα κορυφή. Διασχίζουμε τον βρόχο 0, 1, 2, ... φορές:

$$\varepsilon, \quad x_1, \quad x_1 x_2, \quad \dots \quad (x_i \in L(f)).$$

Η περίπτωση μηδέν επαναλήψεων

Η διαδρομή $u \xrightarrow{\varepsilon} z \xrightarrow{\varepsilon} v$ γράφει ε . Δεν απαιτείται να διασχίσουμε τον βρόχο.

Οι βασικές ετικέτες και το τέλος της κατασκευής



Συνεισφέρει το σύμβολο $a \in \Sigma$.



Επιτρέπει μετάβαση χωρίς σύμβολο.



Η ακμή διαγράφεται: δεν παράγει καμία λέξη.

Τερματισμός

Κάθε αντικατάσταση αφαιρεί έναν τελεστή από τις ετικέτες. Άρα έπειτα από πεπερασμένα βήματα όλες οι ακμές του $G(r)$ έχουν ετικέτα στο $\Sigma \cup \{\varepsilon\}$.

Το θεώρημα που δίνει νόημα στην κατασκευή

Θεώρημα (Αναπαράσταση μιας κανονικής έκφρασης)

Για κάθε κανονική έκφραση r και κάθε συμβολοσειρά x ,

$$x \in L(r) \iff \begin{array}{l} \text{υπάρχει διαδρομή στο } G(r) \text{ από την αρχική} \\ \text{στην τελική κορυφή, της οποίας οι ετικέτες γράφουν } x. \end{array}$$

Τρεις σημαντικές λεπτομέρειες

Αρκεί **μία** τέτοια διαδρομή. Δεν χρειάζεται να είναι μοναδική! Οι κορυφές και οι ακμές μπορούν να επαναληφθούν. Οι ε -ακμές δεν προσθέτουν σύμβολα.

Για $x = \varepsilon$: αναζητούμε διαδρομή μόνο με ε -ακμές.

Γιατί η κατασκευή είναι σωστή;

Το αναλλοίωτο του παραδείγματος

Σε κάθε ενδιάμεσο γράφημα ισχύει:

$$x \in L(r) \iff \text{υπάρχει διαδρομή από την αρχική στην τελική κορυφή} \\ \text{με ετικέτες } r_1, \dots, r_k \text{ και } x \in L(r_1) \dots L(r_k).$$

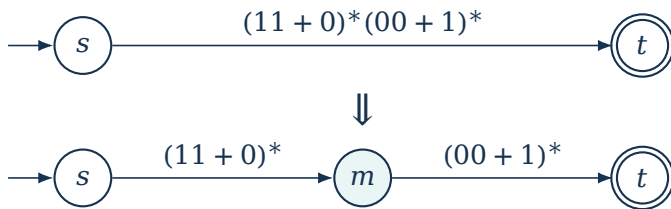
Αρχικά: υπάρχει μόνο η ακμή με ετικέτα r , οπότε η ισοδυναμία είναι άμεση.

Σε κάθε αντικατάσταση: η επιλογή, η διαδοχή και η επανάληψη διατηρούν ακριβώς τις λέξεις που μπορεί να συνεισφέρει το αντίστοιχο τμήμα. Ένας παλιός τρόπος διέλευσης μετατρέπεται σε νέο και αντίστροφα.

Στο τέλος: οι ετικέτες είναι σύμβολα ή ε . Η διαγραφή ακμών με ετικέτα $\emptyset$ δεν αφαιρεί καμία λέξη, αφού $L(\emptyset) = \emptyset$.

Παράδειγμα: Ξεκινάμε από τη συνένωση

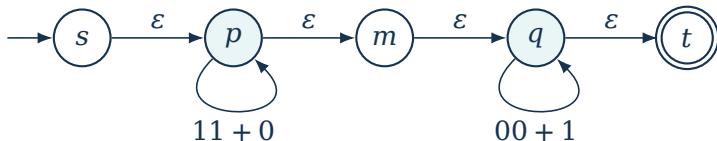
$$r = (11 + 0)^*(00 + 1)^*.$$



Ανάγνωση της έκφρασης

Πρώτα κομμάτια **11** ή **0**, και έπειτα κομμάτια **00** ή **1**. Οποιοδήποτε από τα δύο τμήματα μπορεί να είναι κενό.

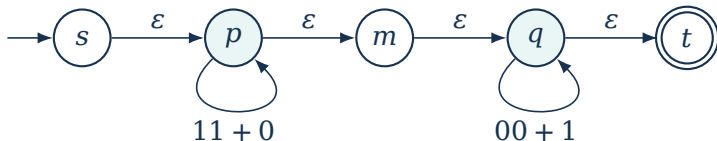
Παράδειγμα: αναπτύσσουμε τα δύο αστέρια



Οι δύο περιοχές επανάληψης

Στην p επαναλαμβάνουμε λέξεις του $L(11 + 0)$. Όταν φύγουμε προς τη q , επαναλαμβάνουμε λέξεις του $L(00 + 1)$.

Παράδειγμα: αναπτύσσουμε τα δύο αστέρια



Οι δύο περιοχές επανάληψης

Στην p επαναλαμβάνουμε λέξεις του $L(11 + 0)$. Όταν φύγουμε προς τη q , επαναλαμβάνουμε λέξεις του $L(00 + 1)$.

Απομένουν δύο ενώσεις και δύο συνενώσεις μέσα στις ετικέτες.

$11 + 0 \rightarrow$ βρόχος 11 ή βρόχος 0, $00 + 1 \rightarrow$ βρόχος 00 ή βρόχος 1.

Παράδειγμα: από τις ενώσεις στις απλές ακμές

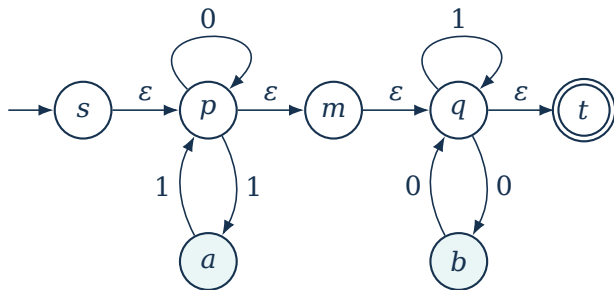
Στην αριστερή περιοχή:

ο βρόχος 11 γίνεται δύο ακμές με ετικέτα 1 και μία νέα κορυφή a . **Στη δεξιά**

περιοχή: ο βρόχος 00

γίνεται δύο ακμές με ετικέτα 0 και μία νέα κορυφή b . Τώρα κάθε

ακμή φέρει 0, 1 ή ε .



Η δομή παραμένει ίδια

Μία περιφορά $p \rightarrow a \rightarrow p$ γράφει **11**. Μία περιφορά $q \rightarrow b \rightarrow q$ γράφει **00**.

Πότε μπορούμε να συρρικνώσουμε μια ε -ακμή;

Ασφαλής κανόνας συρρίκνωσης (ικανές συνθήκες)

Η ακμή $u \xrightarrow{\varepsilon} v$ μπορεί να συρρικνωθεί σε μία κορυφή, αν:

- είναι η **μοναδική εξερχόμενη** ακμή της u και η u **δεν είναι τελική**, ή
- είναι η **μοναδική εισερχόμενη** ακμή της v και η v **δεν είναι αρχική**.

Η νέα κορυφή κληρονομεί τυχόν αρχική ή τελική ιδιότητα.



Τι πρέπει να δείξουμε: Τα δύο γραφήματα δέχονται ακριβώς τις ίδιες συμβολοσειρές **Τι θέλω τώρα από εσάς:** Να καταλάβετε τις συνθήκες, ΟΧΙ ΑΠΟΜΝΗΜΟΝΕΥΣΗ.

Δεν απομνημονεύουμε: ελέγχουμε τις νέες επιλογές

Για τη συγχώνευση των άκρων της $u \xrightarrow{\varepsilon} v$ αρκεί **μία ολόκληρη περίπτωση**:

A. Οι επιλογές στο u

Μοναδική έξοδος: μόνο $u \xrightarrow{\varepsilon} v$.

Μη τελικό u : δεν σταματώ εδώ.

Άρα μια αποδεκτή διαδρομή **πρέπει να συνεχίσει στο v** .

B. Η άφιξη στο v

Μοναδική είσοδος: ήρθα από το u .

Μη αρχικό v : δεν ξεκινώ εδώ.

Άρα **μόλις ήμουν στο u** , χωρίς να διαβάσω σύμβολο.

(A1 και A2) ή (B1 και B2)

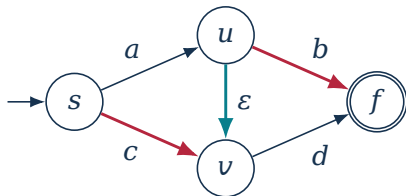
Η ερώτηση πριν από κάθε συγχώνευση

«Μπορεί τώρα να γίνει δεκτή μια λέξη που πριν δεν γινόταν;»

A1. Γιατί χρειάζεται μοναδική έξοδος από το u ;

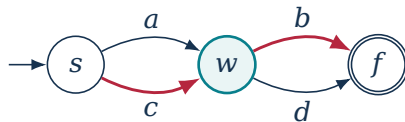
Το u **δεν είναι τελικό**, αλλά έχει και δεύτερη έξοδο: $u \xrightarrow{b} f$.

Πριν: G



$$L(G) = \{ab, ad, cd\}.$$

Μετά: G'



$$L(G') = L(G) \cup \{\text{cb}\}.$$

Νέα λέξη: cb

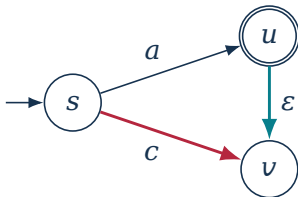
Πριν, το c μας έφερνε στο v : η έξοδος b ανήκε στο u . Τώρα επιτρέπεται $s \xrightarrow{c} w \xrightarrow{b} f$.

Η περίπτωση B δεν μας σώζει: στο v εισέρχονται δύο ακμές.

A2. Γιατί το u δεν πρέπει να είναι τελικό;

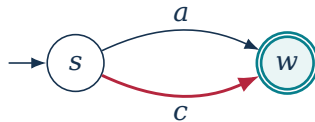
Η $u \xrightarrow{\varepsilon} v$ είναι η **μοναδική έξοδος** του u , αλλά το u είναι **τελικό**.

Πριν: G



$$L(G) = \{a\}.$$

Μετά: G'



$$L(G') = \{a, c\}.$$

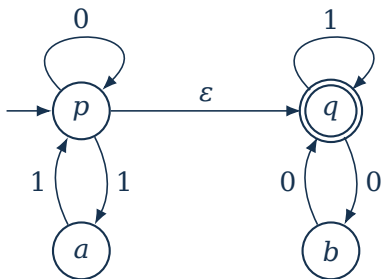
Ο τερματισμός είναι και αυτός επιλογή!

Στο u μπορούσαμε να **σταματήσουμε** χωρίς την ε -μετάβαση. Το τελικό w δίνει αυτή τη δυνατότητα και σε όποιον έφτανε στο v .

Η περίπτωση B δεν εφαρμόζεται: το v έχει δύο εισερχόμενες ακμές.

Τις υπόλοιπες θα τις αφήσουμε ως άσκηση, ώστε να δουλέψετε μόνοι σας τη λογική πίσω από τις συνθήκες του θεωρήματος. Στόχος δεν είναι να απομνημονεύσετε τις συνθήκες, αλλά να καταλάβετε γιατί χρειάζονται.

Παράδειγμα: το απλοποιημένο γράφημα



Συρρικνώσαμε τις επιτρεπτές ε -ακμές του προηγούμενου σχήματος. Η γλώσσα παραμένει

$$L((11 + 0)^*(00 + 1)^*).$$

Έλεγχοι:

$$\varepsilon \in L(r), \quad \mathbf{11001} \in L(r), \quad \mathbf{10} \notin L(r).$$

Για τη **11001**: κύκλος 11, μετά η ε -ακμή, κύκλος 00 και τέλος βρόχος 1.

Γιατί μένει η μεσαία ε -ακμή;

Καμία από τις δύο συνθήκες δεν ισχύει. Αν συγχωνεύαμε p, q , θα επιτρέπαμε μεταξύ άλλων τη **10**.

Την επόμενη εβδομάδα: πεπερασμένα αυτόματα

Κανονική έκφραση $\Rightarrow$ ε -NFA

Σήμερα το κατασκευάσαμε **ως γράφημα**. Στις επόμενες διαλέξεις θα το ορίσουμε ως **αυτόματο**:

καταστάσεις + μεταβάσεις + είσοδος + αποδοχή.

Την επόμενη εβδομάδα: πεπερασμένα αυτόματα

Κανονική έκφραση $\Rightarrow$ ε -NFA

Σήμερα το κατασκευάσαμε **ως γράφημα**. Στις επόμενες διαλέξεις θα το ορίσουμε ως **αυτόματο**:

καταστάσεις + μεταβάσεις + είσοδος + αποδοχή.

Και μετά;

Θα δούμε ότι τα

DFA, NFA, ε -NFA, κανονικές εκφράσεις περιγράφουν **ακριβώς την ίδια κλάση γλωσσών**: τις κανονικές γλώσσες.