# Κανονικές Γλώσσες

Διδάσκοντες: **Φ. Αφράτη, Δ. Φωτάκης**
Επιμέλεια διαφανειών: **Δ. Φωτάκης**

Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών

Εθνικό Μετσόβιο Πολυτεχνείο

# Κανονικές Γλώσσες

- Κανονική γλώσσα αν παράγεται από κανονική γραμματική.
- Παραγωγές $P \subseteq (V - \Sigma) \times \Sigma^*((V - \Sigma) \cup \varepsilon)$
  - Παραγωγές μορφής: $A \to w \mid wB,\ \ w \in \Sigma^*, A, B \in V - \Sigma$
  - Ένα μη-τερματικό αριστερά.
  - Στα δεξιά, το πολύ ένα μη-τερματικό στο τέλος.

$V = \{0, 1, A, B, S\}$,
$\Sigma = \{0, 1\}$,
$S$

| Παραγωγές $P_1$ |
|---|
| $S \to 1S \mid 0A$ |
| $A \to 1S \mid 0B$ |
| $B \to 1S \mid 0B \mid 0$ |

| Παραγωγές $P_2$ |
|---|
| $S \to 1S \mid 1 \mid 0A$ |
| $A \to 1A \mid 0B$ |
| $B \to 1B \mid 0B \mid 0$ |

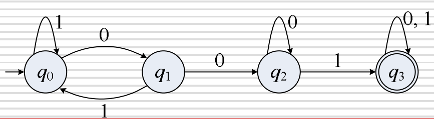
$V' = \{0, 1, S\}$,
$\Sigma = \{0, 1\}$,
$S$

| Παραγωγές $P_3$ |
|---|
| $S \to 01S \mid 010S \mid \varepsilon$ |



# Κανονικές Γλώσσες και Αυτόματα

- Αν γλώσσα αποφασίζεται από (D)FA, τότε παράγεται από κανονική γραμματική.
  - DFA $M(Q, \Sigma, \delta, s, F)$ και **L = L(M)**.
  - **Κανονική** γραμματική G για την L:
    - $V = Q \cup \Sigma$ (**μη-τερματικά** αντιστοιχούν σε **καταστάσεις**)
    - $S = s$
    - Παραγωγές $P = \{p \to \sigma q : \delta(p, \sigma) = q\} \cup \{f \to \varepsilon : f \in F\}$
  - Με επαγωγή αποδεικνύεται (εύκολα) ότι **L = L(G)**



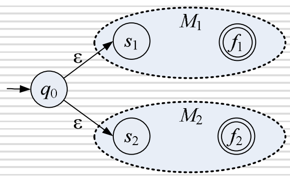



# Κανονικές Γλώσσες και Αυτόματα

- Κάθε κανονική γλώσσα αποφασίζεται από (N)FA.
  - Κανονική γραμματική G(V, Σ, S, P) που παράγει **L = L(G)**.
    - Χ.β.γ παραγωγές $A \to \varepsilon \mid \sigma \mid \sigma B,\ \ \sigma \in \Sigma, A, B \in V - \Sigma$
  - Μη-ντετερμινιστικό αυτόματο:
    - $Q = (V - \Sigma) \cup \{f\},\ s = S,\ F = \{f\}$
    - Σχέση μετάβασης Δ: $(A, \sigma, B) \in \Delta$, κανόνας $A \to \sigma B$
      $(A, \sigma, f) \in \Delta$, κανόνας $A \to \sigma$
      $(A, \varepsilon, f) \in \Delta$, κανόνας $A \to \varepsilon$
  - Με επαγωγή αποδεικνύεται (εύκολα) ότι **L = L(M)**
- Κλάση κανονικών γλωσσών ταυτίζεται με κλάση γλωσσών που αποφασίζονται από DFA (και NFA).

# Κλειστότητα

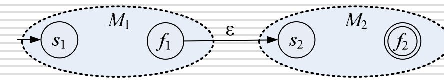
- Η κλάση των κανονικών γλωσσών είναι **κλειστή** ως προς:
  - Ένωση $L_1 \cup L_2$






# Κλειστότητα

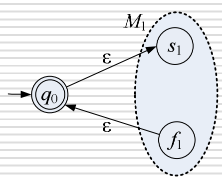
- Η κλάση των κανονικών γλωσσών είναι **κλειστή** ως προς:
  - Ένωση $L_1 \cup L_2$
  - Παράθεση $L_1 L_2$






# Κλειστότητα

- Η κλάση των κανονικών γλωσσών είναι **κλειστή** ως προς:
  - Ένωση $L_1 \cup L_2$
  - Παράθεση $L_1 L_2$
  - Kleene star $L^*$






# Κλειστότητα

- Η κλάση των κανονικών γλωσσών είναι **κλειστή** ως προς:
  - Ένωση $L_1 \cup L_2$
  - Παράθεση $L_1 L_2$
  - Kleene star $L^*$
  - Συμπλήρωμα
    - DFA και εναλλαγή τελικών – μη τελικών.
  - Τομή $L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}}$

# Κλειστότητα

- Η κλάση των κανονικών γλωσσών είναι **κλειστή** ως προς:
  - Τομή $L_1 \cap L_2$
  - Έστω DFA $M_1(Q_1, \Sigma, \delta_2, s_1, F_1)$ με $L_1 = L(M_1)$ και DFA $M_2(Q_1, \Sigma, \delta_2, s_2, F_2)$ με $L_2 = L(M_2)$.
  - Ορίζουμε DFA $M'(Q', \Sigma, \delta', s', F')$ με $L(M') = L_1 \cap L_2$
    - Καταστάσεις $Q' = Q_1 \times Q_2$
    - Αρχική κατάσταση $s' = (s_1, s_2)$
    - Τελικές καταστάσεις $F' = \{(p, q): p \in F_1$ **και** $q \in F_2\}$
    - Συνάρτηση μετάβασης $\delta((p, q), \sigma) = (\delta_1(p, \sigma), \delta_2(q, \sigma))$
  - Διαφορά $L(M_1) - L(M_2)$
    - Μόνη αλλαγή: Τελικές $= \{(p, q): p \in F_1$ **και** $q \notin F_2\}$



# Κανονικές Εκφράσεις

- Κανονική έκφραση αλφάβητου Σ :
  1. Το $\varnothing$ και κάθε $\sigma \in \Sigma$ είναι ΚΕ.
  2. Αν α και β είναι ΚΕ, τότε $(\alpha \cup \beta)$ είναι ΚΕ.
  3. Αν α και β είναι ΚΕ, τότε $(\alpha\,\beta)$ είναι ΚΕ.
  4. Αν α είναι ΚΕ, τότε $\alpha^*$ είναι ΚΕ.
  5. Τίποτα άλλο δεν είναι ΚΕ.
- Παραδείγματα ΚΕ του $\Sigma = \{0, 1\}$: (παραλείπουμε παρενθέσεις)
  - ε γιατί εκφράζεται σαν $\varnothing^*$
  - $(0 \cup 1)$, $(0 \cup 1)1^*$, $((0 \cup 1)1(0 \cup 1)0)^*$
  - $(10 \cup 01)^* \cup (0001 \cup 1000)^*$, $(1 \cup 1^*)^*$
  - $0^*10^*010^*(10^* \cup \varnothing^*)$



# ... και Αντίστοιχες Γλώσσες

- Αν $\alpha$ μια ΚΕ, $\mathcal{L}(\alpha)$ είναι η αντίστοιχη γλώσσα.
  1. $\mathcal{L}(\emptyset) = \emptyset$, $\mathcal{L}(\sigma) = \{\sigma\}$
  2. $\mathcal{L}((\alpha \cup \beta)) = \mathcal{L}(\alpha) \cup \mathcal{L}(\beta)$
  3. $\mathcal{L}((\alpha\beta)) = \mathcal{L}(\alpha)\mathcal{L}(\beta)$
  4. $\mathcal{L}(\alpha^*) = \mathcal{L}(\alpha)^*$
- Μια γλώσσα μπορεί να αναπαρίσταται από άπειρες κανονικές εκφράσεις.



# Παραδείγματα

- Ποια γλώσσα αντιστοιχεί στην ΚΕ: $(a \cup b)^*$

$$\begin{aligned} \mathcal{L}((a \cup b)^*) &= \mathcal{L}((a \cup b))^* && \text{(κανόνας 4)} \\ &= (\mathcal{L}(a) \cup \mathcal{L}(b))^* && \text{(κανόνας 2)} \\ &= (\{a\} \cup \{b\})^* && \text{(κανόνας 1)} \\ &= \{a, b\}^* \end{aligned}$$

- Ποια γλώσσα αντιστοιχεί στην ΚΕ: $((a \cup b)^*(ba))$

$$\begin{aligned} \mathcal{L}(((a \cup b)^*(ba))) &= \mathcal{L}((a \cup b)^*)\mathcal{L}((ba)) && \text{(κανόνας 3)} \\ &= \mathcal{L}((a \cup b))^*(\mathcal{L}(b)\mathcal{L}(a)) && \text{(κανόνες 4 και 3)} \\ &= (\mathcal{L}(a) \cup \mathcal{L}(b))^*\{b\}\{a\} && \text{(κανόνες 2 και 1)} \\ &= \{a, b\}^*\{ba\} && \text{(κανόνας 1)} \end{aligned}$$

## Παραδείγματα

- Ποια γλώσσα αντιστοιχεί στην ΚΕ: $(a \cup (ba))^*$

$$
\begin{aligned}
\mathcal{L}((a \cup (ba))^*) &= \mathcal{L}((a \cup (ba)))^* && (\text{κανόνας } 4)\\
&= (\mathcal{L}(a) \cup \mathcal{L}((ba)))^* && (\text{κανόνας } 2)\\
&= (\mathcal{L}(a) \cup (\mathcal{L}(b)\mathcal{L}(a)))^* && (\text{κανόνας } 3)\\
&= (\{a\} \cup (\{b\}\{a\}))^* && (\text{κανόνας } 1)\\
&= \{a, ba\}^*
\end{aligned}
$$

  - Κάθε $b$ ακολουθείται από $a$
    (ή δεν περιέχει δύο συνεχόμενα $b$).
- Γλώσσα $(\varepsilon \cup 1 \cup 11)(0 \cup 01 \cup 011)^*$
  - Δεν περιέχει 111.



## Παραδείγματα

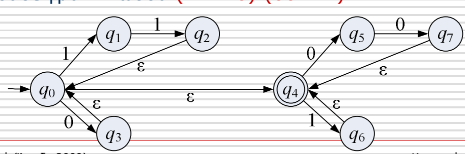
- ΚΕ για $L = \{w \in \{a,b\}^* : w \text{ περιέχει } ba \text{ και τελειώνει σε } b\}$
  $(a \cup b)^* ba (a \cup b)^* b$
- ΚΕ για $L = \{w \in \{0,1\}^* : w \text{ περιέχει ζυγό αριθμό } 0\}$
  $(1^*01^*0)^*1^*$ ή $(1 \cup 01^*0)^*$
- $L = \{w \in \{0,1\}^* : w \text{ τελειώνει με 1 και δεν περιέχει } 00\}$
  $(1 \cup 01)(1 \cup 01)^*$
- ΚΕ για $L = \{w \in \{0,1\}^* : w \text{ περιέχει μία εμφάνιση του } 00\}$
  $(1 \cup 01)^* 00 (1 \cup 10)^*$
- ΚΕ για $L = \{w \in \{0,1\}^* : w \text{ δεν περιέχει } 110\}$
  $(0 \cup 10)^* 1^*$



## Κανονικές Εκφράσεις και Αυτόματα

- Κάθε γλώσσα που αναπαρίσταται από κανονική έκφραση αποφασίζεται από (N)FA (και είναι κανονική).
  - Απόδειξη με επαγωγή στην πολυπλοκότητα της ΚΕ.
  - Βάση: ορίζουμε (στοιχειώδη) NFA για $\emptyset$, $\{\sigma\}$, και $\{\varepsilon\}$.


  - Βήμα: για ένωση, παράθεση, * βλ. προηγούμενες κατασκευές.
  - Παράδειγμα: Γλώσσα $(11 \cup 0)^*(00 \cup 1)^*$






## Κανονικές Εκφράσεις και Αυτόματα

- Κάθε κανονική γλώσσα (που αποφασίζεται από (D)FA) αναπαρίσταται από κανονική έκφραση.
  - Αλγόριθμος δυναμικού προγραμματισμού (παρόμοιος με αλγόριθμο Warshall) παράγει κανονική έκφραση από περιγραφή DFA.
- Τα παρακάτω είναι ισοδύναμα:
  - Μια γλώσσα παράγεται από κανονική γραμματική.
  - Μια γλώσσα αναπαρίσταται από κανονική έκφραση.
  - Μια γλώσσα αποφασίζεται από DFA
  - Μια γλώσσα αποφασίζεται από NFA.

# ΚΕ και DFA

- Κάθε γλώσσα που γίνεται δεκτή από DFA αναπαρίσταται από κανονική έκφραση.
  - Έστω $M(Q, \Sigma, \delta, s, F)$ με $Q = \{ q_1, \ldots, q_n \}$ και $s = q_1$
  - Θα ορίσουμε κανονική έκφραση $R$: $L(R) = L(M)$.
  - Για κάθε $i, j = 1, \ldots, n$ και $k = 0, \ldots, n$, ορίζουμε

$R^{[k]}(i, j) =$ σύνολο συμβ/ρών $\Sigma^*$ που οδηγούν $M$ από $q_i$ σε $q_j$
χωρίς να περάσει από ενδιάμεση κατάσταση με δείκτη $> k$
Αρχική $q_i$ και τελική $q_j$ μπορούν να έχουν δείκτη $> k$

$$R^{[n]}(i, j) = \{ w \in \Sigma^* : (q_i, w) \vdash_M^* (q_j, \varepsilon) \}$$

$$L(M) = \bigcup \{ R^{[n]}(1, j) : q_j \in F \}$$

  - Αν $R^{[k]}(i, j)$ αναπαρίστανται από κανονικές εκφράσεις, $L(M)$ αναπαρίσταται από κανονική έκφραση (ένωση).



# ΚΕ και DFA

- $R^{[k]}(i, j)$ με κανονικές εκφράσεις: επαγωγή στο $k$.

$$R^{[0]}(i, j) = \begin{cases} \{ \sigma \in \Sigma : \delta(q_i, \sigma) = q_j \} & \text{αν } i \neq j \\ \{\varepsilon\} \cup \{ \sigma \in \Sigma : \delta(q_i, \sigma) = q_i \} & \text{αν } i = j \end{cases}$$

$$R^{[k]}(i, j) = R^{[k-1]}(i, j) \cup R^{[k-1]}(i, k)\, R^{[k-1]}(k, k)^*\, R^{[k-1]}(k, j)$$

  - Επίσης βλ. αλγόριθμο Floyd-Warshall για συντομότερα μονοπάτια και μεταβατική κλειστότητα.
  - Δυναμικός προγραμματισμός (bottom-up):
    - Συνδυάζουμε γλώσσες που αποδέχονται περιορισμένα τμήματα του αυτομάτου για να βρούμε γλώσσα που αποδέχεται όλο το αυτόματο.



# Παρατηρήσεις

- Κάθε πεπερασμένη γλώσσα είναι κανονική.
  - Απόδειξη (εύκολη) με επαγωγή στον #συμβ/ρών.
- Κανονική γλώσσα που αναπαρίσταται από ΚΕ χωρίς * (ισοδύναμα από DFA χωρίς «κύκλο»);
  - Είναι πεπερασμένη.
  - Ο τελεστής * (ισοδύναμα, ο κύκλος στο DFA) μοναδικός τρόπος δημιουργίας άπειρου πλήθους συμβ/ρών.



# Δύο Χαρακτηριστικά Κανονικών Γλωσσών

- Η μνήμη για αναγνώριση συμβ/ράς κανονικής γλώσσας εξαρτάται από γλώσσα **αλλά όχι** συμβολοσειρά.
  - Ένδειξη ότι $\{ 1^n 0^n : n \geq 0 \}$ **δεν** είναι κανονική.
- Άπειρες κανονικές γλώσσες: DFA με κύκλους.
  - Απλή επαναληπτική δομή / περιοδικότητα.
  - Ένδειξη ότι $\{ 1^n : n \text{ πρώτος} \}$ **δεν** είναι κανονική.
- Ανάγκη (μαθηματικού) εργαλείου για **απόδειξη** ότι γλώσσα είναι μη κανονική.

# Λήμμα Άντλησης (Pumping Lemma)

- Έστω L μια άπειρη κανονική γλώσσα. Υπάρχει $k \geq 1$ τ.ω. κάθε $w \in L$, $|w| \geq k$, γράφεται $w = xyz$, $y \neq \varepsilon$, $|xy| \leq k$, και $xy^n z \in L$ για κάθε $n \geq 0$.
  - DFA M(Q, Σ, δ, s, F) με $k$ καταστάσεις που αποφασίζει L.
  - Έστω $w \in L : |w| = \ell \geq k$. Υπολογισμός M(w):
    $(q_0, w_1 w_2 \ldots w_\ell) \vdash_M (q_1, w_2 \ldots w_\ell) \vdash_M \cdots \vdash_M (q_\ell, \varepsilon)$
  - … διέρχεται από ίδια κατάσταση τουλάχιστον 2 φορές. Υπάρχουν $0 \leq i < j \leq k$, ώστε $q_i = q_j$ .
  - Έστω $x = w_1 \ldots w_i, y = w_{i+1} \ldots w_j, z = w_{j+1} \ldots w_\ell$
  - M δέχεται $xy^n z$ : x οδηγεί M στην $q_i$ , $y^n$ κάνει n κύκλους με αφετηρία και κατάληξη $q_i$ , και z οδηγεί M σε τελική.



# Μη Κανονικότητα Γλωσσών

- Ν.δ.ο. $L = \{ 0^n 1^n : n \geq 0\}$ **δεν** είναι κανονική.
  - L άπειρη. Άρα αν είναι κανονική, ισχύει το Λ. Άντλησης.
  - Αν L κανονική, υπάρχει $k > 0$, τ.ω. για κάθε $w \in L$, $|w| \geq k$, υπάρχουν x, y, z συμ/ρές, $w = xyz$, $|xy| \leq k$, $y \neq \varepsilon$, και $x y^n z \in L$ για κάθε $n \geq 0$.
  - Πρέπει $x y z \in L$ (n = 1).
    - Αν y αποτελείται μόνο από 0, $x y^0 z \notin L$ (πιο πολλά 1).
    - Αν y αποτελείται μόνο από 1, $x y^0 z \notin L$ (πιο πολλά 0).
    - Αν y αποτελείται από 0 ακολουθούμενα από 1, $x y^2 z \notin L$ (γιατί $0^+1^+0^+1^+ \ldots$ ).
  - Άρα L δεν είναι κανονική.



# Μη Κανονικότητα Γλωσσών

- Ν.δ.ο. $L = \{1^n : n$ πρώτος$\}$ **δεν** είναι κανονική.
  - L άπειρη. Άρα αν είναι κανονική, ισχύει το Θ. Άντλησης.
  - Αν L κανονική, υπάρχουν x, y, z συμ/ρές, $y \neq \varepsilon$, ώστε $x y^n z \in L$ για κάθε $n \geq 0$.
    - Δηλαδή $|x| + n|y| + |z|$ **πρώτος**.
    - **Δεν** ισχύει για **n =** $|x| + 2|y| + |z| + 2$, αφού:

$$|x| + (|x| + 2|y| + |z| + 2)|y| + |z| = (|x| + 2|y| + |z|)(|y| + 1)$$

    που δεν είναι πρώτος!
  - L **δεν** είναι κανονική.



# Μη Κανονικότητα Γλωσσών

- Ν.δ.ο. $L = \{w \in \{0, 1\}^*$: w έχει ίδιο αριθμό 0 και 1$\}$ **δεν** είναι κανονική.
  - $L \cap 0^*1^* = \{ 0^n 1^n : n \geq 0\}$, που **δεν** είναι κανονική.
  - $0^*1^*$ κανονική.
  - Κανονικές γλώσσες κλειστές ως προς τομή.
  - Άρα L δεν είναι κανονική.