Πεπερασμένα Αυτόματα και Κανονικές Γλώσσες

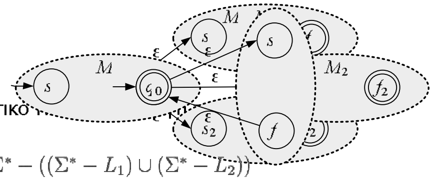
Δημήτρης Φωτάκης

Τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων

Πανεπιστήμιο Αιγαίου

Κλειστότητα

- Η κλάση των γλωσσών που γίνονται δεκτές από πεπερασμένα αυτόματα είναι **κλειστή** ως προς:
  - Ένωση $L_1 \cup L_2$
  - Παράθεση $L_1 L_2$
  - Kleene star L^*
  - Συμπλήρωμα
    - Ντετερμινιστικό
  - Τομή
    $L_1 \cap L_2 = \Sigma^* - ((\Sigma^* - L_1) \cup (\Sigma^* - L_2))$



Κλειστότητα

- Η κλάση των γλωσσών που γίνονται δεκτές από πεπερασμένα αυτόματα είναι **κλειστή** ως προς:
  - Τομή $L(M_1) \cap L(M_2)$
    - Καρτεσιανό γινόμενο καταστάσεων.
    - Αρχική: (s_1, s_2)
    - Τελικές = $\{(p, q)$: $p \in F_1$ **και** $q \in F_2\}$
    - $\delta((p, q), \sigma) = (\delta(p, \sigma), \delta(q, \sigma))$
  - 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 $\emptyset$ και κάθε $\sigma \in \Sigma$ είναι ΚΕ.
  2. Αν α και β είναι ΚΕ, τότε (α ∪ β) είναι ΚΕ.
  3. Αν α και β είναι ΚΕ, τότε (α β) είναι ΚΕ.
  4. Αν α είναι ΚΕ, τότε α* είναι ΚΕ.
  5. Τίποτα άλλο δεν είναι ΚΕ.

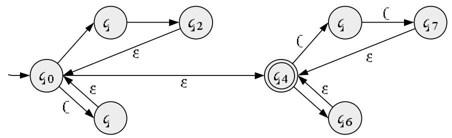
Κανονικές Γλώσσες

- Γλώσσες που αναπαριστώνται από ΚΕ.
- Αν α μια ΚΕ, $\mathcal{L}(\alpha)$ είναι η αντίστοιχη κανονική 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. $\mathcal{L}(\alpha^*) = \mathcal{L}(\alpha)^*$
- Ελάχιστη κλάση που περιέχει ∅, {σ}, ∀σ ∈ Σ, και είναι κλειστή ως προς ένωση, παράθεση, Kleene star (*).
- Θ.δ.ο. γλώσσα είναι κανονική **ανν** γίνεται δεκτή από πεπερασμένο αυτόματο.

ΚΕ και ΠΑ

- Κανονικές γλώσσες γίνονται δεκτές από πεπερασμένα αυτόματα.
  - Ορίζουμε (στοιχειώδη) ΠΑ που δέχονται ∅, {σ}, και {ε}. Για ένωση, παράθεση, *: προηγούμενες κατασκευές.

  - Παράδειγμα: Γλώσσες (11 ∪ 0)*(00 ∪ 1)* , (10 ∪ 110)*



ΚΕ και ΠΑ

- Κάθε γλώσσα που γίνεται δεκτή από πεπερασμένο αυτόματο είναι κανονική.
  - Έστω M(K, Σ, δ, s, F) με K = { q_1, …, q_n } και s = q_1
  - Θα ορίσουμε κανονική έκφραση *R*: *L*(*R*) = *L*(*M*).
  - Για κάθε *i*, *j* = 1, …, *n* και *k* = 0, …, *n*, ορίζουμε

$$R(i,j,k) = \text{σύνολο συμβ/ρών } \Sigma^* \text{ που οδηγούν } M \text{ από } q_i \text{ σε } q_j$$
$$\text{χωρίς να περάσει από ενδιάμεση κατάσταση με δείκτη} > k$$
$$\text{Αρχική } q_i \text{ και τελική } q_j \text{ μπορούν να έχουν δείκτη} > k$$
$$R(i,j,n) = \{w \in \Sigma^* : (q_i, w) \vdash_M^* (q_j, \varepsilon)\}$$
$$L(M) = \bigcup\{R(1,j,n) : q_j \in F\}$$

  - Αν *R*(*i*, *j*, *k*) κανονικές, *L*(*M*) κανονική (ένωση κανονικών).

ΚΕ και ΠΑ

- *R*(*i*, *j*, *k*) κανονικές: επαγωγή στο *k* με βάση ορισμό.

$$R(i,j,0) = \begin{cases} \{\sigma \in \Sigma : \delta(q_i,\sigma) = q_j\} & \text{αν } i \neq j \\ \{\varepsilon\} \cup \{\sigma \in \Sigma : \delta(q_i,\sigma) = q_j\} & \text{αν } i = j \end{cases}$$
$$R(i,j,k) = R(i,j,k-1) \cup R(i,k,k-1)R(k,k,k-1)^* R(k,j,k-1)$$

  - 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.

Δύο Χαρακτηριστικά Κανονικών Γλωσσών

- Η μνήμη για αναγνώριση συμβ/ράς κανονικής γλώσσας εξαρτάται από γλώσσα **αλλά όχι** συμβολοσειρά.
  - Ένδειξη ότι $\{1^n0^n : n > 0\}$ **δεν** είναι κανονική.
- Άπειρες κανονικές γλώσσες: ΠΑ με κύκλους (*).
  - Απλή επαναληπτική δομή / περιοδικότητα.
  - Ένδειξη ότι $\{1^n$: n πρώτος$\}$ **δεν** είναι κανονική.
- Ανάγκη (μαθηματικού) εργαλείου για **απόδειξη** ότι γλώσσα είναι μη κανονική.

Θεώρημα Άντλησης (Pumping Lemma)

- Έστω L μια άπειρη κανονική γλώσσα. Υπάρχει $k > 1$ τ.ω. κάθε $w \in L$, $|w| > k$, γράφεται $w = xyz$, $y \neq \varepsilon$, $|xy| \leq k$, και $xy^nz \in L$ για κάθε $n > 0$.
  - Ντετ. ΠΑ M(K, Σ, δ, s, F) με k καταστάσεις που δέχεται L.
  - Έστω $w \in L : |w| = \ell \geq k$. Υπολογισμός M(w):
    $$(q_0, w_1w_2 \dots w_\ell) \vdash_M (q_1, w_2 \dots w_\ell) \vdash_M \cdots \vdash_M (q_\ell, \varepsilon)$$
  - … διέρχεται από ίδια κατάσταση τουλάχιστον 2 φορές. Υπάρχουν $0 < i < j < k$, ώστε $q_i = q_j$.
  - Έστω $x = w_1 \dots w_i, y = w_{i+1} \dots w_j, z = w_{j+1} \dots w_\ell$
  - M δέχεται xy^nz : 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

- Έστω L μια **άπειρη** κανονική γλώσσα. Υπάρχουν συμβ/ρές x, y, z με $y \neq \varepsilon$ ώστε $xy^nz \in L$ για κάθε $n > 0$.
- Ν.δ.ο. $L = \{1^n0^n : n > 0\}$ **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 $n > 0$.
  - 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 $0^+1^+0^+1^+ \dots$).
  - 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 $n > 0$.
    - Δηλαδή $|x| + n|y| + |z|$ **πρώτος**.
    - **Δεν** ισχύει για $\mathbf{n} = |x| + 2|y| + |z| + 2$, 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 ∈ {0, 1}*: w έχει ίδιο αριθμό 0 και 1} **δεν** είναι κανονική.
  - $L \cap 1^*0^* = \{1^n0^n : n > 0\}$, που **δεν** είναι κανονική.
  - 1^*0^* κανονική.
  - Κανονικές γλώσσες κλειστές ως προς τομή.
  - Άρα L δεν είναι κανονική.

Αλγόριθμοι για ΠΑ

- Υπάρχουν αλγόριθμοι που απαντούν στα παρακάτω ερωτήματα για ΠΑ:
  - Δίνεται ΝΠΑ M. Ισοδύναμο ντετερμινιστικό ΠΑ;
  - Δίνεται ΝΠΑ M. Κανονική έκφραση γλώσσας L(M);
  - Κανονική έκφραση R. ΝΠΑ M με L(R) = L(M);
  - Δίνεται ΠΑ M και συμβ/ρά w. $w \in L(M)$;
  - Δίνεται ΠΑ M. $L(M) = \emptyset$; ($L(M) = \Sigma^*$;)
  - Δίνεται ΠΑ M. $L(M) = \infty$;
  - Δίνονται ΠΑ M_1 και M_2. $L(M_1) \subseteq L(M_2)$;
  - Δίνονται ΠΑ M_1 και M_2. $L(M_1) = L(M_2)$;