Γλώσσες χωρίς Συμφραζόμενα

Δημήτρης Φωτάκης

Τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων

Πανεπιστήμιο Αιγαίου

Γραμματικές

- **Αυτόματα** αναγνωρίζουν γλώσσες
  («αλγοριθμική» αναπαράσταση).
- **Γραμματικές** παράγουν γλώσσες
  («παραγωγική» αναπαράσταση – επίσης βλ. κανονικές εκφράσεις).
  - Ξεκινούν από αρχικό σύμβολο.
  - Προχωρούν εφαρμόζοντας κανόνες (όσο είναι δυνατόν).
  - Επιλογή κανόνων: όχι πλήρως καθορισμένη λειτουργία.
    (ίδια γραμματική παράγει πολλές συμβ/ρές).
  - Δύσκολο αυστηρή γραμματική για «φυσικές» γλώσσες.
  - Αυστηρή γραμματική για γλώσσες προγραμματισμού.
  - Γραμματικές **χωρίς συμφραζόμενα**.

Γραμματική

- … είναι μία τετράδα $G(V, \Sigma, R, S)$ όπου:
  - V : **αλφάβητο**
  - $\Sigma \subseteq V$: σύνολο **τερματικών** συμβόλων.
    - V – Σ : **μη τερματικά** σύμβολα, μόνο σε κανόνες.
    - Συμβολοσειρές γλώσσας: μόνο τερματικά σύμβολα.
  - $R \subseteq V^* \times V^*$: σύνολο **κανόνων** (μεταγραφή μη-τερματικών).
  - $S \in V - \Sigma$: **αρχικό** (μη-τερματικό) σύμβολο.

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

Κανόνες R
$S \to 0A1$
$A1 \to 0A11$
$0A \to 00A1$
$A \to \varepsilon$

Γραμματική χωρίς Συμφραζόμενα

- Κανόνες $R \subseteq (V - \Sigma) \times V^*$
  (μόνο ένα τερματικό σύμβολο στα αριστερά).
  - Κανόνες εφαρμόζονται ανεξάρτητα από συμφραζόμενα του τερματικού συμβόλου (context-free).
  - Ξεκινάμε με S
  - Ενόσω συμβ/ρά έχει μη-τερματικό A:
    - Κανόνα $A \to w$, $w \in V^*$
    - Αντικατάσταση A με w.

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

Κανόνες R
$S \to 0S1$
$S \to \varepsilon$

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

Κανόνες R
$S \to 0S0 \mid 1S1$
$S \to 0 \mid 1 \mid \varepsilon$

Γλώσσα χωρίς Συμφραζόμενα

- Έστω u, v, w $\in$ V* για γραμματική G(V, Σ, R, S).
- Αν A $\to$ w κανόνας, τότε uAv παράγει uwv : $uAv \Rightarrow uwv$
- Σχέση u παράγει v σε μηδέν ή περισσότερα βήματα $u \Rightarrow^* v$ αν υπάρχουν $k \geq 0$ και $u_1, \ldots, u_k$, $u_k = v$, ώστε
  $u \Rightarrow u_1 \Rightarrow \cdots \Rightarrow u_k = v$ (παραγωγή *k* βημάτων)
- **Γλώσσα G** : σύνολο συμβ/ρών Σ* που παράγονται από αρχικό σύμβολο S (σε πεπερασμένο #βημάτων).
  $$L(G) = \{w \in \Sigma^* : S \Rightarrow^* w\}$$
- **Γλώσσα L χωρίς συμφραζόμενα** αν υπάρχει γραμματική G χωρίς συμφραζόμενα: **L = L(G).**

Παράδειγμα

- Γραμματική χωρίς συμφραζόμενα G(V, Σ, R, S):

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

Κανόνες R
$S \to 0S1$
$S \to \varepsilon$

- Γλώσσα $L(G) = \{0^n 1^n : n \geq 0\}$
- Υπάρχουν γλώσσες χωρίς συμφραζόμενα που **δεν** είναι κανονικές.

Παράδειγμα

- Γραμματική χωρίς συμφραζόμενα G(V, Σ, R, S):

$V = \{(,), S\}$, $\Sigma = \{(,)\}$, S

Κανόνες R
$S \to (S) \mid SS$
$S \to \varepsilon$

$$L(G) = \{w \in \{(,)\}^* : w \text{ έχει σωστά ζυγισμένες παρενθέσεις}\}$$

Παράδειγμα

- Γραμματική χωρίς συμφραζόμενα για γλώσσα $L = \{u \in \{0, 1\}^* : u = ww^R\}$ (παλίνδρομα άρτιου μήκους).

Κανόνες R
$S \to 0S0 \mid 1S1 \mid \varepsilon$

- Γραμματική χωρίς συμφραζόμενα για γλώσσα όλων των συμβ/ρών που **δεν** είναι παλινδρομικές.

Κανόνες R
$S \to 0S0 \mid 1S1 \mid A$
$A \to 1B0 \mid 0B1$
$B \to 0B \mid 1B \mid \varepsilon$

Εκφραστικότητα

- Γλώσσες χωρίς συμφραζόμενα περιγράφουν:
  - Μικρά υποσύνολα φυσικών γλωσσών.
  - Αριθμητικές εκφράσεις.
  - **Γλώσσες προγραμματισμού** (C, C++, Pascal, …).
  - Κανονικές γλώσσες.
    - Κλάση κανονικών γλωσσών είναι **γνήσιο υποσύνολο** κλάσης γλωσσών χωρίς συμφραζόμενα.

$V = \{E, T, F, (,), \times, +, x, y, z, \ldots\}$
$\Sigma = \{(,), \times, +, x, y, z, \ldots\}$
E

Κανόνες R
$E \to T \mid E + T$
$T \to F \mid T \times F$
$F \to (E) \mid x \mid y \mid z \mid \ldots$

Κανονική Γραμματική

- Κανόνες R ⊆ (V − Σ) × Σ*((V − Σ) ∪ ε)
  - Κανόνες μορφής: $A \to w \mid wB,\ w \in \Sigma^*, A, B \in V - \Sigma$
  - Χωρίς συμφραζόμενα: μόνο ένα μη-τερματικό αριστερά.
  - Στα δεξιά, το πολύ ένα μη-τερματικό στο τέλος.

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

Κανόνες R
$S \to 1S \mid 0A$
$A \to 1S \mid 0F$
$F \to 1S \mid 0F \mid \varepsilon$

$(1 \cup 0)^*00$

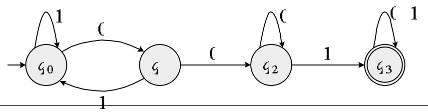
$V = \{0, 1, S\}$, $\Sigma = \{0, 1\}$, S

Κανόνες R
$S \to 01S \mid 010S \mid \varepsilon$

$(01 \cup 010)^*$

Κανονικές Γραμματικές και Κανονικές Γλώσσες

- Αν μια γλώσσα είναι κανονική, τότε παράγεται από κανονική γραμματική.
  - Ντετερμινιστικό ΠΑ M(K, Σ, δ, s, F) που αναγνωρίζει ΚΓ L.
  - Κανονική γραμματική G:
    - V = K ∪ Σ (**μη-τερματικά** αντιστοιχούν σε **καταστάσεις**)
    - S = s
    - Κανόνες $R = \{p \to \sigma q : \delta(p, \sigma) = q\} \cup \{f \to \varepsilon : f \in F\}$
  - Με επαγωγή αποδεικνύεται (εύκολα) ότι **L = L(G)**



Κανονικές Γραμματικές και Κανονικές Γλώσσες

- Αν μια γλώσσα παράγεται από κανονική γραμματική, τότε είναι κανονική.
  - Κανονική γραμματική G(V, Σ, R, S) που παράγει L.
    - Χ.β.γ κανόνες $A \to \varepsilon \mid \sigma \mid \sigma B,\ \sigma \in \Sigma, A, B \in V - \Sigma$
  - Μη-ντετερμινιστικό αυτόματο:
    - $K = (V - \Sigma) \cup \{f\},\ s = S,\ F = \{f\}$
    - 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)**

Κλειστότητα

- Οι γλώσσες χωρίς συμφραζόμενα είναι κλειστές ως προς ένωση και παράθεση (και Kleene star).
  - Έστω ΓχΣ $G_1(V_1, \Sigma_1, R_1, S_1)$ και $L_1 = L(G_1)$.
    Έστω ΓχΣ $G_2(V_2, \Sigma_2, R_2, S_2)$ και $L_2 = L(G_2)$.
    Θεωρούμε ότι $(V_1 - \Sigma_1) \cap (V_2 - \Sigma_2) = \emptyset$
  - Ένωση $L_1 \cup L_2$:
    - Νέο αρχικό σύμβολο S, νέοι κανόνες $S \to S_1 \mid S_2$
    - $G(V_1 \cup V_2 \cup \{S\}, \Sigma_1 \cup \Sigma_2, R_1 \cup R_2 \cup \{S \to S_1 \mid S_2\}, S)$
  - Παράθεση $L_1\, L_2$:
    - Νέο αρχικό σύμβολο S, νέος κανόνας $S \to S_1S_2$
    - $G(V_1 \cup V_2 \cup \{S\}, \Sigma_1 \cup \Sigma_2, R_1 \cup R_2 \cup \{S \to S_1S_2\}, S)$

Κλειστότητα

- Οι γλώσσες χωρίς συμφραζόμενα είναι κλειστές ως προς Kleene star (και ένωση και παράθεση).
  - Έστω ΓχΣ G(V, Σ, R, S) και L = L(G).
  - Kleene star L^*:
    - Νέο αρχικό σύμβολο S’, νέοι κανόνες $S' \to S'\, S \mid \varepsilon$
    - $G(V \cup \{S'\}, \Sigma, R \cup \{S' \to S'\, S \mid \varepsilon\}, S')$
- Άσκηση: Ν.δ.ο. $L = \{a^i b^j c^k : j = i + k\}$ είναι ΓχΣ.
  - Έστω $L_1 = \{a^i b^i : i \geq 0\}$ και $L_2 = \{b^k c^k : k \geq 0\}$
  - $L = L_1\, L_2$ (παράθεση των δύο γλωσσών).

Αριστερότερες Παραγωγές

- Έστω u, v, w $\in V^*$ για γραμματική G(V, Σ, R, S).
- Αν $A \to w$ κανόνας, τότε uAv παράγει uwv : $uAv \Rightarrow uwv$
- Σχέση u παράγει v σε μηδέν ή περισσότερα βήματα $u \Rightarrow^* v$ αν υπάρχουν $k \geq 0$ και $u_1, \ldots, u_k$, $u_k = v$, ώστε
  $u \Rightarrow u_1 \Rightarrow \cdots \Rightarrow u_k = v$ (παραγωγή *k* βημάτων)
- Αριστερότερη παραγωγή: σε κάθε βήμα το αριστερότερο μη-τερματικό αντικαθίσταται με κανόνα.
- Αν $u \Rightarrow^* v$ υπάρχει τουλάχιστον μία αριστερότερη παραγωγή με την οποία από το u παράγεται το v.

Διφορούμενες Γραμματικές

- Αν $u \Rightarrow^* v$ υπάρχει τουλάχιστον μία αριστερότερη παραγωγή με την οποία από το u παράγεται το v.
- Αν για κάποια u, v υπάρχουν δύο ή περισσότερες αριστερότερες παραγωγές, η γραμματική λέγεται **διφορούμενη** ή **ασαφής**.
- Διφορούμενη γραμματική: $V = \{(,), S\}$, $\Sigma = \{(,)\}$

Κανόνες R
$S \to (S) \mid SS \mid \varepsilon$

  - Δύο αριστερότερες παραγωγές για ( ).

$$S \Rightarrow (S) \Rightarrow ()$$
$$S \Rightarrow SS \Rightarrow S \Rightarrow (S) \Rightarrow ()$$

Διφορούμενες Γραμματικές

- Διφορούμενες γραμματικές: συντακτική ανάλυση;
  - Μετατρέπουμε τη διφορούμενη γραμματική σε ισοδύναμη μη-διφορούμενη.
  - Π.χ. $V = \{(,), S\}$, $\Sigma = \{(,)\}$

Κανόνες R
$S \to (S)S \mid \varepsilon$

  - **Εγγενώς διφορούμενες** γραμματικές **δεν** μπορούν να μετατραπούν σε μη-διφορούμενες.
  - Γραμματικές γλωσσών προγραμματισμού **δεν** είναι διφορούμενες.