Γενικά περί γραμματικών

Οι γραμματικές είναι συστήματα που παράγουν γλώσσες.

Γραμματική: Γραμματική (γενικά) είναι μια τετράδα $G(V, \Sigma, R, S)$ όπου

- V είναι αλφάβητο,
- Σ είναι επίσης αλφάβητο με $\Sigma \subseteq V$,
- R είναι το σύνολο κανόνων, $R \subseteq V^* \times V^*$, και
- S είναι το αρχικό σύμβολο, $S \in V - \Sigma$.

Το Σ λέγεται σύνολο τερματικών συμβόλων, και το $V - \Sigma$ λέγεται σύνολο μη τερματικών συμβόλων (ή μεταβλητών).

Παράδειγμα γραμματικής:

- $V = \{0, 1, S, A\}$,
- $\Sigma = \{0, 1\}$,
- $R = \left\{ \begin{array}{rcl} S & \rightarrow & 0A1, \\ A1 & \rightarrow & 0A11, \\ 0A & \rightarrow & 00A1, \\ A & \rightarrow & \varepsilon \end{array} \right\}$

Διαισθητικά, το S «παράγει» όλες τις συμβολοσειρές της γλώσσας. Τα υπόλοιπα μη τερματικά σύμβολα «παράγουν» υποσυμβολοσειρές.

Γραμματικές χωρίς συμφραζόμενα (context-free)

Θα μελετήσουμε την ειδική κατηγορία των γραμματικών, τις context-free (χωρίς συμφραζόμενα) γραμματικές. Οι κανόνες τών γραμματικών χωρίς συμφραζόμενα είναι της μορφής $(V - \Sigma) \times V^*$ (ενώ των γενικών γραμματικών είναι της μορφής $V^* \times V^*$).

Παράδειγμα γραμματικής χωρίς συμφραζόμενα:

- $V = \{0, 1, S\}$,
- $\Sigma = \{0, 1\}$,
- $R = \left\{ \begin{array}{rcl} S & \rightarrow & 0S1, \\ S & \rightarrow & \varepsilon \end{array} \right\}$

Γλώσσες χωρίς συμφραζόμενα

Έστωσαν $u, v, w \in V^*$ για μια γραμματική G.

Αν $A \rightarrow w$ είναι κανόνας της G τότε λέμε ότι η uAv παράγει uwv και το συμβολίζουμε με

$$uAv \Rightarrow uwv.$$

Ορίζουμε επίσης τη σχέση u παράγει v σε μηδέν ή περισσότερα βήματα και το συμβολίζουμε $u \Rightarrow^* v$, αν υπάρχουν $u_1, \ldots, u_k$ όπου $k \geq 0$ τέτοια ώστε

$$u \Rightarrow u_1 \Rightarrow \cdots \Rightarrow u_k \Rightarrow v.$$

Ορίζουμε τη γλώσσα $L(G)$ μιας γραμματικής $G(V, \Sigma, R, S)$ σαν το σύνολο των συμβολοσειρών του Σ^* που παράγονται από το αρχικό σύμβολο S:

$$L(G) = \{w : w \in \Sigma^* \text{ και } S \Rightarrow^* w\}.$$

Οι γλώσσες χωρίς συμφραζόμενα είναι οι γλώσσες κάποιας γραμματικής χωρίς συμφραζόμενα.

Παράδειγμα

Η γλώσσα της γραμματικής

- $V = \{0, 1, S\}$,
- $\Sigma = \{0, 1\}$,
- $R = \left\{ \begin{array}{lcl} S & \to & 0S1, \\ S & \to & \varepsilon \end{array} \right\}$

είναι $L(G) = \{0^n 1^n : n \geq 0\}$.

Συμπέρασμα: Υπάρχουν γλώσσες χωρίς συμφραζόμενα που δεν είναι κανονικές.

Παράδειγμα

Η γλώσσα της γραμματικής

- $V = \{(,), S\}$,
- $\Sigma = \{(,)\}$,
- $R = \left\{ \begin{array}{lcl} S & \to & SS, \\ S & \to & (S), \\ S & \to & \varepsilon \end{array} \right\}$

είναι $L(G) = \{w : w \text{ αποτελείται από σωστά ζυγισμένες παρενθέσεις}\}$.

Παράδειγμα

Να βρείτε γραμματική χωρίς συμφραζόμενα για τη γλώσσα
$L = \{ww^R : w \in \{0, 1\}^*\}$. (Η γλώσσα των *παλινδρόμων άρτιου μήκους*).

$$\begin{array}{lcl} S & \longrightarrow & \varepsilon \\ S & \longrightarrow & 0S0 \\ S & \longrightarrow & 1S1 \end{array}$$

Για συντομία μερικές φορές γράφουμε

$$S \longrightarrow \varepsilon \mid 0S0 \mid 1S1$$

Κανονικές γλώσσες και γραμματικές

Όπως είδαμε υπάρχουν γλώσσες χωρίς συμφραζόμενα που δεν είναι κανονικές. Τώρα θα δείξουμε ότι το σύνολο των γλωσσών χωρίς συμφραζόμενα περιέχει τις κανονικές γλώσσες. Δηλαδή

$$\text{ΚΑΝΟΝ. ΓΛΩΣΣΕΣ} \subset \text{ΓΛΩΣΣΕΣ ΧΩΡΙΣ ΣΥΜΦΡ.}$$

Κανονική γραμματική: Μια γραμματική είναι κανονική αν οι κανόνες της έχουν τη μορφή $(V-\Sigma)\times\Sigma^*((V-\Sigma)\cup\varepsilon)$.

Παράδειγμα:

- $V=\{0,1,S\}$,
- $\Sigma=\{0,1\}$,
- $R=\left\{\begin{array}{lcl} S & \rightarrow & 01S, \\ S & \rightarrow & 010S, \\ S & \rightarrow & \varepsilon \end{array}\right\}$

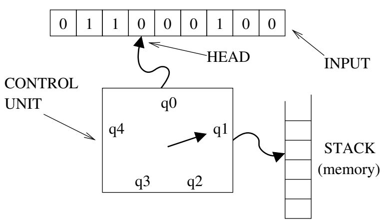
Η γλώσσα αυτής της γραμματικής είναι: $L(G)=(01\cup 010)^*$.

Θεώρημα 1 *Μια γλώσσα είναι κανονική αν και μόνο αν είναι γλώσσα μιας κανονικής γραμματικής.*

Η απόδειξη είναι εύκολη. Για κάθε (μη) ντετερμινιστικό πεπερασμένο αυτόματο μπορούμε να κατασκευάσουμε μια κανονική γραμματική όπου το σύνολο των μη τερματικών συμβόλων είναι οι καταστάσεις του αυτομάτου. Οι κανόνες της γραμματικής αντιστοιχούν στις μεταβάσεις του αυτομάτου. Π.χ., μια μετάβαση $\delta(q,0)=p$ αντιστοιχίζεται στον κανόνα $q\rightarrow 0p$. Για κάθε τελική κατάσταση f του αυτομάτου προσθέτουμε τον κανόνα $f\rightarrow\varepsilon$.

Αντιστρόφως, για κάθε κανονική γραμματική μπορούμε να κατασκευάσουμε ένα (μη ντετερμινιστικό) πεπερασμένο αυτόματο, όπου οι καταστάσεις είναι το σύνολο των μη τερματικών συμβόλων της γραμματικής. Αντιστοιχίζουμε τους κανόνες της γραμματικής σε μεταβάσεις, όπως κάναμε στο ευθύ του θεωρήματος.

Αυτόματα Στοίβας (Pushdown Automata)



Αυτόματο στοίβας $M(K,\Sigma,\Gamma,\Delta,s,F)$

$$\begin{array}{lcl} K & = & \text{σύνολο καταστάσεων (πεπερασμένο)} \\ \Sigma & = & \text{αλφάβητο εισόδου} \\ s & = & \text{αρχική κατάσταση } s\in K \\ F & = & \text{σύνολο τελικών καταστάσεων, } F\subseteq K \end{array}\left.\vphantom{\begin{array}{l}K\\ \Sigma\\ s\\ F\end{array}}\right\}\text{Όπως στα πεπερασμένα αυτόματα}$$

$$\begin{array}{lcl} \Gamma & = & \text{αλφάβητο στοίβας} \\ \Delta & = & \text{σχέση μετάβασης, } \Delta\subseteq(K\times(\Sigma\cup\{\varepsilon\})\times\Gamma^*)\times(K\times\Gamma^*) \end{array}$$

$((p,u,\beta),(q,\gamma))\in\Delta$: Το M είναι στην κατάσταση p, η κορυφή της στοίβας είναι β, διαβάζει από την είσοδο u και μεταβαίνει στην κατάσταση q αντικαθιστώντας το β με το γ.

Ένα αυτόματο στοίβας M δέχεται τη συμβολοσειρά w αν και μόνο αν, ξεκινώντας από την αρχική κατάσταση με άδεια στοίβα, επεξεργαστεί το w και καταλήξει σε κάποια τελική κατάσταση του F με άδεια στοίβα πάλι.

Δηλαδή

$$(s,w,\varepsilon)\vdash^*_M(q,\varepsilon,\varepsilon),\ \text{με } q\in F$$

Προσοχή! Οι παραπάνω τριάδες αντιστοιχούν σε συνολικές καταστάσεις (configurations) του αυτομάτου στοίβας. Σε μια συνολική κατάσταση η τρίτη συνιστώσα απεικονίζει τα περιεχόμενα ολόκληρης της στοίβας. Στα ορίσματα της Δ η τρίτη συνιστώσα απεικονίζει μια *συμβολοσειρά* που βρίσκεται στην κορυφή της στοίβας.

Η γλώσσα ενός αυτομάτου στοίβας είναι το σύνολο των συμβολοσειρών που δέχεται το αυτόματο.

$$L(M)=\{w : M \text{ δέχεται } w\}$$

Παράδειγμα

Έστω το αυτόματο στοίβας με σχέση μετάβασης

$$\Delta = \{ \begin{array}{l} ((s,0,\varepsilon),(s,0)) \\ ((s,1,\varepsilon),(s,1)) \\ ((s,\#,\varepsilon),(f,\varepsilon)) \\ ((f,0,0),(f,\varepsilon)) \\ ((f,1,1),(f,\varepsilon))\} \end{array} \quad \text{και } F = \{f\}$$

Το αυτόματο δέχεται τις συμβολοσειρές που είναι της μορφής $w\#w^R$. Η γλώσσα δηλαδή του αυτομάτου είναι η $L = \{w\#w^R : w \in \{0,1\}^*\}$.

Παράδειγμα

$$L = \{w : w \text{ έχει ζυγισμένες αγκύλες } \}$$

$$S \longrightarrow \varepsilon \,|\, [S] \,|\, SS$$

$$(s, [, \varepsilon)(s, [)$$
$$(s,], [)(s, \varepsilon)$$

Κατάσταση	**Εισοδος**	**Στοίβα**
s	[[ ]] [ ]	ε
s	[ ]] [ ]	[
s	]] [ ]	[[
s	] [ ]	[
s	[ ]	ε
s	]	[
s	ε	ε

Γραμματικές χωρίς συμφραζόμενα και αυτόματα στοίβας

Θεώρημα 2 *Οι γλώσσες των αυτομάτων στοίβας είναι ακριβώς οι γλώσσες χωρίς συμφραζόμενα.*

Το θεώρημα προκύπτει από τα Θεωρήματα 3 και 4:

Θεώρημα 3 *Για κάθε γραμματική χωρίς συμφραζόμενα G υπάρχει αυτόματο στοίβας M, τέτοιο ώστε $L(M) = L(G)$.*

Θεώρημα 4 *Για κάθε αυτόματο στοίβας M υπάρχει γραμματική G χωρίς συμφραζόμενα, τέτοια ώστε $L(G) = L(M)$.*

Οι αποδείξεις παραλείπονται για την ώρα.

Παράρτημα: απόδειξη Θεωρ. 3

Απόδειξη: Έστω $G(V, \Sigma, R, S)$ μια γραμματική χωρίς συμφραζόμενα. Κατασκευάζουμε το εξής αυτόματο στοίβας M που δέχεται τη γλώσσα που παράγει η G. Το M έχει μόνο δύο καταστάσεις, την αρχική κατάσταση s και την τελική κατάσταση f. Το αλφάβητο της στοίβας Γ είναι το V. Τέλος η σχέση μετάβασης του M είναι:

1. $((s,\varepsilon,\varepsilon),(f,S))$
2. $((f,\varepsilon,A),(f,v))$ για κάθε κανόνα της G: $A \to v$
3. $((f,\sigma,\sigma),(f,\varepsilon))$ για κάθε $\sigma \in \Sigma$.

Το M δηλαδή δέχεται μια συμβολοσειρά w ως εξής:

- Στην αρχή η στοίβα έχει το σύμβολο S (βήμα τύπου 1).
- Κάθε φορά που η κορυφή της στοίβας είναι **μη τερματικό σύμβολο** το αντικαθιστά μιμούμενο τον τρόπο που η G παράγει τη w (βήμα τύπου 2).
- Κάθε φορά που η κορυφή της στοίβας είναι **τερματικό σύμβολο** πρέπει να το 'ταιριάξει' με ίδιο σύμβολο του **input** (βήμα τύπου 3).

Παράδειγμα: Για τη γραμματική G με κανόνες $S \to \varepsilon \,|\, [S] \,|\, SS$ το αυτόματο στοίβας που κατασκευάζεται έχει τις μεταβάσεις:

1. $((s,\varepsilon,\varepsilon),(f,S))$
2. $((f,\varepsilon,S),(f,\varepsilon))$
   $((f,\varepsilon,S),(f,[S]))$
   $((f,\varepsilon,S),(f,SS))$
3. $((f,[,[),(f,\varepsilon))$
   $((f,],]),(f,\varepsilon))$