

Γραμματικές

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

Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών

Εθνικό Μετσόβιο Πολυτεχνείο



Αναπαράσταση Γλωσσών

- Απαρίθμηση στοιχείων (πεπερασμένες).
- Χαρακτηριστική ιδιότητα.
 - $L_1 = \{w \in \{0, 1\}^* : w \text{ έχει ίδιο αριθμό } 0 \text{ και } 1\}$, $L_2 = \{0^n 1^n : n \in \mathbb{N}\}$
- Ως **αποτέλεσμα πράξεων** (π.χ. ένωση, τομή, παράθεση, *)
(π.χ. **κανονικές εκφράσεις**).
- **Υπολογιστική μηχανή** (αλγόριθμο) που **αποφασίζει** πότε είσοδος ανήκει στη γλώσσα και πότε όχι.
 - DFA, NFA, αυτόματα στοίβας, LBA, μηχανές Turing.
 - «Πολυπλοκότητα» γλώσσας καθορίζεται από **υπολογιστικές δυνατότητες «απλούστερης» μηχανής** που την αποφασίζει.
- **Γραμματική**: μηχανισμός **παραγωγής** των στοιχείων της γλώσσας.
 - «Πολυπλοκότητα» γλώσσας καθορίζεται από **δομή «απλούστερης» γραμματικής**.

Γραμματικές

- **Γραμματική**: μηχανισμός «παραγωγής» γλώσσας.
 - Ξεκινά από **αρχικό σύμβολο**.
 - Προχωρά **εφαρμόζοντας παραγωγές** (ή κανόνες) όσο αυτό είναι δυνατόν.
 - **Επιλογή** κανόνων: όχι πλήρως καθορισμένη λειτουργία. (ίδια γραμματική **παράγει πολλές** συμβ/ρές).
 - **Δύσκολο** αυστηρή γραμματική για «**φυσική**» γλώσσα.
 - Σχετικά απλές αυστηρές γραμματικές για **γλώσσες προγραμματισμού**.

Γραμματική Δομής-Φράσης

- ... είναι μία τετράδα $G(V, T, S, P)$ όπου:
 - V : **αλφάβητο**. $T \subseteq V$: σύνολο **τερματικών** συμβόλων.
 - $N = V - T$: **μη τερματικά** σύμβολα, μόνο σε παραγωγές.
 - Συμβολοσειρές γλώσσας: **μόνο τερματικά** σύμβολα.
 - $S \in V - T$: **αρχικό** (μη-τερματικό) σύμβολο.
 - $P \subseteq V^*(V - T)V^* \times V^*$: σύνολο **παραγωγών** (ή κανόνων) για τη μεταγραφή (μη-τερματικών συμβόλων). Παραγωγές P
 - Γράφουμε $a \rightarrow \beta$ αντί $(a, \beta) \in P$.

$$V = \{0, 1, 2, S, A, B\},$$

$$T = \{0, 1, 2\},$$

$$S$$

$$S \rightarrow \varepsilon$$

$$S \rightarrow 0SAB$$

$$BA \rightarrow AB$$

$$0A \rightarrow 01$$

$$1A \rightarrow 11$$

$$1B \rightarrow 12$$

$$2B \rightarrow 22$$

Παραδείγματα

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

Παραγωγές P_1
$S \rightarrow 0B \mid 1A \mid \varepsilon$
$A \rightarrow 0 \mid 0S \mid 1AA$
$B \rightarrow 1 \mid 1S \mid 0BB$

Παραγωγές P_2
$S \rightarrow 1S \mid 1 \mid 0A$
$A \rightarrow 1A \mid 0B$
$B \rightarrow 1B \mid 0S \mid 0$

$$V' = \{0, 1, S\},$$
$$T = \{0, 1\},$$
$$S$$

Παραγωγές P_3
$S \rightarrow 0S0 \mid 1S1 \mid 1 \mid 0 \mid \varepsilon$

Παραγωγές P_4
$S \rightarrow 0S1 \mid \varepsilon$

Σχέσεις Παραγωγής

- Ορίζουμε σχέση **u παράγει άμεσα v** στο V^* : $u \Rightarrow v$.
 - Έστω $\pi, \kappa, z_0, z_1 \in V^*$ τέτοια ώστε $u = \pi z_0 \kappa$ και $v = \pi z_1 \kappa$.
 - **u παράγει άμεσα v** αν υπάρχει παραγωγή $z_0 \rightarrow z_1$
- Ορίζουμε σχέση **u παράγει v** στο V^* : $u \Rightarrow^* v$.
 - **u παράγει v** αν υπάρχουν $k \geq 1$ και $u_1, \dots, u_{k-1} \in V^*$ τέτοια ώστε
$$u \Rightarrow u_1 \Rightarrow \dots \Rightarrow u_{k-1} \Rightarrow v \quad (\text{παραγωγή } k \text{ βημάτων})$$
- **Γλώσσα γραμματικής G** : σύνολο συμβολοσειρών T^* που παράγονται από αρχικό σύμβολο S (σε πεπερασμένο #βημάτων).
$$L(G) = \{w \in T^* : S \Rightarrow^* w\}$$
- Αντιστοιχία με μη ντετερμινιστικό υπολογισμό:
 - $w \in L(G)$ ανν υπάρχει παραγωγή από S .

Τύποι Γραμματικών

- **Ιεραρχία** γραμματικών ανάλογα με δομή κανόνων παραγωγής (Chomsky, 50's).
- Τύπος 0: Γραμματικές **χωρίς περιορισμούς**.
 - Γλώσσες αναγνωρίσιμες (αναδρομικά απαριθμήσιμες) από μηχανές Turing.
- Τύπος 1: Γραμματικές **με συμφραζόμενα (context-sensitive)**.
 - \forall παραγωγή $a \rightarrow \beta$, $|\beta| \geq |a|$ (non-contracting).
 - $S \rightarrow \varepsilon$ μόνο αν S δεν εμφανίζεται σε δεξιό μέλος παραγωγής.
 - Γλώσσες που αποφασίζονται από LBA.

Τύποι Γραμματικών

- Τύπος 2: Γραμματικές χωρίς συμφραζόμενα (context-free).
 - Παραγωγές $P \subseteq (V - T) \times V^*$
 - Παραγωγές μορφής $A \rightarrow w$, όπου $w \in V^*$
 - Μη τερματικά σύμβολα αντικαθίσταται ανεξάρτητα από συμφραζόμενα (context-free).
 - Γλώσσες που αποφασίζονται από (μη ντετερμινιστικά) αυτόματα στοίβας.
 - Συντακτικό γλωσσών προγραμματισμού (Pascal, C, C++, ...) περιγράφεται από γραμματικές χωρίς συμφραζόμενα.

Τύποι Γραμματικών

- Τύπος 3: Κανονικές γραμματικές (regular).
 - Παραγωγές $P \subseteq (V - T) \times T^*((V - T) \cup \epsilon)$
 - Παραγωγές μορφής $A \rightarrow w \mid wB$, όπου $w \in T^*$, $A, B \in V - T$.
 - Γλώσσες που αποφασίζονται από πεπερασμένα αυτόματα.
 - Γλώσσες που αναπαρίστανται με κανονικές εκφράσεις.
 - Λεκτική ανάλυση βασίζεται σε κανονικές γραμματικές.
- Γλώσσα τύπου k αν παράγεται από γραμματική τύπου k και όχι γραμματική τύπου $k + 1$.
 - Οι 4 κλάσεις γλωσσών είναι μη κενές.
 - Τύπου 3 \subset Τύπου 2 \subset Τύπου 1 \subset Τύπου 0.