ΕΠΛ 211:
Θεωρία Υπολογισμού και Πολυπλοκότητας

Διάλεξη 7: Ασυμφραστικές Γλώσσες (Γλώσσες Ελεύθερες Συμφραζομένων)

Τι θα κάνουμε σήμερα…

- Εισαγωγικά
- Ασυμφραστικές Γραμματικές (2.1)
- Τυπικός Ορισμός Της Ασυμφραστικής Γραμματικής (2.1.1)
- Σχεδίαση Ασυμφραστικών Γραμματικών (2.1.3)
- Πολυτροπία (2.1.4)
- Κανονική Μορφή Chomsky (2.1.5)

Ασυμφραστικές Γλώσσες

- Παρατηρήσαμε ότι η γλώσσα $B=\{0^{\kappa}1^{\kappa} \mid \kappa \geq 0\}$
  - Δεν αναγνωρίζεται από κανένα πεπερασμένο αυτόματο
  - Δεν περιγράφεται από καμιά κανονική έκφραση
- Υπάρχει δυνατότερο υπολογιστικό μοντέλο που να μπορεί να αναγνωρίζει την B;
  - **Ασυμφραστικές Γραμματικές**
    - **Περιγραφή ιδιοτήτων αναδρομικού χαρακτήρα**
  - **Αυτόματα με στοίβα**

Ασυμφραστικές Γλώσσες

Ορισμός:
Οι γλώσσες που παράγονται από ασυμφραστικές γραμματικές ονομάζονται ασυμφραστικές γλώσσες (context free languages).

Ασυμφραστικές Γραμματικές (CFG)

$$
\begin{array}{rcl}
A & \rightarrow & 0A1 \\
A & \rightarrow & B \\
B & \rightarrow & \#
\end{array}
$$

- Συλλογή **Κανόνων Αντικατάστασης**
  - Κάθε κανόνας σε ξεχωριστή γραμμή
- Μορφή κανόνα αντικατάστασης:
  - **Μη-Τερματικό Σύμβολο** (μεταβλητή) ⟶ **Λέξη**
  - Λέξη αποτελείται από τερματικά και μη-τερματικά σύμβολα
    - π.χ. 0A1 – 0,1 τερματικά σύμβολα και A μη-τερματικό
- **Εναρκτήρια Μεταβλητή**
  - Αριστερό μέλος πρώτου κανόνα

Παραγωγή Λέξεων

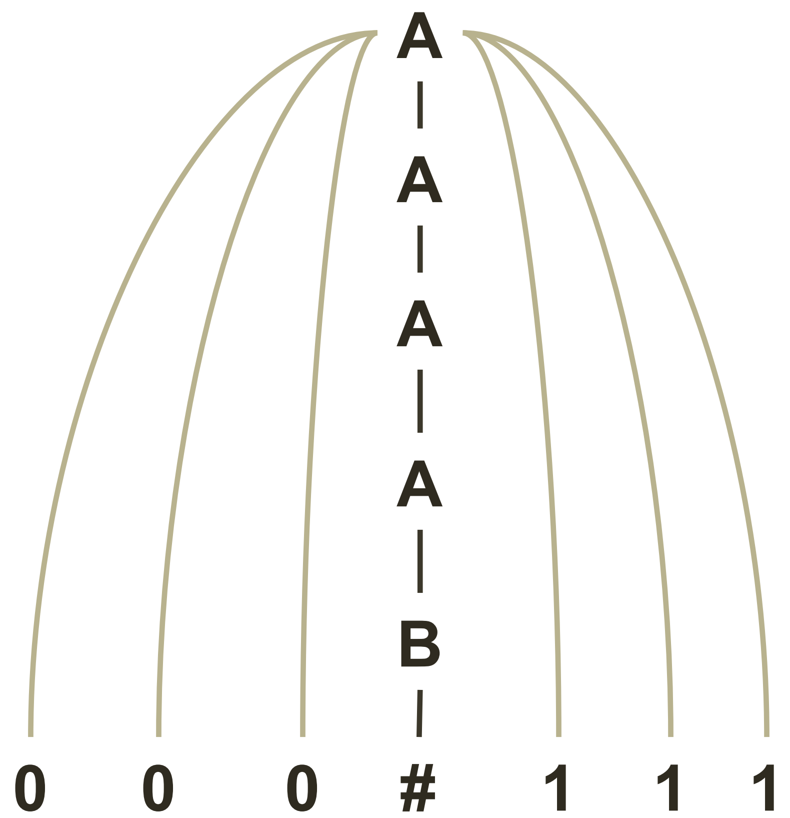
- Μια γραμματική περιγράφει μια γλώσσα αν μας επιτρέπει να παραγάγουμε όλες τις λέξεις της γλώσσας

- Βήματα παραγωγής λέξης:
  1. Γράφουμε την εναρκτήρια μεταβλητή
  2. Επιλέγουμε ένα μη-τερματικό σύμβολο που υπάρχει στη λέξη που έχουμε γραμμένη και το αντικαθιστούμε με το δεξί μέλος ενός κανόνα που ξεκινά από αυτό.
  3. Επαναλαμβάνουμε το 2 μέχρις ότου η λέξη να περιέχει μόνο τερματικά σύμβολα.

Παράδειγμα

- Παραγωγή λέξης 000#111 από G1
  - A
  - 0A1
  - 00A11
  - 000A111
  - 000B111
  - 000#111

Συντακτικό Δέντρο



$L(G1)=\{0^n\#1^n \mid n\geq 0\}$

Συγχώνευση Κανόνων

- Μπορούμε να συγχωνεύσουμε όλους τους κανόνες που έχουν στο αριστερό τους μέλος την ίδια μεταβλητή
  - Π.χ. A -> 0A1 και A-> B τότε γράφουμε A-> 0A1 | B

Τυπικός Ορισμός CFG

Ορισμός

***Ασυμφραστική γραμματική* είναι μια τετράδα (V, Σ, R, S)**

1. ***V* = πεπερασμένο σύνολο *μη-τερματικών* συμβόλων**
2. ***Σ* = πεπερασμένο σύνολο *τερματικών* συμβόλων, ξένο προς το V,**
3. ***R* = Το σύνολο των *κανόνων*, καθένας από τους οποίους αποτελείται από *μια μεταβλητή* και *μια λέξη* από μεταβλητές και τερματικά σύμβολα**
4. **$S \in V$ = εναρκτήρια μεταβλητή**

Τυπικός Ορισμός CFG

- Έστω u, v, w λέξεις από μεταβλητές και τερματικά σύμβολα, και A->w κανόνας της γραμματικής.

- Τότε uAv **αποδίδει** τη λέξη uwv, $uAv \Rightarrow uwv$

- Η u **παράγει** τη λέξη v, $u \Rightarrow^* v$ εάν
  - $u = v$
  - Υπάρχει ακολουθία $u_1 \ldots u_k, k \geq 0$ *τέτοια ώστε*

$$u_1 \Rightarrow u_2 \Rightarrow \ldots \Rightarrow u_k \Rightarrow v$$

- **Γλώσσα της γραμματικής** G είναι το σύνολο

$$L(G) = \{w \in \Sigma^* | S \Rightarrow^* w\}$$

Παραδείγματα

- G_3=({S}, {a,b}, R, S) όπου R το σύνολο των κανόνων¨
  - S-> aSb | SS | e
  - abab, aababb $\in$ L(G_3)

- G_4=({<ΕΚΦΡΑΣΗ>,<ΟΡΟΣ>,<ΠΑΡΑΓΟΝΤΑΣ>}, {a,+, x,(,)}, R,<ΕΚΦΡΑΣΗ>) και R οι κανόνες:
  - <ΕΚΦΡΑΣΗ> -> <ΕΚΦΡΑΣΗ> + <ΟΡΟΣ> | <ΟΡΟΣ>
<ΟΡΟΣ> -> <ΟΡΟΣ> x <ΠΑΡΑΓΟΝΤΑΣ> | <ΠΑΡΑΓΟΝΤΑΣ>
<ΠΑΡΑΓΟΝΤΑΣ> -> (<ΕΚΦΡΑΣΗ>) | a

  - Πως σχηματίζονται τα a+a x a, (a+a) x a?

Σχεδίαση CFG

- Κατασκευή γραμματικής για σύνθετη γλώσσα
  - Διαιρούμε την γλώσσα σε απλούστερα μέρη
  - Κατασκευάζουμε μια γραμματική για κάθε μέρος
  - Προσθέτουμε τον κανόνα $S \rightarrow S_1|S_2|\ldots|S_k$
    - S_i οι εναρκτήριες μεταβλητές των επιμέρους γραμματικών
  - Π.χ. $L = \{0^n1^n|n \geq 0\} \cup \{1^n0^n|n \geq 0\}$

$$S_1 \rightarrow 0S_11|e \qquad S_2 \rightarrow 1S_20|e$$

$$S \rightarrow S_1|S_2$$

- Γραμματική για κανονική γλώσσα που ξέρουμε το αυτόματο που την αναγνωρίζει
  - Για κάθε μετάβαση δ(q_i, α)=q_j κανόνας $R_i \rightarrow aR_j$
  - Για κάθε κατάσταση αποδοχής q_i κανόνας $R_i \rightarrow e$
  - Εναρκτήρια κατάσταση η R_0 (q_0 εναρκτήρια αυτομάτου)

Σχεδίαση CFG

- Αλληλένδετες υπολέξεις
  - Εμφάνιση μιας υπολέξης με βάση πόσες φορές εμφανίζεται μια δεύτερη υπολέξη
    π.χ. $\{0^{\kappa}1^{\kappa} \mid \kappa \geq 0\}$
  - Χρησιμοποιούμε κανόνα τησ μορφής $R \rightarrow uRv$
    - Όπου u η πρώτη υπολέξη
    - Και v η υπολέξη που εξαρτάται στο u

- Αναδρομικές δομές
  - Εμφανίζονται ως τμήματα άλλων δομών η ακόμη και του εαυτού τους
    π.χ.
    <ΕΚΦΡΑΣΗ> -> <ΕΚΦΡΑΣΗ> + <ΟΡΟΣ> | <ΟΡΟΣ>
    <ΠΑΡΑΓΟΝΤΑΣ> -> (<ΕΚΦΡΑΣΗ>) | a

Πολυτροπία

- Κάποια γραμματική παράγει την ίδια λέξη με περισσότερους από ένα τρόπους
  - Περισσότερα από ένα συντακτικά δέντρα
  - π.χ. $E \rightarrow E + E | E \times E | (E) | a$
    Πως παράγεται η α + α x α;
- **Εξ αριστερών παραγωγή**
  - Σε κάθε βήμα αντικαθιστούμε την αριστερότερη από τις εναπομείναντες μεταβλητές

Ορισμός

Για κάθε ασυμφραστική γραμματική G και για κάθε λέξη w, λέμε ότι η w παράγεται πολύτροπα στην G εάν υπάρχουν για αυτήν περισσότερες από μια εξ αριστερών παραγωγές στην G. Η G ονομάζεται πολύτροπη αν παράγει κάποια λέξη πολύτροπα.

Κανονική μορφή Chomsky

Ορισμός

Μια ασυμφραστική γραμματική G βρίσκεται σε κανονική μορφή Chomsky εάν κάθε κανόνας της βρίσκεται σε μια από τις μορφές

$$\begin{array}{ccc} A & \rightarrow & BC \\ A & \rightarrow & a \end{array}$$

όπου α οποιοδήποτε τερματικό σύμβολο και B,C οποιοδήποτε μεταβλητή πλην της εναρκτήριας. Επιτρέπεται και ο κανόνας $S \rightarrow e$.

Ερωτήσεις;