Γλώσσες όχι χωρίς Συμφραζόμενα

Δημήτρης Φωτάκης

Τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων

Πανεπιστήμιο Αιγαίου

Περιοδικότητα ΓχΣ

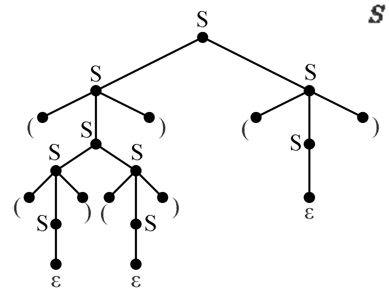
- (Άπειρη) κανονική γλώσσα: μεγάλη συμβ/ρά οδηγεί ΠΑ σε ίδια κατάσταση («κύκλος»).
  - Επανάληψη τμήματος συμβ/ράς που αντιστοιχεί σε κύκλο οδηγεί σε τελική κατάσταση (συμβ/ρά της γλώσσας).
- (Άπειρη) ΓχΣ L παράγεται από γραμματική G(V, Σ, R, S).
  - Έστω συμβ/ρά $w = uvxyz$: κατά την παραγωγή της εφαρμόζονται δύο παραγωγές για μη-τερματικό Α:
    $S \Rightarrow^* uAz \Rightarrow^* uvAyz \Rightarrow^* uvxyz$ με $A \Rightarrow^* vAy$ και $A \Rightarrow^* x$
  - Γλώσσα χωρίς συμφραζόμενα: για κάθε $n \geq 0$, εφαρμογή 1ης n φορές και μετά 2ης δίνει $uv^n xy^n z \in L$

$$S \Rightarrow^* uAz \Rightarrow^* uvAyz \Rightarrow^* uv^2Ay^2z \Rightarrow^* \cdots \Rightarrow^* uv^n xy^n z$$

Συντακτικά Δέντρα (parse trees)

- … απεικονίζει την παραγωγή συμβ/ράς από γραμματική G(V, Σ, R, S):
  - Ρίζα επιγράφεται με αρχικό σύμβολο S.
  - Κόμβοι επιγράφονται με σύμβολα του V.
    - Ενδιάμεσοι κόμβοι επιγράφονται με **μη-τερματικά**.
    - Φύλλα επιγράφονται με **τερματικά** ή ε.
  - (Ενδιάμεσος) κόμβος με επιγραφή Α και παράθεση επιγραφών παιδιών του $u \in V^*$: εφαρμογή κανόνα $A \to u$.
  - Παράθεση επιγραφών φύλλων από αριστερά προς δεξιά δίνει τη συμβ/ρά που παράγεται από συντακτικό δέντρο.

Παράδειγμα



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

(()())()

Μεγάλες Συμβολοσειρές

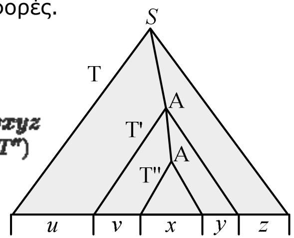
- Γραμματική χωρίς συμφραζόμενα G(V, Σ, R, S):
  - **Εύρος** $\varphi(G)$: μέγιστος #συμβόλων σε δεξιό μέλος κανόνα.
    - Κάθε κόμβος συντακτικού δέντρου έχει $\leq \varphi(G)$ παιδιά.
  - Παραγόμενη συμβ/ρά από συντακτικό δέντρο ύψους h έχει μήκος $\leq \varphi(G)^h$
  - Κάθε συμβ/ρά με μήκος $> \varphi(G)^{|V - \Sigma|}$ παράγεται από συντακτικό δέντρο ύψους $\geq |V - \Sigma| + 1$.
    - Υπάρχει κλάδος με $\geq |V - \Sigma| + 2$ σύμβολα, 1 τερματικό.
    - Υπάρχει κλάδος όπου κάποιο μη-τερματικό σύμβολο εμφανίζεται τουλάχιστον δύο φορές.

Μεγάλες Συμβολοσειρές

- Γραμματική χωρίς συμφραζόμενα G(V, Σ, R, S). Κάθε συμβ/ρά w, $|w| > \varphi(G)^{|V - \Sigma|}$ παράγεται από συντακτικό δέντρο με κλάδο όπου κάποιο μη-τερματικό εμφανίζεται τουλάχιστον δύο φορές.
  - Έστω $w = uvxyz$
  - Παραγωγή αναλύεται:

$$A \Rightarrow^* vAy \ (T')$$
$$S \Rightarrow^* uAz \Rightarrow^* uvAyz \Rightarrow^* uvxyz$$
$$S \Rightarrow^* uAz \qquad A \Rightarrow^* x \ (T'')$$

  - Για κάθε #εφαρμογών 1ης παραγωγής παίρνουμε συμβ/ρά στη γλώσσα L(G).

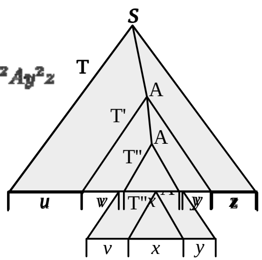


«Φούσκωμα» Συμβολοσειράς

- Κάθε συμβ/ρά w, $|w| > \varphi(G)^{|V - \Sigma|}$, γράφεται $w = uvxyz$ ώστε $uv^n xy^n z \in L(G)$ για κάθε $n \geq 0$.

$$S \Rightarrow^* uAz \Rightarrow^* uvAyz \Rightarrow^* uvxyz = w$$

  - $n = 0$: $S \Rightarrow^* uAz \Rightarrow^* uxz$
  - $n = 1$: εξ' ορισμού.
  - $n = 2$: $S \Rightarrow^* uAz \Rightarrow^* uvAyz \Rightarrow^* uv^2Ay^2z$
  - … κοκ.



Θεώρημα Άντλησης (Pumping Lemma)

- Για κάθε γλώσσα L χωρίς συμφραζόμενα, υπάρχει $k \geq 1$ ώστε κάθε $w \in L$, $|w| > k$, γράφεται $w = uvxyz$:
  1. για κάθε $n \geq 0$, $uv^n xy^n z \in L$.
  2. $|vy| > 0$ (τουλ. ένα από τα v, y μη-κενό).
  3. $|uxy| \leq k$
- Απόδειξη:
  - Αποδείξαμε το (1).
  - Για (2), συντακτικό δέντρο με ελάχιστο αριθμό κόμβων.
    - Κάθε κανόνας συνεισφέρει στην τελική συμβ/ρά.
  - Για (3), θεωρούμε δύο κατώτερες εμφανίσεις A στον αντίστοιχο κλάδο.

Εφαρμογές

- Γλώσσα $L = \{a^n b^n c^n : n \geq 0\}$ δεν είναι χωρίς συμφραζόμενα.
  - Έστω *k* ο φυσικός αριθμός του λήμματος άντλησης.
  - Θεωρούμε w = $a^k b^k c^k$ = uvxyz .
  - Αν vy περιέχει τρία διαφορετικά σύμβολα, στη uv^nxy^nz εμφανίζονται σύμβολα εκτός σειράς.
  - Αν vy περιέχει λιγότερα από τρία διαφορετικά σύμβολα, στη uv^nxy^nz υπάρχει διαφορετικός αριθμός a, b, c.
  - Διαισθητικά, χρειάζονται δύο στοίβες για την αναγνώριση αυτής της γλώσσας.

Εφαρμογές

- Γλώσσα $L = \{a^n : n \text{ είναι πρώτος}\}$ δεν είναι χωρίς συμφραζόμενα.
  - Θεωρούμε w = uvxyz ώστε |w| πρώτος.
  - Έστω |vy| = t και |uxz| = r.
  - Ο αριθμός r+nt **δεν** είναι **πρώτος** για κάθε n ≥ 0. (π.χ. n = r + t + 1 ⇒ (r+t)(t+1)).
  - Για αλφάβητα ενός συμβόλου, γλώσσες χωρίς συμφραζόμενα είναι κανονικές.

Εφαρμογές

- Γλώσσα $L = \{w \in \{a, b, c\} : w \text{ έχει ίσο αριθμό } a, b \text{ και } c\}$ δεν είναι χωρίς συμφραζόμενα.
  - Η γλώσσα $L_1 = \{a^n b^n c^n : n \geq 0\}$ είναι τομή της L με κανονική γλώσσα a* b* c*.
  - Αν L ήταν ΓχΣ, θα ήταν και η L_1 . Άτοπο.

Μη-Κλειστότητα

- Η κλάση των γλωσσών χωρίς συμφραζόμενα **δεν** είναι κλειστή ως προς συμπλήρωμα και τομή.
  - $L_1 = \{a^n b^n c^m : n \geq 0, m \geq 0\}$ χωρίς συμφραζόμενα.
  - $L_2 = \{a^m b^n c^n : m \geq 0, n \geq 0\}$ χωρίς συμφραζόμενα.
  - Τομή $L_1 \cap L_2 = \{a^n b^n c^n : n \geq 0\}$ **δεν** είναι χωρίς συμφραζόμενα.
  - Μη-κλειστότητα ως προς συμπλήρωμα: $L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}}$