Parikh’s Theorem

Ορισμοί

- Γραμματική G=(V,T,P,S)
- Γλώσσες χωρίς συμφραζόμενα (κανόνες της μορφής: $\mathrm{A} \longrightarrow \alpha, \mathrm{A} \in \mathrm{V}$)
- Γραμμικό σύνολο:
  Ένα υποσύνολο Q του $\mathbb{N}^n$ λέγεται γραμμικό όταν υπάρχουν στοιχεία $\mathrm{\alpha}, \beta_1, \beta_2, \dots, \beta_m$ του $\mathbb{N}^n$ τέτοια ώστε:
  $$Q = \{x | x = \mathrm{\alpha} + n_1\beta_1 + n_2\beta_2 + \cdots + n_m\beta_m, \text{όπου } n_i \epsilon \mathbb{N}\}$$
- Ημιγραμμικό σύνολο:
  Ένα υποσύνολο J του $\mathbb{N}^n$ λέγεται ημιγραμμικό αν είναι πεπερασμένη ένωση γραμμικών συνόλων(του $\mathbb{N}^n$).

Ορισμοί

Ορίζουμε το mapping Φ της γλώσσας L με $\mathrm{T_L} = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$ ως εξής:

Φ(ε)=(0,0,0,…,0,0)

Φ(α_i)=(0,0,..,0,1,0,…,0,0)

Φ(xy)=Φ(x)+Φ(y)

Θεώρημα(Parikh 61&66)

Το mapping μιας C.F. γλώσσας L είναι ημιγραμμικό σύνολο.

Παραδείγματα

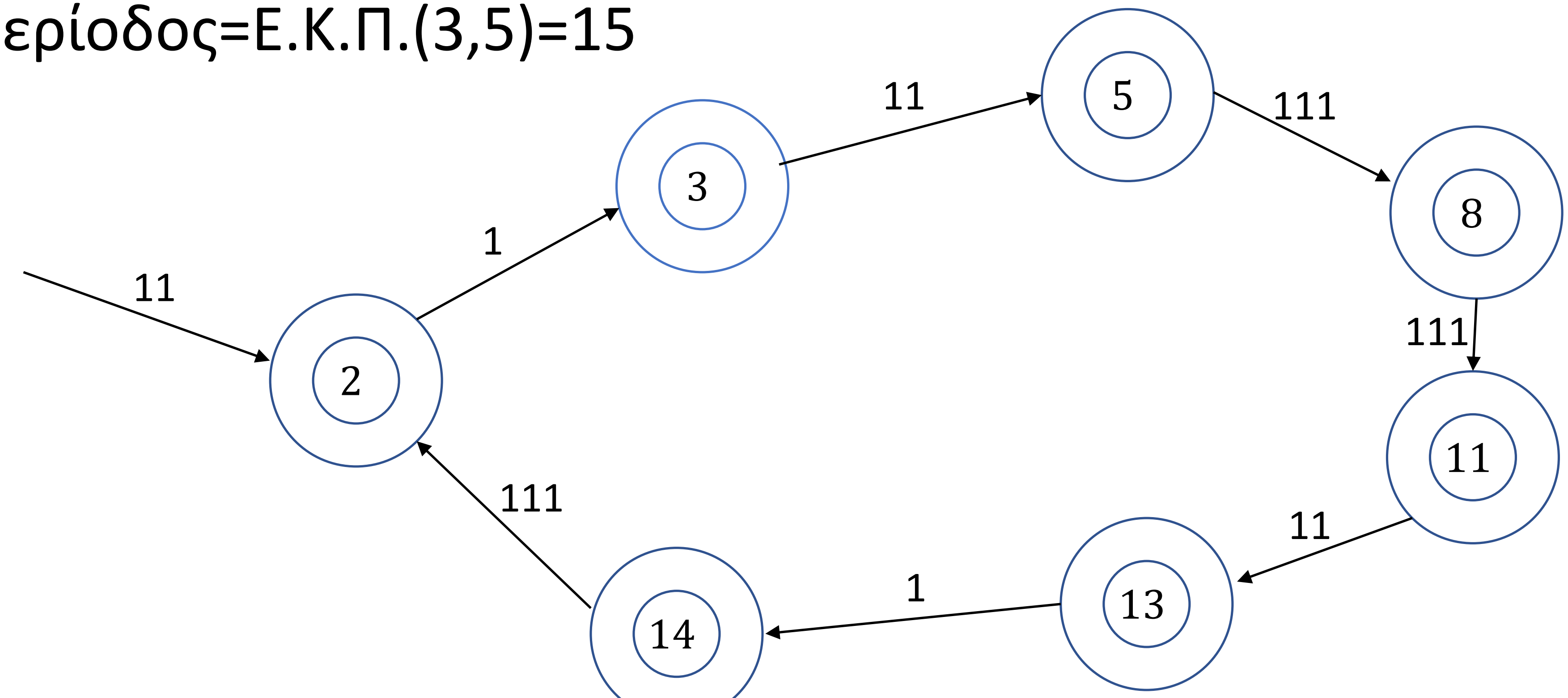
- L=$\{a^n b^n\}$
- L=$\{a^m b^n$| m>n≥0}
- L=$\{w \in \{a, b\}^* | \#a \neq \#b\}$

Πορίσματα

- Κάθε CF γλώσσα έχει ισοδύναμη κανονική γλώσσα ως προς την αντιστοιχία μηκών των strings τους.
- Περιοδικότητα των μηκών μιας CF γλώσσας.
- Μια CF γλώσσα με 1 τερματικό σύμβολο είναι κανονική.
- Ικανή, αλλά όχι αναγκαία συνθήκη.

Παράδειγμα

- Έστω T={1} και το σύνολο AUB με A={2+3k}, B={5l-2, l>0}
- A={2,5,8,11,14,17,20,...}
- B={3,8,13,18,23...}
- AUB={2,3,5,8,11,13,14,17,18,20,23...}
- Περίοδος=Ε.Κ.Π.(3,5)=15

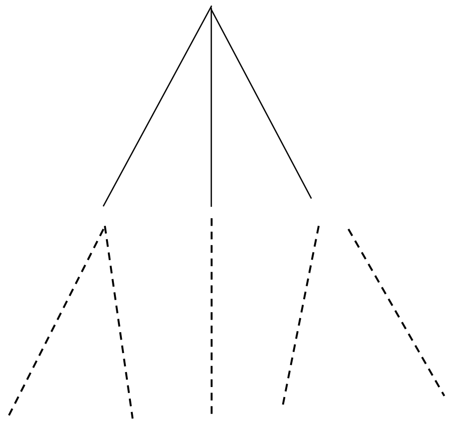


Απόδειξη

Έστω η γλώσσα L(G).

- Επιλέγουμε κάποια $A\epsilon\mathrm{V}$ και μαζί με τους κανόνες παραγωγής τους σχηματίζουμε τη γλώσσα L'(G).
- Η L μπορεί να γραφεί ως πεπερασμένη ένωση τέτοιων γλωσσών:
$$L(G)=L_1(G) \cup \cdots \cup L_k(G)$$
- Ορίζουμε τα σύνολα:
    1. $\mathrm{R_A} = \{\text{strings w τα οποία περιέχουν μόνο τερματικά σύμβολα και το σύμβολο A και η ρίζα τους είναι το A}\}$
    2. $\mathrm{T_A} = \{$όλα τα τερματικά τα οποία χρησιμοποιούν όλα τα μη

Πώς μπορεί να παραχθεί ένα string ;



w=$a_1 \dots \dots a_2 .. a_1 .. \ a_k \dots \dots a_m$

- Εξαιτίας του P.L. μπορούμε σε κάθε μη-τερματικό κόμβο B του δέντρου(που εμφανίζεται δύο φορές) να αντικαταστήσουμε ένα δέντρο με ένα άλλο δέντρο με ρίζα το B, το οποίο να είναι είτε κενό είτε μεγαλύτερο.
- Έτσι αν το αρχικό string ήταν της μορφής *z=uvwxy* και στη συνέχεια μετατραπεί σε $z' = uv^i wx^i y$, για το mapping του θα ισχύει ότι:

  $$Φ(z')=Φ(u)+i \cdot Φ(v)+Φ(w)+i \cdot Φ(x)+Φ(y)$$

  και άρα είναι ένας γραμμικός συνδυασμός των ήδη παραγόμενων mappings.

Κάνοντας αυτό για κάθε υποσύνολο της γλώσσας, η ένωση:

$$L(G)=L_1(G) \cup \cdots \cup L_k(G)$$

θα είναι ημιγραμμικό σύνολο.