# Parikh’s Theorem

Παρουσίαση: Γιώργος Αυδής

# 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($\varepsilon$)=(0,0,0,…,0,0)

Φ($\alpha_i$)=(0,0,..,0,1,0,…,0,0)

Φ(xy)=Φ(x)+Φ(y)

Ουσιαστικά το Φ μετράει το πλήθος των εμφανίσεων ενός τερματικού συμβόλου σε μια λέξη.

Π.χ. Φ(abbba)=(2,3), Φ(abbba)=(2,3,0), Φ(babab)=(2,3,0), όπου τα δύο τελευταία είναι σε γλώσσα με τερματικά τα a, b, c.

# Θεώρημα(Parikh 61&66)

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

Το mapping μιας γλώσσας είναι το σύνολο των mappings των λέξεών της:
Φ(L)={Φ(w), $w \in L$}

# Παραδείγματα

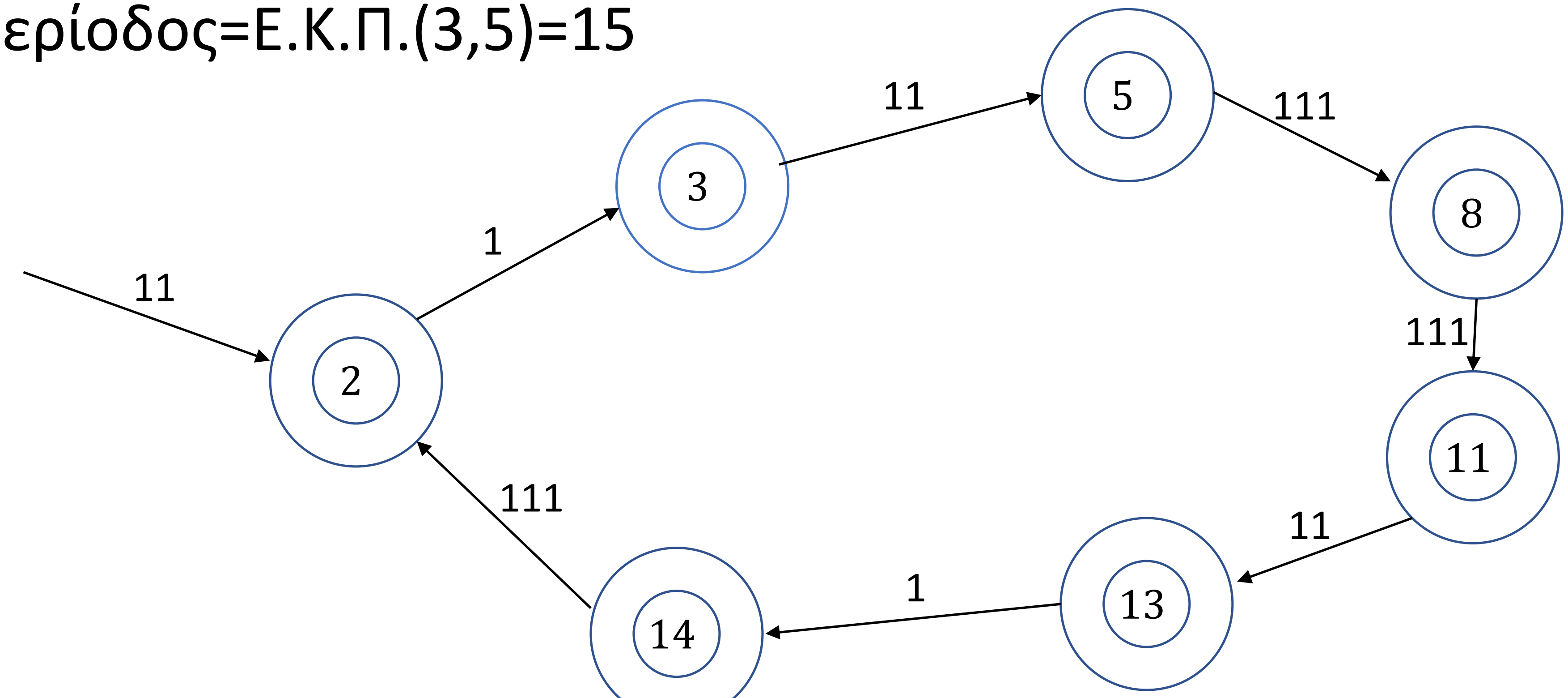
- $L_1$={$\mathrm{a}^{\mathrm{n}}\mathrm{b}^{\mathrm{n}}$}
- $L_2$ ={$\mathrm{a}^{\mathrm{m}}\mathrm{b}^{\mathrm{n}}$| m>n≥0}
- $L_3$ ={$\mathrm{w} \in \{\mathrm{a}, \mathrm{b}\}^*$| $\#\mathrm{a} \neq \#\mathrm{b}$}
- $L_4 = \{a^n b^n c^n\}$ ;
- $L_5 = \{a^{2^n}, n \in \mathbb{N}\}$ ;

# 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={3+5l}
- 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

Ας πάμε στη σελίδα 195/373 των σημειώσεων του μαθήματος για να δούμε το pumping lemma για context free languages.

Εμείς θα χρησιμοποιήσουμε το $n' = 2^{2k}$ ώστε να υπάρχει η ιδιότητα της γραμμικότητας που θέλουμε (δεν κάνουμε pump μέχρι πάνω-πάνω, εξήγηση στο σχήμα).

Το mapping όλων των λέξεων με $|w| \leq n'$ είναι ένα ημιγραμμικό σύνολο.

Αφού κουτσουρέψαμε το δέντρο, το αφήνουμε να ξαναμεγαλώσει (με ακριβώς την αντίστροφη διαδικασία).

Σχήμα


"Κόβουμε" το δέντρο μέχρι το μήκος να γίνει μικρότερο του $n' = 2^{2k}$.

Για όλες τις λέξεις $g$ με μήκος $\leq n'$, βρίσκω τα $\Phi(g)$.

Για να παράγω την "παλιά" (μεγάλη) λέξη κάνω τα αντίστροφα βήματα.

Όταν από το $u'v'w'x'y$ πάω στο $u'(v')^i w'(x')^i y$, το mapping είναι το $\phi(u') + i\,\phi(v') + \phi(w') + i\,\phi(x') + \phi(y)$, το οποίο είναι γρ. συνδυασμός του προηγούμενου και άρα θα ανήκει στο ημιγραμμικό σύνολο που ορίζεται από τα $\Phi(g), |g| \leq n'$.

Κάθε λέξη προκύπτει ως τέτοιος γρ. συνδυασμός εξαιτίας του P.L.

Γιατί πρέπει $n' = 2^{2k}$ και όχι $n' = n = 2^k$; (δυνατότητα παραγωγής όλων των συμβολοσειρών)

Δηλαδή τα mappings των λέξεων της context free γλώσσας είναι ένα ημιγραμμικό σύνολο!

Edit: παραγάγω και όχι παράξω στην 9