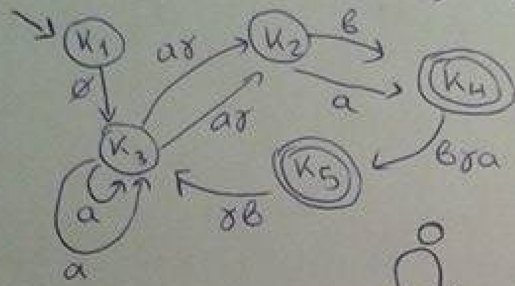
Ομαλές Γραμματικές - Πεπερασμένα αυτόματα

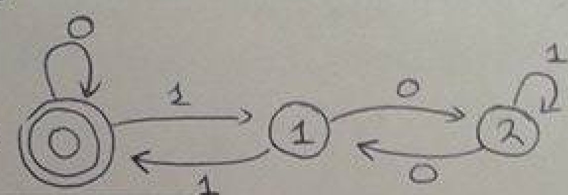
4άδα $G = \langle \Sigma, \Sigma_K, I, \Delta \rangle$ Σ = αλφάβητο, τερματικά, Σ_K = παραγωγικά, I = αρχικό

Δ = κανόνες : $K \to \varnothing$ ή $K \to \lambda K'$

παράδειγμα:
Αποδεκτικός περίπατος αααγββγαγβαγα



Δυαδικές παραστάσεις πολλαπλασίων του 3 } =>



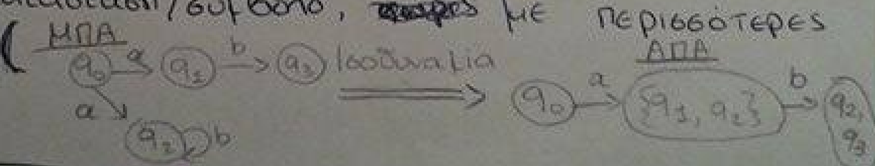
π.χ : $0011 = 3$
$1111 = 15$
$1001 = 9$

Ισοδυναμία αιτιοκρατικών και μη

Για οποιοδήποτε μη αιτιοκρατικό αυτόματο με n καταστάσεις υπάρχει ένα αιτιοκρατικό με $\leq 2^n$ καταστάσεις για την ίδια ακριβώς γλώσσα.

Μη αιτιοκρατικά αυτόματα

- υπάρχουν μεταβάσεις οφειλόμενες σε μια λέξη $\lambda \in \Sigma^*$ και όχι σε ακριβώς ένα σύμβολο $\sigma \in \Sigma$. (Ισοδυναμία -> σπάμε τη λέξη σε επιμέρους διαδρομές).
- υπάρχουν ζεύγη : κατάσταση/σύμβολο χωρίς καμία επόμενη κατάσταση. (Ισοδυναμία -> χρησιμοποιούμε μία extra κατάσταση (απορριπτική)).
- υπάρχουν κενές μεταβάσεις ($X \xrightarrow{\sigma} K \xrightarrow{\varnothing} K' \overset{\text{Ισοδύν.}}{\Longrightarrow} X \xrightarrow{\sigma} K'$).
- υπάρχουν ζεύγη κατάσταση/σύμβολο, με περισσότερες από μία καταστάσεις (



Κλειστότητα - Ομαλές γλώσσες - αυτόματα

Ιδιότητα, σύμφωνα με την οποία, 2 μέρη ενός συνόλου μετά από μία "πράξη" παράγουν πάλι μέρος του ίδιου συνόλου.

- Ένωση: $L(A) \cup L(B)$ q' ενώνουμε με μία τυχαία αρχική από q_A, q_B
- Παράθεση: $L(A) \| L(B)$ q' ενώνουμε τελικούς q_A με αρχικές q_B π.χ ($\{a, b, ab\} \| \{c, bc\} \to \{ac, abc, bc, bbc, abbc\}$)
- συμπλήρωμα: $\Sigma^* - L(A)$ q' : τελική q_A με αρχική q_B και τελική q_B με αρχική q_A
- τομή: $L(A) \cap L(B)$
- κατοπτρική : $L(A)^R$ q' : αντιστρέφουμε τις φορές όλων των μεταβάσεων.

π.χ ($K_\sigma \to K'$ $\Longrightarrow$ $K'_\sigma \to K$)

* (q' είναι το παραγόμενο αυτόματο)

Λήμμα άντλησης ομαλών γλωσσών

Εάν μια γλώσσα $L(A)$ αναγνωρίζεται από αιτιοκρατικό αυτόματο, (δηλαδή είναι ομαλή), τότε:

- κάθε λέξη επαρκώς μεγάλου μήκους $\lambda \in L(A)$ $(|\lambda| \geq |\Sigma_K|)$ μπορεί να τριχοτομηθεί ως: $\underline{\lambda = a\mu^{(\kappa)}\tau}$, όπου a, μ, τ λέξεις του Σ^* και $\mu \neq \varnothing$ ($\underline{n\ |\mu| \geq 1}$)
- κάθε λέξη της μορφής $a\mu^{(i)}\tau$, $i = 0, 1, 2 \ldots$ ανήκει στη γλώσσα $L(A)$

<u>Παράδειγμα:</u>

Γλώσσα $L = \{x^{(v)} y^{(v)} : v \geq 0\}$ $(x \neq y)$

<u>Για να αποδείξω πως L δεν είναι ομαλή:</u>

Έστω L ομαλή, άρα ικανοποιεί το λήμμα για τις ομαλές και μπορεί να τριχοτομηθεί ως: $\lambda = a\mu^{(\kappa)}\tau$ και όλες οι λέξεις της μορφής $a\mu^{(\kappa)}\tau$, $\kappa \geq 0$ θα πρέπει να ανήκουν στη γλώσσα L, θα είναι δηλαδή όλες οι λέξεις της μορφής $a^{(v)}b^{(v)}$ $(\forall v)$ όμως το μεσαίο επαναλαμβανόμενο μέρος μ θα είναι $2+1$ ειδών:

- αποτελείται μόνο από X. Αδύνατο γιατί $\lambda = a\mu^{(\kappa)}\tau$ έχει ίσο πλήθος x και y, ενώ $\lambda = a\mu^{(\kappa+1)}\tau$ θα έχει περισσότερα x από y.
- αποτελείται μόνο από y. Ίδιο με πάνω (αδύνατο)
- αποτελείται από x και y. Αδύνατο γιατί π.χ η λέξη $a\mu^{(2)}\tau$ δεν έχει όλα τα x στην αρχή και όλα τα y στο τέλος ($\mu\mu = \ldots xy \ldots xy$)

$\Rightarrow$ Άτοπο, άρα L δεν είναι ομαλή!!

Άσκηση - Ασυμφραστικές Γραμματικές

<u>Σχεδιάστε τις γραμματικές</u>

I. $\underline{\{\lambda \in \{a, b\}^* : (\text{πλήθος } a \text{ στη } \lambda) = 2 \times (\text{πλήθος } b \text{ στη } \lambda)\}}$

<u>Λύση</u>

$I \to aA_{-1} \mid bA_{+2} \mid \varnothing$

$A_{+1} \to aI \mid bA_{+1}A_{+2}$, $A_{+2} \to aA_{+1} \mid bA_{+2}A_{+2}$

$A_{-1} \to aA_{2} \mid bA_{-1}$, $A_{-2} \to IbI \mid A_{+1}bA_{-1}$

II. $\underline{\{\lambda \in \{a, b, \gamma\}^* : \text{η } \lambda \text{ έχει τη μορφή } xyx^{(R)}, x \in \{a, b\}^*\}}$

<u>Λύση</u> $I \to aIa \mid bIb \mid \gamma I\gamma \mid \gamma$

III. $\underline{\{\lambda \in \{a, b\}^* : \lambda = \lambda^{(R)}\}}$

<u>Λύση</u>

$I = aIa \mid bIb \mid a \mid b \mid \varnothing$

Myhill-Nerode

Έστω μια γλώσσα $L \subseteq \Sigma^*$ και έστω η εξής σχέση $\approx_L$ μεταξύ των λέξεων του Σ^* προσδιορισμένη με βάση την γλώσσα L: $(\lambda \approx_L \lambda') \Leftrightarrow (\forall \tau \in \Sigma^* (\lambda\tau \in L \Leftrightarrow \lambda'\tau \in L))$. Τότε

- Η σχέση $\approx_L$ είναι σχέση ισοδυναμίας επί των λέξεων του Σ^*
- Η γλώσσα L αναγνωρίζεται από κάποιο αυτόματο εάν και μόνο εάν το πλήθος των κλάσεων ισοδυναμίας της $\approx_L$ είναι πεπερασμένο.

<u>Παράδειγμα:</u> Έστω γλώσσα Λ επί του $\{a, b\}$ που αποτελείται από όλες τις λέξεις εκτός από όλες όσες περιέχουν την λέξη $\underline{abbbbb}$. Η κατά Myhill-Nerode κλάση ισοδυναμίας της λέξης $\mu = bba\underline{bbb}$ ως προς Λ είναι:

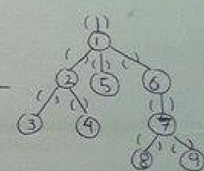
S = «όλες οι λέξεις της μορφής $X\ \underline{abbb}$ όπου x μια οποιαδήποτε λέξη της Λ»

<u>Ρόλος Myhill-Nerode:</u> μικρότερος αλγόριθμος-αυτόματο που επιλύει ένα πρόβλημα.

Ισορροπημένες παρενθέσεις

<u>έρριζο δένδρο</u>

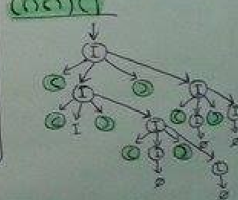
$$((()())()((()())))$$



(Γραμματική) - Κανόνας:

$\{I \to (I)I \mid \varnothing\}$

<u>δένδρο για τη λέξη</u> $\underline{(()())()}$



$\boxed{L = \{x^{(v)} y^{(v)} : v \geq 0\}}$ (ίσο πλήθος x, y)

Γραμματική: Τερματικά, Σ $\{a, b\}$

Παραγωγικά, Σ_K $\{I, A, B\}$

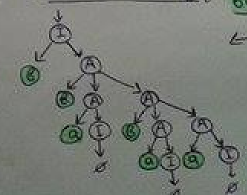
Αρχικό σύμβολο I

Κανόνες: $I \to aB \mid bA \mid \varnothing$

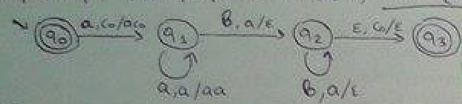
$A \to aI \mid bAA$

$B \to bI \mid aBB$

δένδρο: <u>όσα a τόσα b</u>: $bbabaa$



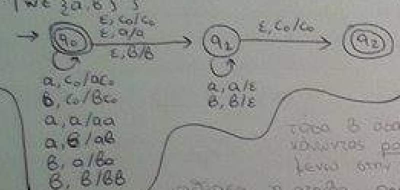
Αυτόματα με στοίβα Βάζουμε μια στοίβα για να επιλύσουμε το πρόβλημα καταμέτρησης πλήθους, π.χ $L=\{a^n b^n, n\geq 0\}$



Από την q_0 βλέπω a και ξεκινώ κάποια πρώτη θέση στοίβας με a (c0/ac0) και πάω στην q_1 όπου για κάθε a, βάζω στη στοίβα που είχε a (a,a/aa) ώσπου με b πάω στην q_2 και για κάθε b μέχρι να πάρω τόσα b όσα a είχε η στοίβα κάνοντας pop τα a (a/ε) μένω στην q_2 και όταν αδειάσει η στοίβα πάω στην q_3

π.χ 2

$L=\{ww^R \mid w\in\{a,b\}^*\}$



Αριστερή αναδρομή: $Q \to QY$ ← Q επαναφέρει τον εαυτό του και δεν μπορούμε να ξέρουμε αν κ' πότε θα σταματήσει.

Μετατροπή: ασυμφραστική γραμματική $\xrightarrow{\text{σε}}$ αυτιοκρατική μορφή (δηλ να αναγνωρίζεται από αιτιοκρ. αυτόματο).

1ο τέχνασμα: $Q\to \lambda Q_1$ και $Q\to\lambda Q_2$, $\lambda\in\Sigma^*$

γίνεται: $\Leftrightarrow$ $Q\to\lambda Q'$ και $Q'\to Q_1|Q_2$

2ο τέχνασμα: $Q\to QY \mid X$ (παράγει λέξεις πχ: XY...YYY.)

γίνεται $\Rightarrow$ $Q\to X|XQ'$, $Q'\to YQ'|\varnothing$

Chomsky

Κάθε ασυμφραστική γραμματική G επιδέχεται μια ισοδύναμη διατύπωση G' (L(G)=L(G')) στην οποία οι κανόνες έχουν την εξής μορφή:
- $Q\to Q_1Q_2$ { κάθε παραγωγικό σύμβολο είτε αντικαθίσταται από 2 παραγωγικά ή από τερματικό
- $Q\to\lambda$ { δηλαδή μια λέξη λ (ίσως κενή).

Λήμμα Άντλησης ασυμφραστικών ~~γραμματικών~~ γλωσσών

L(G). G ασυμφραστική γραμματική, υπάρχει ένας αριθμός, όριο n(G), ώστε $\forall\ \lambda\in L(G)$ με μήκος τουλ n(G) να ισχύουν:
- η λέξη λ γράφεται ως xaybz, έτσι ώστε κάθε λέξη της μορφής $xa^{(k)}yb^{(k)}z$, $k\geq 0$, να ανήκει στην γλώσσα $\{xa^{(k)}yb^{(k)}z : k\geq 0\}\subseteq L(G)$
- Ένα τουλάχιστον των τμημάτων a ή b είναι μη κενό, δηλαδή $|ab|\geq 1$.

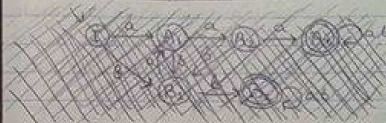
$R^+ = R\cdot R^* =$ επανάληψη 1 ή περισσότερες φορές. $\boxed{OY}$

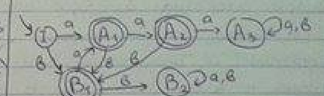
$|n\geq 1|$ $L=(ab)^+$

μη ντετερμινιστικό

Γλώσσα: $\{\lambda\in\{a,b\}^*$: η λέξη λ δεν περιέχει ούτε την υπολέξη aaa ούτε την bb$\}$

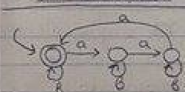
Αιτιοκρατικό αυτόματο:



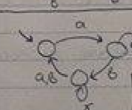


κάθε a πρέπει να κλείνει σαν να ήταν αριστερή παρένθεση από ένα b, σαν να ήταν δεξιά παρένθεση (3 επίπεδα παρενθέσεων)

μη-αιτιοκρατικό (πρώτη κατάσταση έχει λούπ προς α επόμενη)



το πλήθος των a είναι πολλαπλάσιο του 3. αιτιοκρατικό



Αποδέχεται την κενή γλώσσα (δεν έχει τερματική κατάσταση) (μη αιτιοκρατικό)

Η τελευταία κατάσταση δεν χρησιμεύει σε τίποτα, το αυτόματο αποδέχεται οποιαδήποτε λέξη αιτιοκρατικό

Δείξτε ότι η γλώσσα είναι ομαλή (χωρίς αυτόματο)

$L=\{\lambda\in\{a,b\}^*$: η λέξη λ δεν περιέχει ούτε την υπολέξη aaa ούτε bb$\}$

Απάντηση: Η γλώσσα των λέξεων που περιέχουν έστω μια φορά το «aaa» είναι η $\Sigma^*aaa\Sigma^*$. Το συμπλήρωμα της δεν περιέχει καμία εμφάνιση του aaa και είναι η γλώσσα $\Lambda_a=(\Sigma^*-(\Sigma^*aaa\Sigma^*))$. Αναλόγως έχουμε και την $\Lambda_b=(\Sigma^*-(\Sigma^*bb\Sigma^*))$, χωρίς κανένα bb. Η γλώσσα L είναι η $\Lambda_a\cap\Lambda_b$. Από τους κανόνες κλειστότητας η L προκύπτει ομαλή.

κλειστότητα ως προς την ένωση!

Έστω ότι η L γλώσσα επί του αλφαβήτου Σ είναι ομαλή, και $M \subseteq \Sigma^*$
Δείξτε ότι και η εξής γλώσσα θα είναι ομαλή:

• Πρόθεμα $(L) \equiv \{\lambda : \lambda \in L$, και υπάρχει τέλος $\tau \in \Sigma^*$ για τη λ ώστε $\lambda\tau \in L\}$

Απάντηση:

Εξετάζουμε για κάθε κατάσταση K του αυτόματου εάν σε αυτό υπάρχει έστω μία διαδρομή από την K έως μία αποδεκτική κατάσταση (μία διαδρομή που θα έγραφε μία λέξη «τ»). Τρέπουμε αυτή την κατάσταση σε αποδεκτική, διότι αν φτάσουμε σε αυτή τη λέξη λ τότε υπάρχει η συνέχεια τ ώστε να αποδεχθούμε τη λέξη $\lambda\tau$.

δεν μπορεί να συγκρίνει 2 αριθμούς (να δει πλήθος α και β ώστε να τα βγάλει ίσα)

• $L_? = \{x^{(v)} y^{(v)} : v \geq 0\}$ $(x \neq y)$ → μη ομαλή γλώσσα

• $L = \{\lambda : [^{(v)}]^{(v)} : v \geq 1\}$ → μη ομαλή γλώσσα (κλειστότητα)

• $L = \{\lambda : [^{(\kappa)}]^{(\lambda)} : \kappa, \lambda \geq 0\}$ → ομαλή γλώσσα

• $L_{πρώτοι} = \{a^{(p)} : p$ πρώτος αριθμός$\}$ → μη ομαλή γλώσσα (λήμμα)

• $L_+ = \{\lambda : \lambda \in \Sigma^*$ λ παριστά μια σωστή πρόσθεση δυαδικών$\}$ → ομαλή γλώσσα
αποδείξη με αυτόματο με καταστάσεις για ύπαρξη ή όχι κρατουμένου και θεώρημα κλειστότητας ως προς κατοπτρισμό $L^{(R)}$ γιατί διαβάζει από δεξιά → αριστερά.

• $L_\times = \{\lambda : \lambda \in \Sigma^*$ λ παριστά σωστό πολλ/σμο δυαδικών$\}$ → μη ομαλή γλώσσα
ίδιο με $\{x^{(v)} y x^{(v)} : v \geq 0\}$ → μη ομαλό

■ Δεν υπάρχει διαγνωστικό των ολικών προγραμμάτων

■ Ο τερματισμός ενός προγράμματος π επί δεδομένων δ είναι αλγοριθμικά άγνωστος

■ Υπάρχει συνάρτηση $\Sigma : \varphi^* \to \varphi^*$, ολικώς ορισμένη, υπολογίσιμη, αλλά δεν υπολογίζεται από καμία πρωτογενώς αναδρομική περιγραφή.

S = ασυμφραστική γλώσσα, L = ομαλή γλώσσα

$S - L$ = πάντα ασυμφραστική
• (ασυμφραστικές κλειστές ως προς συμπληρώματα με ομαλές)
$L - S$ = μπορεί να είναι ομαλή