

Θεωρία Υπολογισμού

Ιδιότητες των Γλωσσών χωρίς Συμφραζόμενα

2003-04

Ιδιότητες Κλειστότητας

- **Θεώρημα.** Οι γλώσσες χωρίς συμφραζόμενα είναι κλειστές ως προς τις πράξεις ένωση, παράθεση και κλειστότητα.

Απόδειξη:

- Έστω γλώσσες χωρίς συμφραζόμενα L_1 και L_2 . Έστω γραμματικές χωρίς συμφραζόμενα $G_1 = (V_1, \Sigma_1, R_1, S_1)$ και $G_2 = (V_2, \Sigma_2, R_2, S_2)$ τέτοιες ώστε $L_1 = L(G_1)$ και $L_2 = L(G_2)$.

- Υποθέτουμε ότι τα $V_1 \setminus \Sigma_2$ και $V_2 \setminus \Sigma_1$ είναι ξένα μεταξύ τους.

- Ένωση. Έστω S ένα νέο σύμβολο και

$$G = (V_1 \cup V_2 \cup \{S\}, \Sigma_1 \cup \Sigma_2, R_1 \cup R_2 \cup \{S \rightarrow S_1, S \rightarrow S_2\}, S).$$

$$\text{Τότε } L(G_1) \cup L(G_2) = L(G).$$

Συνέχεια της Απόδειξης

- Οι μοναδικοί κανόνες που αφορούν το S είναι οι $S \rightarrow S_1$ και $S \rightarrow S_2$. Άρα $S \Rightarrow_G^* w$, όπου $w \in (\Sigma_1 \cup \Sigma_2)^*$, ανν $S_1 \Rightarrow_G^* w$ ή $S_2 \Rightarrow_G^* w$.
- Και εφόσον οι G_1 και G_2 έχουν ξένα μεταξύ τους σύνολα μη τελικών συμβόλων, η τελευταία διάζευξη είναι ισοδύναμη με $S_1 \Rightarrow_{G_1}^* w$ ή $S_2 \Rightarrow_{G_2}^* w$.
- Ισοδύναμα $w \in L(G_1) \cup L(G_2)$.

Συνέχεια της Απόδειξης

- **Παράθεση:** Η $L(G_1)L(G_2)$ παράγεται από τη γραμματική

$$(V_1 \cup V_2 \cup \{S\}, \Sigma_1 \cup \Sigma_2, R_1 \cup R_2 \cup \{S \rightarrow S_1 S_2\}, S).$$

- **Κλειστότητα:** Η $L(G_1)^*$ παράγεται από την

$$(V_1, \Sigma_1, R_1 \cup \{S_1 \rightarrow e, S_1 \rightarrow S_1 S_1\}, S_1).$$

Ιδιότητες Κλειστότητας

- **Θεώρημα:** Η τομή μιας γλώσσας χωρίς συμφραζόμενα με μια κανονική γλώσσα είναι γλώσσα χωρίς συμφραζόμενα.
- **Απόδειξη:** Αν η L είναι μια γλώσσα χωρίς συμφραζόμενα και η R είναι μια κανονική γλώσσα, τότε $L = L(M_1)$ για κάποιο αυτόματο στοίβας $M_1 = (K_1, \Sigma_1, \Gamma_1, \Delta_1, s_1, F_1)$ και $R = L(M_2)$ για κάποιο ντετερμινιστικό πεπερασμένο αυτόματο $M_2 = (K_2, \Sigma_2, \delta_2, s_2, F_2)$.
- Θα συνδυάσουμε αυτές τις δύο μηχανές σε ένα μόνο αυτόματο στοίβας M που εκτελεί παράλληλα τους υπολογισμούς των M_1 και M_2 και δέχεται μια συμβολοσειρά αν και τα δύο θα τη δεχόταν.

Συνέχεια της Απόδειξης

- Ειδικότερα, έστω $M = (K, \Sigma, \Gamma, \Delta, s, F)$, όπου

$$K = K_1 \times K_2$$

$$\Sigma = \Sigma_1 \cup \Sigma_2$$

$$\Gamma = \Gamma_1$$

$$s = (s_1, s_2)$$

$$F = F_1 \times F_2$$

Συνέχεια της Απόδειξης

- Η σχέση μετάβασης Δ ορίζεται ως εξής. Για κάθε μετάβαση

$$((q_1, u, \beta), (p_1, \gamma)) \in \Delta_1$$

και κάθε κατάσταση q_2 του M_2 τέτοια ώστε

$$(q_2, u) \vdash_{M_2}^* (p_2, e)$$

προσθέτουμε τη μετάβαση

$$(((q_1, q_2), u, \beta), ((p_1, p_2), \gamma))$$

στο Δ .

- **Παρατήρηση:** Για να βρούμε την p_2 αρκεί να προσομοιώσουμε το M_2 για $|u|$ βήματα με είσοδο u .

Παράδειγμα

- Έστω η γλώσσα L με αλφάβητο $\{a, b\}$ που αποτελείται από όλες τις συμβολοσειρές με ίσο αριθμό a και b που δεν περιέχουν τις υποσυμβολοσειρές $abaa$ ή $babb$.

- Η L είναι γλώσσα χωρίς συμφραζόμενα, αφού αποτελεί τομή της γλώσσας

$$\{w \in \{a, b\}^* : w \text{ έχει ίσο αριθμό } a \text{ και } b\}.$$

με την κανονική γλώσσα $\overline{\{a, b\}^* \{abaa, babb\} \{a, b\}^*}$.

Μερικοί Ορισμοί

- **Ορισμός.** Έστω $G = (V, \Sigma, R, S)$ μια γραμματική χωρίς συμφραζόμενα. Με το $\phi(G)$ συμβολίζουμε το μεγαλύτερο αριθμό συμβόλων στο δεξιό μέρος των κανόνων του R .
- **Ορισμός.** Αν A είναι ένα σύνολο θα ονομάζουμε πληθικό αριθμός του A (συμβολισμός: $|A|$) τον αριθμό των στοιχείων του.
- **Παραδείγματα:** $|\{a, b, c\}| = 3, |\emptyset| = 0$

Μερικοί Ορισμοί

- **Ορισμός.** Ένα μονοπάτι σε ένα συντακτικό δέντρο είναι μία ακολουθία διακριτών κόμβων όπου ο καθένας συνδέεται με τον προηγούμενο του με ένα ευθύγραμμο τμήμα, ο πρώτος κόμβος είναι η ρίζα και ο τελευταίος κόμβος είναι ένα φύλλο.
- Το **μήκος** ενός μονοπατιού είναι ο αριθμός των ευθυγράμμων τμημάτων του (δηλ. ο αριθμός των κόμβων μείον ένα.)
- Το **ύψος** ενός συντακτικού δέντρου είναι το μήκος του μεγαλύτερου μονοπατιού του.

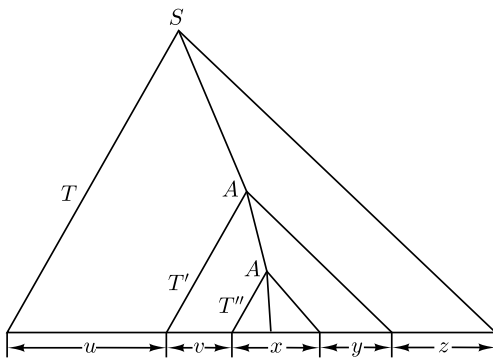
Ένα Θεώρημα Άντλησης

- **Λήμμα:** Έστω G μια γραμματική χωρίς συμφραζόμενα. Το προϊόν ενός συντακτικού δέντρου της G με ύψος h έχει μήκος το πολύ $\phi(G)^h$.
- **Απόδειξη:** Επαγωγή ως προς h .
- **Θεώρημα:** Έστω G μια γραμματική χωρίς συμφραζόμενα. Τότε κάθε συμβολοσειρά $w \in G$ με μήκος μεγαλύτερο από $\phi(G)^{|V \setminus \Sigma|}$ μπορεί να ξαναγραφεί σαν $w = uvxyz$ με τέτοιο τρόπο ώστε είτε η v είτε η y είναι μη κενές και η uv^nxy^nz ανήκει στην $L(G)$ για κάθε $n \geq 0$.

Απόδειξη

- Έστω w μια τέτοια συμβολοσειρά, και T ένα συντακτικό δέντρο με ρίζα S και προϊόν w , το οποίο έχει το **μικρότερο αριθμό φύλλων** ανάμεσα σε όλα τα συντακτικά δέντρα με την ίδια ρίζα και το ίδιο προϊόν.
- Επειδή το προϊόν του T είναι μεγαλύτερο από $\phi(G)^{|V \setminus \Sigma|}$, συνεπάγεται ότι το T έχει ένα μονοπάτι με μήκος τουλάχιστον $|V \setminus \Sigma| + 1$ δηλαδή με τουλάχιστον $|V \setminus \Sigma| + 2$ κόμβους.
- Σ' αυτό το μονοπάτι μόνο ένας κόμβος περιέχει τερματικό σύμβολο – οι υπόλοιποι κόμβοι περιέχουν μη τερματικά σύμβολα. Αφού λοιπόν έχουμε περισσότερους κόμβους από τερματικά σύμβολα, αυτό το μονοπάτι περιλαμβάνει δύο κόμβους που έχουν το ίδιο μη τερματικό σύμβολο A .

Συνέχεια της Απόδειξης



Συνέχεια της Απόδειξης

- Στην προηγούμενη εικόνα δείχνουμε το συντακτικό δέντρο T με $w = uvxyz$.
- Τώρα είναι φανερό ότι το μέρος του υποδέντρου T' που δεν περιέχει το T'' μπορεί να επαναληφθεί όσες φορές θέλουμε (μηδέν φορές ή περισσότερες) για να παράγουμε άλλα συντακτικά δέντρα με προϊόν της μορφής uv^nxy^nz , $n \geq 0$.
- Το μόνο που μένει τώρα να αποδείξουμε είναι ότι $vy \neq \epsilon$. Αλλά αν ήταν $vy = \epsilon$ τότε υπάρχει ένα συντακτικό δέντρο με ρίζα S και προϊόν w με λιγότερα φύλλα από το T (αυτό το δέντρο προκύπτει αν παραλείψουμε από το δέντρο T το T' και προσθέσουμε το T''). Αυτό όμως δεν είναι δυνατό από την αρχική μας υπόθεση.

Εφαρμογές του Θεωρήματος Άντλησης

- Η γλώσσα $L = \{a^n b^n c^n : n \geq 0\}$ δεν είναι γλώσσα χωρίς συμφραζόμενα.

Απόδειξη:

- Ας υποθέσουμε $L = L(G)$ για κάποια γραμματική G χωρίς συμφραζόμενα.
- Έστω $n > \phi(G)^{|V \setminus \Sigma|/3}$. Τότε η συμβολοσειρά $w = a^n b^n c^n$ ανήκει στην $L(G)$ και έχει μήκος μεγαλύτερο από $\phi(G)^{|V \setminus \Sigma|}$.
- Από το θεώρημα άντλησης έχουμε ότι η w έχει μια αναπαράσταση της μορφής $uvxyz$ τέτοια ώστε είτε η v είτε η y είναι μη κενές και $uv^i xy^i z$ ανήκει στην $L(G)$ για κάθε $i \geq 0$.

Συνέχεια της Απόδειξης

- Τώρα υπάρχουν οι εξής περιπτώσεις:
 - Η συμβολοσειρά vy περιέχει όλα τα σύμβολα a, b και c . Τότε είτε η v είτε η y περιέχει τουλάχιστον δύο σύμβολα από τα $\{a, b, c\}$ οπότε η $uv^2 xy^2 z$ περιέχει δύο σύμβολα που δεν είναι στη σειρά που δίνεται από τον ορισμό της γλώσσας. Αδύνατο!
 - Η συμβολοσειρά vy περιέχει κάποια αλλά όχι όλα τα σύμβολα a, b και c . Τότε η συμβολοσειρά $uv^2 xy^2 z$ δεν περιέχει ίσο αριθμό από a, b και c . Αδύνατο!

Παράδειγμα

- Η γλώσσα

$$L = \{a^n : n \text{ είναι πρώτος αριθμός}\}$$

δεν είναι γλώσσα χωρίς συμφραζόμενα.

- Απόδειξη: Χρησιμοποιείστε το θεώρημα άντλησης.

Ιδιότητες Μη Κλειστότητας

- **Θεώρημα:** Οι γλώσσες χωρίς συμφραζόμενα δεν είναι κλειστές ως προς την τομή και το συμπλήρωμα.

- **Απόδειξη.**

- **Τομή:** Είναι φανερό ότι οι γλώσσες

$$\{a^n b^n c^m : m, n \geq 0\}$$

και

$$\{a^m b^n c^n : m, n \geq 0\}$$

είναι γλώσσες χωρίς συμφραζόμενα.

- Η τομή των δύο αυτών γλωσσών είναι η γλώσσα $\{a^n b^n c^n : n \geq 0\}$ η οποία δεν είναι γλώσσα χωρίς συμφραζόμενα.

Συνέχεια της Απόδειξης

- **Συμπλήρωμα:** Αν οι γλώσσες χωρίς συμφραζόμενα ήταν κλειστές ως προς το συμπλήρωμα, θα ήταν κλειστές και ως προς την τομή επειδή

$$L_1 \cap L_2 = \Sigma^* \setminus ((\Sigma^* \setminus L_1) \cup (\Sigma^* \setminus L_2)).$$

Αλγοριθμικές Ιδιότητες

- **Θεώρημα:** Υπάρχει αλγόριθμος που για μια γραμματική G και συμβολοσειρά w απαντά στο ερώτημα: “Ανήκει η w στην $L(G)$;”
- **Παρατήρηση:** Το παραπάνω ερώτημα είναι πολύ σημαντικό. Ο αλγόριθμος που θα παρουσιάσουμε μπορεί να χρησιμοποιηθεί από ένα συντακτικό αναλυτή (parser) για τη γλώσσα που παράγεται από την G .

Η κανονική μορφή Chomsky

- **Ορισμός.** Μια γραμματική χωρίς συμφραζόμενα $G = (V, \Sigma, R, S)$ είναι σε κανονική μορφή Chomsky αν $R \subseteq (V \setminus \Sigma) \times V^2$.
- **Διαισθητικά:** Μια γραμματική χωρίς συμφραζόμενα είναι σε κανονική μορφή Chomsky αν το δεξιό μέρος κάθε κανόνα έχει μήκος 2.
- **Παράδειγμα:** Η γραμματική $G = (\{S, N, a, b, c\}, \{a, b, c\}, R, S)$ όπου

$$R = \{ S \rightarrow aN, N \rightarrow NN, N \rightarrow bc \}$$

είναι σε κανονική μορφή Chomsky.

Παράδειγμα

- Έστω η γραμματική με κανόνες

$$S \rightarrow SS, S \rightarrow (S), S \rightarrow e.$$

- Η γραμματική αυτή δεν είναι σε κανονική μορφή Chomsky.

Η κανονική μορφή Chomsky

- **Θεώρημα.** Για κάθε γραμματική χωρίς συμφραζόμενα G υπάρχει μια γραμματική χωρίς συμφραζόμενα G' σε κανονική μορφή Chomsky τέτοια ώστε $L(G') = L(G) \setminus (\Sigma \cup \{e\})$.

Η κανονική μορφή Chomsky

- Θα δώσουμε ένα αλγόριθμο για τη μετατροπή μιας γραμματικής χωρίς συμφραζόμενα $G = (V, \Sigma, R, S)$ σε κανονική μορφή Chomsky.
- Υπάρχουν τρία είδη κανόνων που παραβιάζουν τους περιορισμούς της κανονικής μορφής Chomsky:
 - κανόνες που έχουν δεξιό μέρος με μήκος ≥ 3 .
 - κανόνες της μορφής $A \rightarrow e$.
 - κανόνες της μορφής $A \rightarrow x$ όπου $x \in V$.
- Τα 3 βήματα του αλγόριθμου απαλείφουν κανόνες όπως τους παραπάνω.

Η κανονική μορφή Chomsky

- **Βήμα 1:** Απαλοιφή κανόνων που έχουν δεξιά μέρος με μήκος ≥ 3 .
- Έστω ο κανόνας

$$A \rightarrow B_1 B_2 \dots B_n$$

όπου $B_1, B_2, \dots, B_n \in V$ και $n \geq 3$.

- Αντικαθιστούμε αυτό τον κανόνα με τους ακόλουθους $n - 1$ νέους κανόνες:

$$A \rightarrow B_1 A_1, A_1 \rightarrow B_2 A_2, \dots, A_{n-2} \rightarrow B_{n-1} B_n$$

όπου A_1, \dots, A_{n-2} είναι νέα μη τερματικά σύμβολα.

Παράδειγμα

- Έστω η γραμματική με κανόνες

$$S \rightarrow SS, S \rightarrow (S), S \rightarrow e.$$

- Ο προβληματικός κανόνας είναι ο $S \rightarrow (S)$. Μπορεί να αντικατασταθεί από τους ισοδύναμους κανόνες

$$S \rightarrow (S_1, S_1 \rightarrow S).$$

Η κανονική μορφή Chomsky

- **Βήμα 2:** Απαλοιφή κανόνων της μορφής $A \rightarrow e$.
- Πρώτα υπολογίζουμε το σύνολο μη τερματικών συμβόλων που μπορούν να παράγουν την κενή συμβολοσειρά:

$$\mathcal{E} = \{A \in V \setminus \Sigma : A \Rightarrow^* e\}$$

- Ο υπολογισμός του \mathcal{E} μπορεί να γίνει με τον ακόλουθο αλγόριθμο:
 $\mathcal{E} := \emptyset$

while υπάρχει κανόνας $A \rightarrow \alpha$ και $\alpha \in \mathcal{E}^*$ **do**
 πρόσθεσε το A στο \mathcal{E}

- Αφού υπολογίσουμε το σύνολο \mathcal{E} κάνουμε τα εξής:
 - Απαλείφουμε όλους τους κανόνες της μορφής $A \rightarrow e$.
 - Για κάθε κανόνα της μορφής $A \rightarrow BC$ ή $A \rightarrow CB$ με $B \in \mathcal{E}$ και $C \in V$ προσθέτουμε στη γραμματική τον κανόνα $A \rightarrow C$.

Παράδειγμα

- Συνεχίζουμε με τη γραμματική με κανόνες

$$S \rightarrow SS, S \rightarrow (S_1, S_1 \rightarrow S), S \rightarrow e.$$

- Παρατηρήστε ότι $\mathcal{E} = \{S\}$. Αρα το νέο σύνολο κανόνων που μας δίνει το βήμα 2 είναι:

$$S \rightarrow SS, S \rightarrow (S_1, S_1 \rightarrow S), S \rightarrow S, S_1 \rightarrow S).$$

Ο κανόνας $S \rightarrow S$ μπορεί να παραλειφθεί.

Η κανονική μορφή Chomsky

- **Βήμα 3:** Απαλοιφή κανόνων της μορφής $A \rightarrow x$ όπου $x \in V$.
- Για κάθε $A \in V$ υπολογίζουμε το σύνολο $\mathcal{D}(A)$ των συμβόλων που μπορούν να παραχθούν από το A :

$$\mathcal{D}(A) = \{B \in V : A \Rightarrow^* B\}$$

- Ο υπολογισμός του $\mathcal{D}(A)$ μπορεί να γίνει με τον ακόλουθο αλγόριθμο:
 $\mathcal{D}(A) := \{A\}$

while υπάρχει κανόνας $B \rightarrow C$ και $B \in \mathcal{D}(A)$ και $C \notin \mathcal{D}(A)$ **do**
 πρόσθεσε το C στο $\mathcal{D}(A)$

- **Παρατήρηση:** Αν a είναι ένα τερματικό σύμβολο τότε $\mathcal{D}(A) = \{a\}$.

Η κανονική μορφή Chomsky

- Αφού υπολογίσουμε το σύνολο $\mathcal{D}(A)$ κάνουμε τα εξής:
 - Απαλείφουμε όλους τους κανόνες της μορφής $A \rightarrow x$ όπου $x \in V$.
 - Αντικαθιστούμε κάθε κανόνα της μορφής $A \rightarrow BC$ με όλους τους δυνατούς κανόνες $A \rightarrow B'C'$ όπου $B' \in \mathcal{D}(B)$ και $C' \in \mathcal{D}(C)$.
- Προσθέτουμε τους κανόνες $S \rightarrow BC$ για κάθε κανόνα $A \rightarrow BC$ τέτοιο ώστε $A \in \mathcal{D}(S) \setminus \{S\}$.

Παράδειγμα

- Συνεχίζουμε με τη γραμματική με κανόνες

$$S \rightarrow SS, S \rightarrow (S_1, S_1 \rightarrow S), S_1 \rightarrow S).$$

- Παρατηρήστε ότι $\mathcal{D}(S_1) = \{S_1, S\}$ και $\mathcal{D}(A) = \{A\}$ για κάθε $A \in V \setminus \{S_1\}$.
- Αρα το νέο σύνολο κανόνων που μας δίνει το βήμα 3 είναι:

$$S \rightarrow SS, S \rightarrow (S_1, S_1 \rightarrow S), S \rightarrow (S).$$

Η κανονική μορφή Chomsky

- Ας υποθέσουμε ότι έχουμε μια γραμματική χωρίς συμφραζόμενα G που είναι σε κανονική μορφή Chomsky και μια συμβολοσειρά $x = x_1 \dots x_n$ όπου $n \geq 2$.
- Ο αλγόριθμος της επόμενης σελίδας αποφασίζει αν $x \in L(G)$ χρησιμοποιώντας μια μέθοδο δυναμικού προγραμματισμού.

Αλγόριθμος

• Algorithm DECIDE

Input: Γραμματική G και συμβολοσειρά $x = x_1 \cdots x_n$

Method:

for $i := 1$ **to** n **do**

$N[i, i] := \{x_i\}$; όλα τα άλλα $N[i, j]$ είναι αρχικά \emptyset

for $s := 1$ **to** $n - 1$ **do**

for $i := 1$ **to** $n - s$ **do**

for $k := i$ **to** $i + s - 1$ **do**

if υπάρχει κανόνας $A \rightarrow BC \in R$ με $B \in N[i, k]$ και

$C \in N[k + 1, i + s]$ **then**

Πρόσθεσε το A στο $N[i, i + s]$

if $S \in N[1, n]$ **then output** " $x \in L(G)$ "

Ιδιότητες του Αλγόριθμου

- Ο αλγόριθμος DECIDE αποφασίζει αν $x \in L(G)$ αναλύοντας όλες της υποσυμβολοσειρές της x :
 - Για $s = 1$ ο αλγόριθμος αναλύει τις υποσυμβολοσειρές x_1, x_2, \dots, x_n .
 - Για $s = 2$ ο αλγόριθμος αναλύει τις υποσυμβολοσειρές $x_1x_2, x_2x_3, \dots, x_{n-1}x_n$.
 - Τελικά για $s = n - 1$ ο αλγόριθμος ασχολείται με ολόκληρη την συμβολοσειρά x .

Ιδιότητες του Αλγόριθμου

- Σε κάθε στάδιο της ανάλυσης ο αλγόριθμος υπολογίζει το σύνολο των συμβόλων της γραμματικής που μπορούν να παράγουν την αναλύμενη συμβολοσειρά.
- **Πρόταση.** Για κάθε φυσικό αριθμό s τέτοιο ώστε $0 \leq s \leq n$, μετά την s -οστή επανάληψη του αλγόριθμου DECIDE έχουμε

$$N[i, i + s] = \{A \in V : A \Rightarrow^* x_i \cdots x_{i+s}\}$$

για όλα τα $i = 1, \dots, n - s$.

Παράδειγμα

						8)	
					7)	\emptyset	
				6)	\emptyset	\emptyset	
			5	(S	S_1	\emptyset	
		4	(\emptyset	\emptyset	S	S_1	
	3)	\emptyset	\emptyset	\emptyset	\emptyset	\emptyset	
	2	(S	\emptyset	\emptyset	\emptyset	S_1	
1	(\emptyset	\emptyset	\emptyset	\emptyset	\emptyset	S	
	1	2	3	4	5	6	7	8

Τελικός Αλγόριθμος

- Αν G είναι μια γραμματική χωρίς συμφραζόμενα και x είναι μια συμβολοσειρά τότε για να αποφασίσουμε αν $x \in L(G)$ μπορούμε να κάνουμε τα εξής.
 - Μετατρέπουμε τη G σε μια γραμματική χωρίς συμφραζόμενα G' που είναι σε κανονική μορφή Chomsky.
 - Αν $|x| \leq 1$ και κατά την μετατροπή σε κανονική μορφή Chomsky είχαμε απαλείψει τον κανόνα $S \rightarrow x$ τότε $x \in L(G)$. Αλλιώς μπορούμε να χρησιμοποιήσουμε τον αλγόριθμο DECIDE.

Μελέτη

- Κεφάλαιο 3.5 από το βιβλίο:
Harry R. Lewis και Χρίστος Χ. Παπαδημητρίου, Στοιχεία Θεωρίας Υπολογισμού, Τεχνικό Επιμελητήριο Ελλάδας, 1992.