Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Διακριτές Μέθοδοι για την Επιστήμη των Υπολογιστών

Διδάσκοντες: Καθ. Φ. Αφράτη, Λεκτ. Δ. Φωτάκης

4η Σειρά Προτεινόμενων Ασκήσεων, Σχέδιο Λύσεων

Άσκηση 1 (Κανονικές Εκφράσεις). Να γράψετε κανονικές εκφράσεις για τις γλώσσες:

1. $L_1 = \{w \in \{0,1\}^* : \text{η } w \text{ δεν περιέχει τη συμβολοσειρά } 1111\}$
2. $L_2 = \{w \in \{0,1\}^* : \text{η } w \text{ περιέχει μία (ακριβώς) εμφάνιση της συμβολοσειράς } 1111\}$
3. $L_3 = \{w \in \{0,1\}^* : \text{η } w \text{ δεν περιέχει τη συμβολοσειρά } 101\}$

Λύση. Μια κανονική έκφραση που περιγράφει την L_1 είναι:

$$(0 \cup 10 \cup 110 \cup 1110)^*(\varepsilon \cup 1 \cup 11 \cup 111)$$

Μια κανονική έκφραση που περιγράφει την L_2 είναι:

$$(0 \cup 10 \cup 110 \cup 1110)^*1111(0 \cup 01 \cup 011 \cup 0111)^*$$

Για την L_3, πρέπει κάθε εμφάνιση της συμβολοσειράς 10 να ακολουθείται από τουλάχιστον ένα 0, ενώ δεν υπάρχει κάποιος περιορισμός όταν έχουμε μια ακολουθία από 0 ή μια ακολουθία από 1 ή μια ακολουθία από 0 που ακολουθείται από κάποια 1. Έτσι μια κανονική έκφραση που περιγράφει την L_3 είναι:

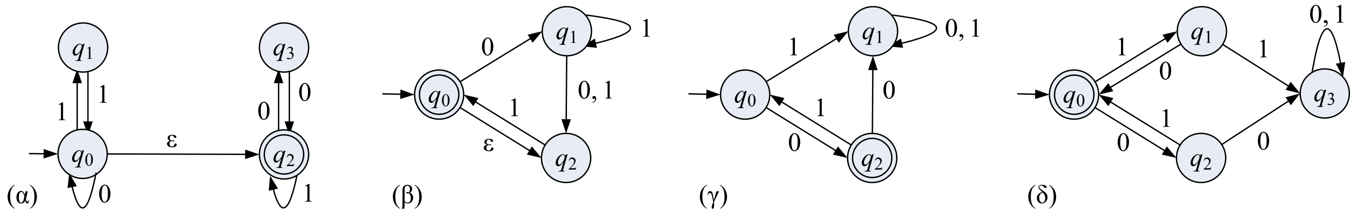
$$0^*1^*(1000^*1^*)^*0^*$$

Άσκηση 2 (Πεπερασμένα Αυτόματα). (α) Να κατασκευάσετε πεπερασμένα αυτόματα (όχι κατ' ανάγκη ντετερμινιστικά) για τις παρακάτω γλώσσες:

1. $L_1 = \{w \in \{0,1\}^* : \text{η } w \text{ περιέχει τουλάχιστον δύο εμφανίσεις της συμβολοσειράς } 1111\}$
2. $L_2 = \{w \in \{0,1\}^* : \text{η } w \text{ έχει άρτιο αριθμό 0 και δεν περιέχει τη συμβολοσειρά } 111\}$

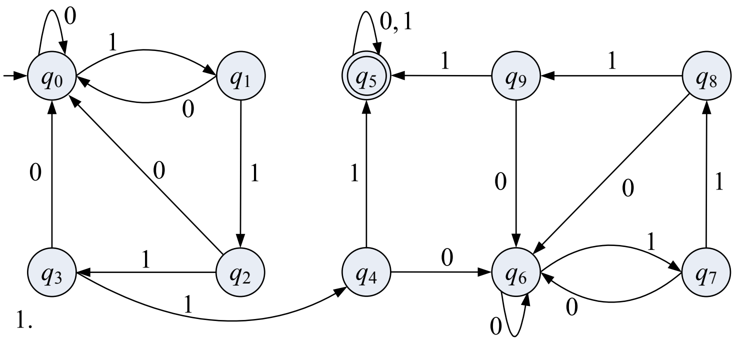
(β) Να μετατρέψετε τα μη ντετερμινιστικά αυτόματα των Σχημάτων 1.α και 1.β σε ντετερμινιστικά.

(γ) Να γράψετε κανονικές εκφράσεις για τις γλώσσες που αναγνωρίζονται από τα πεπερασμένα αυτόματα του Σχήματος 1.

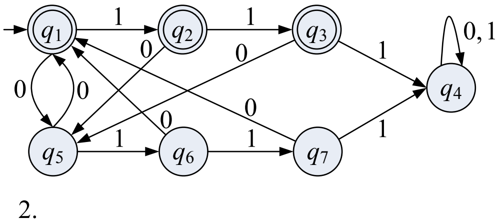


Σχήμα 1. Πεπερασμένα αυτόματα για το Θέμα 2.

Λύση. (α) (Ντετερμινιστικά) αυτόματα για τις γλώσσες L_1 και L_2 φαίνονται στο παρακάτω σχήμα. Το αυτόματο για την γλώσσα L_2 ουσιαστικά προκύπτει θεωρώντας την ‘‘τομή’’ των αυτομάτων για την γλώσσα που περιλαμβάνει τις δυαδικές συμβολοσειρές με άρτιο πλήθος 0 και την γλώσσα που περιλαμβάνει τις δυαδικές συμβολοσειρές που δεν περιέχουν το 111.

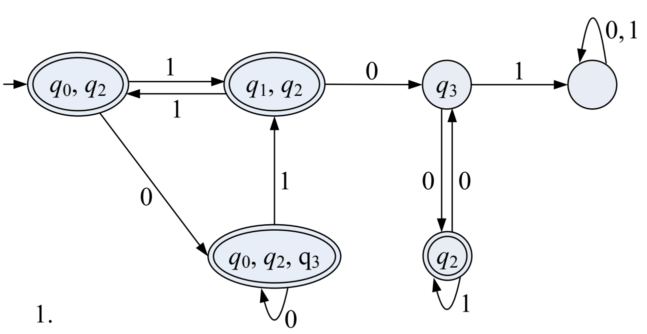


1.

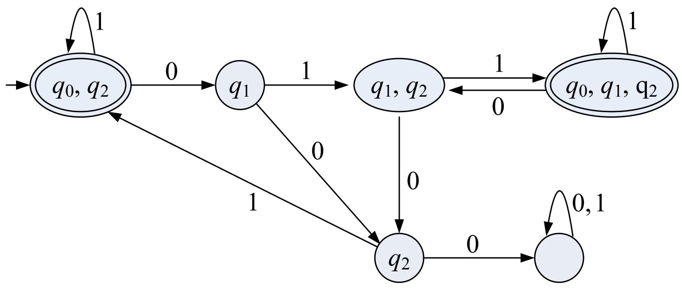


2.

(β) Το αποτέλεσμα της μετατροπής φαίνεται στο παρακάτω σχήμα:



1.



2.

(γ) Για το αυτόματο το Σχήματος 1.α είναι: $(11 \cup 0)^*(00 \cup 1)^*$. Για το αυτόματο το Σχήματος 1.β είναι: $((\varepsilon \cup 01^*(0 \cup 1))1)^*$. Για το αυτόματο το Σχήματος 1.γ είναι: $0(10)^*$. Για το αυτόματο το Σχήματος 1.δ είναι: $(10 \cup 01)^*$. □

Άσκηση 3 (Κανονικές και Μη Κανονικές Γλώσσες). Είναι κανονικές οι παρακάτω γλώσσες; Αν μια γλώσσα δεν είναι κανονική, να το αποδείξετε χρησιμοποιώντας είτε το Λήμμα Άντλησης είτε κάποια ιδιότητα κλειστότητας. Αν μια γλώσσα είναι κανονική, να το αιτιολογήσετε κατάλληλα.

1. $L_1 = \{w \in \{0,1\}^* : \text{το πλήθος των 0 στην } w \text{ είναι διπλάσιο από το πλήθος των 1}\}$
2. $L_2 = \{ww : w \in \{0,1\}^*, |w| \leq 10^{100}\}$
3. $L_3 = \{w \in \{0,1\}^* : \text{η } w \text{ δεν είναι παλινδρομική}\}$
4. $L_4 = \{0^n 1^m : n \neq m\}$
5. Η γλώσσα που παράγεται από την γραμματική $G = (V, \Sigma, R, S)$, όπου $V = \{S, A, B, 0, 1\}$, $\Sigma = \{0,1\}$, και το σύνολο κανόνων είναι $R = \{S \to AA \,|\, B, A \to 0A \,|\, A0 \,|\, 1, B \to 0B00 \,|\, 1\}$.

Λύση. (1) Θα δείξουμε ότι η (άπειρη) γλώσσα L_1 δεν είναι κανονική χρησιμοποιώντας το Λήμμα Άντλησης. Υποθέτουμε λοιπόν ότι η L_1 είναι κανονική, και θα καταλήξουμε σε άτοπο, παρουσιάζοντας μια αρκετά μεγάλη συμβολοσειρά της L_1 για την οποία το Λήμμα Άντλησης δεν ισχύει.

Έστω $k \geq 1$ το pumping length για τη γλώσσα L_1. Θεωρούμε τη συμβολοσειρά $w = 0^{2k}1^k \in L_1$. Σύμφωνα με το Λήμμα Άντλησης, πρέπει να υπάρχουν συμβολοσειρές x, y, z τέτοιες ώστε (i) $w = xyz$, (ii) $y \neq \varepsilon$, (iii) $|xy| \leq k$, και (iv) για κάθε $n \in \mathbb{N}$, $xy^n z \in L_1$. Λόγω των (i), (ii), και (iii), πρέπει $x = 0^{\ell_1}$, για κάποιο $\ell_1 \in \mathbb{N}$, $y = 0^{\ell_2}$, για κάποιο $\ell_2 \in \mathbb{N}$, με $\ell_2 > 0$ και $\ell_1 + \ell_2 \leq k$, και $z = 0^{2k-(\ell_1+\ell_2)}1^k$. Τότε όμως δεν ισχύει το (iv), αφού $xy^0z = 0^{2k-\ell_2}1^k \notin L_1$, επειδή το πλήθος των 0 είναι μικρότερο από το διπλάσιο του πλήθους των 1. Συνεπώς η L_1 δεν είναι κανονική.

(2) Η γλώσσα L_2 είναι κανονική γιατί είναι πεπερασμένη. Γνωρίζουμε ότι κάθε πεπερασμένη γλώσσα είναι κανονική.

(3) Το συμπλήρωμα $\overline{L_3}$ της L_3 είναι η γλώσσα που περιλαμβάνει όλες τις παλινδρομικές δυαδικές συμβολοσειρές. Γνωρίζουμε ότι η γλώσσα $\overline{L_3}$ δεν είναι κανονική. Επειδή οι κανονικές γλώσσες είναι κλειστές ως προς το συμπλήρωμα, ούτε η L_3 είναι κανονική.

(4) Παρατηρούμε ότι $\overline{L_4} \cap 0^*1^* = \{0^n1^n : n \geq 0\}$ (μια συμβολοσειρά που ανήκει στην 0^*1^* και δεν ανήκει στην L_4 πρέπει να αρχίζει με κάποια 0 ακολουθούμενα από το ίδιο πλήθος 1). Επειδή οι κανονικές γλώσσες είναι κλειστές ως προς την τομή και η γλώσσα $\{0^n1^n : n \geq 0\}$ δεν είναι κανονική, η γλώσσα $\overline{L_4}$ δεν είναι κανονική. Συνεπώς, επειδή οι κανονικές γλώσσες είναι κλειστές ως προς το συμπλήρωμα, ούτε η L_4 είναι κανονική.

(5) Παρατηρούμε ότι η γλώσσα $L(G)$ που παράγεται από την γραμματική G είναι η ένωση δύο (ξένων μεταξύ τους) γλωσσών: της L_1 που αναπαρίσταται από την κανονική έκφραση $0^*10^*10^*$ (προκύπτει από τους κανόνες $S \to AA$ και $A \to 0A \,|\, A0 \,|\, 1$, δείτε ότι ο δεύτερος κανόνας παράγει την γλώσσα 0^*10^*), και της $L_2 = \{0^n10^{2n} : n \geq 0\}$ που προκύπτει από τους κανόνες $S \to B$ και $B \to 0B00 \,|\, 1$. Θα δείξουμε ότι η (άπειρη) γλώσσα $L(G) = L_1 \cup L_2$ δεν είναι κανονική χρησιμοποιώντας το Λήμμα Άντλησης. Υποθέτουμε λοιπόν ότι η $L(G)$ είναι κανονική, και θα καταλήξουμε σε άτοπο.

Έστω $k \geq 1$ το pumping length για τη γλώσσα $L(G)$. Θεωρούμε τη συμβολοσειρά $w = 0^k10^{2k} \in L(G)$. Σύμφωνα με το Λήμμα Άντλησης, πρέπει να υπάρχουν συμβολοσειρές x, y, z τέτοιες ώστε (i) $w = xyz$, (ii) $y \neq \varepsilon$, (iii) $|xy| \leq k$, και (iv) για κάθε $n \in \mathbb{N}$, $xy^nz \in L(G)$. Λόγω των (i), (ii), και (iii), πρέπει $x = 0^{\ell_1}$, για κάποιο $\ell_1 \in \mathbb{N}$, $y = 0^{\ell_2}$, για κάποιο $\ell_2 \in \mathbb{N}$, με $\ell_2 > 0$ και $\ell_1 + \ell_2 \leq k$, και $z = 0^{k-(\ell_1+\ell_2)}10^{2k}$. Θεωρούμε την $xy^0z = 0^{k-\ell_2}10^{2k}$. Παρατηρούμε ότι $0^{k-\ell_2}10^{2k} \notin L_1$, γιατί περιέχει μόνο ένα 1, και ότι $0^{k-\ell_2}10^{2k} \notin L_2$, γιατί το πλήθος των μηδενικών πριν το 1 είναι μικρότερο από το ήμισυ του πλήθους των μηδενικών μετά το 1. Άρα $0^{k-\ell_2}10^{2k} \notin L(G)$. Συνεπώς η $L(G)$ δεν είναι κανονική. $\square$

Άσκηση 4 (Γραμματικές). Να διατυπώσετε γραμματικές χωρίς συμφραζόμενα που παράγουν τις παρακάτω γλώσσες. Αν μια γλώσσα είναι κανονική, η αντίστοιχη γραμματική πρέπει να είναι κανονική.

1. $L_1 = \{a^mb^nc^p : p \geq m+n, m, n \geq 0\}$
2. $L_2 = \{w \in \{0,1\}^* :$ η w έχει περιττό μήκος και το μεσαίο της σύμβολο είναι $0\}$
3. $L_3 = \{w \in \{0,1\}^* :$ το πλήθος των 1 στην w είναι διαφορετικό από το πλήθος των $0\}$
4. $L_4 = \{w \in \{0,1\}^* :$ η w δεν περιέχει τρία συνεχόμενα $1\}$

Λύση. (1) Η L_1 είναι γλώσσα χωρίς συμφραζόμενα (αλλά όχι κανονική). Παράγεται από μια γραμματική χωρίς συμφραζόμενα $G_1(V_1, T_1, S, P_1)$ με αλφάβητο $V_1 = \{S, B, C, a, b, c\}$, σύνολο τερματικών συμβόλων $T_1 = \{a, b, c\}$, και το παρακάτω σύνολο κανόνων παραγωγής:

Κανόνες Παραγωγής P_1
$S \to aSc \mid B$
$B \to bBc \mid C$
$C \to Cc \mid \varepsilon$

(2) Η L_2 είναι γλώσσα χωρίς συμφραζόμενα (αλλά όχι κανονική, επιβεβαιώστε το). Παράγεται από μια γραμματική χωρίς συμφραζόμενα $G_2(V_2, T, S, P_2)$ με αλφάβητο $V_2 = \{S, 0, 1\}$, σύνολο τερματικών συμβόλων $T = \{0, 1\}$, και το παρακάτω σύνολο κανόνων παραγωγής:

Κανόνες Παραγωγής P_2
$S \to 0S0 \mid 0S1 \mid 1S0 \mid 1S1 \mid 0$

(3) Η L_3 είναι γλώσσα χωρίς συμφραζόμενα (αλλά όχι κανονική). Παράγεται από μια γραμματική χωρίς συμφραζόμενα $G_3(V_3, T, S, P_3)$ με αλφάβητο $V_3 = \{S, S_{0+}, S_0, A_0, B_0, S_{1+}, S_1, A_1, B_1, 0, 1\}$, και σύνολο τερματικών συμβόλων $T = \{0, 1\}$.

Μπορούμε να καταλήξουμε στο P_3 αν τροποποιήσουμε κατάλληλα το σύνολο κανόνων παραγωγής $P_=$, το οποίο παράγει όλες τις συμβολοσειρές με το ίδιο πλήθος 0 και 1 (και μόνον αυτές):

Κανόνες Παραγωγής $P_=$
$S \to AB \mid BA \mid \varepsilon$
$A \to 0 \mid AS \mid SA$
$B \to 1 \mid BS \mid SB$

Η ιδέα είναι να προσθέσουμε ένα νέο αρχικό σύμβολο με αποστολή την παραγωγή κάποιων επιπλέον 0 (τουλάχιστον ενός). Τα επιπλέον 0 θα πλαισιώνονται από υποσυμβολοσειρές με το ίδιο πλήθος 0 και 1. Έτσι δημιουργούμε το σύνολο κανόνων παραγωγής P_{0+} που παράγει όλες τις συμβολοσειρές όπου το πλήθος των 0 είναι μεγαλύτερο από το πλήθος των 1 (και μόνον αυτές):

Κανόνες Παραγωγής P_{0+}
$S_{0+} \to S_0 0 S_0$
$S_0 \to A_0B_0 \mid B_0A_0 \mid S_{0+} \mid \varepsilon$
$A_0 \to 0 \mid A_0S_0 \mid S_0A_0$
$B_0 \to 1 \mid B_0S_0 \mid S_0B_0$

Δεν είναι δύσκολο να διαπιστώσουμε ότι το σύνολο κανόνων παραγωγής P_{0+} παράγει μόνο συμβολοσειρές όπου το πλήθος των 0 είναι μεγαλύτερο από το πλήθος των 1. Χρειάζεται ένα επαγωγικό επιχείρημα για να επιβεβαιώσουμε ότι το P_{0+} παράγει όλες τις συμβολοσειρές όπου το πλήθος των 0 είναι μεγαλύτερο από το πλήθος των 1.

Αντίστοιχα δημιουργούμε το σύνολο κανόνων παραγωγής P_{1+} που παράγει όλες τις συμβολοσειρές όπου το πλήθος των 1 είναι μεγαλύτερο από το πλήθος των 0 (και μόνον αυτές). Το ζητούμενο σύνολο κανόνων παραγωγής P_3 προκύπτει από την ένωση των P_{0+} και P_{1+}.

Κανόνες Παραγωγής P_3
$S \to S_{0+} \mid S_{1+}$
$S_{0+} \to S_0 0 S_0$
$S_0 \to A_0B_0 \mid B_0A_0 \mid S_{0+} \mid \varepsilon$
$A_0 \to 0 \mid A_0S_0 \mid S_0A_0$
$B_0 \to 1 \mid B_0S_0 \mid S_0B_0$
$S_{1+} \to S_1 1 S_1$
$S_1 \to A_1B_1 \mid B_1A_1 \mid S_{1+} \mid \varepsilon$
$A_1 \to 0 \mid A_1S_1 \mid S_1A_1$
$B_1 \to 1 \mid B_1S_1 \mid S_1B_1$

(4) Η L_4 είναι κανονική γλώσσα και παράγεται από μια κανονική γραμματική $G_4(V_4, T, S, P_4)$ με αλφάβητο $V = \{S, A_1, A_2, 0, 1\}$, σύνολο τερματικών συμβόλων $T = \{0, 1\}$, και το παρακάτω σύνολο κανόνων παραγωγής:

Κανόνες Παραγωγής P_4
$S \;\; \to 0S \mid 1A_1 \mid \varepsilon$
$A_1 \to 0S \mid 1A_2 \mid \varepsilon$
$A_2 \to 0S \mid \varepsilon$

Άσκηση 5. Ποιες από τις παρακάτω γλώσσες είναι χωρίς συμφραζόμενα; Αν μια γλώσσα δεν είναι χωρίς συμφραζόμενα, να το αποδείξετε χρησιμοποιώντας είτε το Λήμμα Άντλησης είτε κάποια ιδιότητα κλειστότητας. Αν μια γλώσσα είναι χωρίς συμφραζόμενα, να το αιτιολογήσετε κατάλληλα.

1. $L_1 = \{0^n 1^m : n \neq m\}$
2. $L_2 = \{0^n 1^n 0^n 1^n : n \geq 0\}$
3. $L_3 = \{auabwbcvc : u, w, v \in \{a, b, c\}^*, |u| = |w| = |v|\}$
4. $L_4 = \{w \in \{0, 1\}^*$: η w είναι παλινδρομική και περιέχει το ίδιο πλήθος από 0 και 1 $\}$

Λύση. (1) Έχουμε αποδείξει ότι η L_1 δεν είναι κανονική. Θα δείξουμε ότι η L_1 είναι γλώσσα χωρίς συμφραζόμενα διατυπώνοντας μια γραμματική $G_1(V_1, T, S, P_1)$ χωρίς συμφραζόμενα που την παράγει. Το αλφάβητο της G_1 είναι $V_1 = \{S, S_0, A, S_1, B, 0, 1\}$, και το σύνολο τερματικών συμβόλων είναι $T = \{0, 1\}$. Για τους κανόνες παραγωγής, εργαζόμαστε όπως στο Θέμα 4.3, δηλ. διατυπώνουμε και συνδυάζουμε δύο σύνολα κανόνων παραγωγής, ένα (με αρχικό σύμβολο S_0) που παράγει την γλώσσα $\{0^n 1^m : n > m\}$, και ένα (με αρχικό σύμβολο S_1) που παράγει την γλώσσα $\{0^n 1^m : n < m\}$.

Κανόνες Παραγωγής P_1
$S \;\; \to S_0 \mid S_1$
$S_0 \to 0S_0 1 \mid A$
$A \;\; \to 0S_0 \mid 0$
$S_1 \to 0S_1 1 \mid B$
$B \;\; \to S_1 1 \mid 1$

(2) Θα δείξουμε ότι η (άπειρη) γλώσσα L_2 δεν είναι γλώσσα χωρίς συμφραζόμενα χρησιμοποιώντας το αντίστοιχο Λήμμα Άντλησης. Υποθέτουμε λοιπόν ότι η L_2 είναι γλώσσα χωρίς συμφραζόμενα, και θα καταλήξουμε σε άτοπο.

Έστω $k \geq 1$ το pumping length για την γλώσσα L_2. Θεωρούμε τη συμβολοσειρά $w = 0^k 1^k 0^k 1^k \in L_2$. Σύμφωνα με το Λήμμα Άντλησης, πρέπει να υπάρχουν συμβολοσειρές u, v, x, y, z τέτοιες ώστε (i) $w = uvxyz$, (ii) $|vy| > 0$, (iii) $|vxy| \leq k$, και (iv) για κάθε $n \in \mathbb{N}$, $uv^n xy^n z \in L_2$. Διακρίνουμε τις παρακάτω περιπτώσεις:

– Αν η v είναι μη κενή και αποτελείται από ένα μόνο σύμβολο, π.χ. η v αποτελείται μόνο από 0, τότε $uv^0 xy^0 z \notin L_2$ γιατί το πλήθος των 0 στο ένα τμήμα της $uv^0 xy^0 z$ είναι διαφορετικό από το πλήθος των 0 στο άλλο τμήμα της (π.χ. αν η v αποτελεί υποσυμβολοσειρά του πρώτου 0^k, η vxy αποτελεί υποσυμβολοσειρά του πρώτου μισού της w επειδή $|vxy| \leq k$, οπότε το πρώτο μισό της $uv^0 xy^0 z$ έχει λιγότερα μηδενικά από το δεύτερο μισό της). Αντίστοιχη είναι η περίπτωση όπου η v είναι μη κενή και αποτελείται μόνο από 1.
– Χειριζόμαστε αντίστοιχα την περίπτωση όπου η y είναι μη κενή και αποτελείται από ένα μόνο σύμβολο.
– Διαφορετικά, τουλάχιστον μία από τις v, y είναι μη κενή και αποτελείται τόσο από 0 όσο και από 1. Τότε $uv^2 xy^2 z \notin L_2$ γιατί στην $uv^2 xy^2 z$ υπάρχουν περισσότερα από δύο σημεία όπου κάποιο 0 ακολουθείται από κάποιο 1.

Συνεπώς η L_2 δεν είναι γλώσσα χωρίς συμφραζόμενα.

(3) Παρατηρούμε ότι $L_3 \cap a^*b^*c^* = \{a^n b^n c^n : n \geq 2\}$. Επειδή η τομή μια γλώσσας χωρίς συμφραζόμενα με μια κανονική γλώσσα είναι γλώσσα χωρίς συμφραζόμενα, και επειδή η $\{a^n b^n c^n : n \geq 2\}$ δεν είναι γλώσσα χωρίς συμφραζόμενα, συμπεραίνουμε ότι η L_3 δεν είναι γλώσσα χωρίς συμφραζόμενα.

(4) Θα δείξουμε ότι η (άπειρη) γλώσσα L_4 δεν είναι γλώσσα χωρίς συμφραζόμενα χρησιμοποιώντας το αντίστοιχο Λήμμα Άντλησης. Υποθέτουμε λοιπόν ότι η L_4 είναι γλώσσα χωρίς συμφραζόμενα, και θα καταλήξουμε σε άτοπο.

Έστω $k \geq 1$ το pumping length για την γλώσσα L_4. Θεωρούμε τη συμβολοσειρά $w = 0^k 1^k 1^k 0^k \in L_4$. Σύμφωνα με το Λήμμα Άντλησης, πρέπει να υπάρχουν συμβολοσειρές u, v, x, y, z τέτοιες ώστε (i) $w = uvxyz$, (ii) $|vy| > 0$, (iii) $|vxy| \leq k$, και (iv) για κάθε $n \in \mathbb{N}$, $uv^n xy^n z \in L_4$. Διακρίνουμε τις παρακάτω περιπτώσεις:

- Αν η vy αποτελείται μόνο από 0 (ή αντίστοιχα μόνο από 1), τότε $uv^2xy^2z \notin L_4$ γιατί περιέχει διαφορετικό πλήθος από 0 και 1.
- Αν η v είναι μη κενή και αποτελείται μόνο από 0 και η y είναι μη κενή και αποτελείται μόνο από 1 (ή αντίστροφα), τότε $uv^2xy^2z \notin L_4$ γιατί δεν είναι παλινδρομική.
- Αν μια μόνο από τις v, y αποτελείται τόσο από 0 όσο και από 1 (η άλλη μπορεί να είναι κενή ή να αποτελείται από ένα μόνο σύμβολο), τότε $uv^2xy^2z \notin L_4$ γιατί δεν είναι παλινδρομική.
- Τέλος παρατηρούμε ότι επειδή $|vxy| \leq k$, δεν είναι δυνατόν αμφότερες οι v και y να είναι μη κενές και να αποτελούνται τόσο από 0 όσο και από 1.

Συνεπώς η L_4 δεν είναι γλώσσα χωρίς συμφραζόμενα. $\square$