ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

Θεωρία Υπολογισμού

Σημειώσεις - Γραμματική Ανεξάρτητη Συμφραζόμενων

Αλέξιος Καπόρης

Τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Αιγαίου**» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Θεωρήστε τη γλώσσα L, της οποίας οι συμβολοσειρές προκύπτουν από τους παρακάτω κανόνες:

1. $\epsilon \in L$.
2. Αν $S \in L$ τότε και $0S1 \in L$.
3. Τίποτα άλλο δεν ανήκει στην L εκτός αν προκύπτει από τους παραπάνω κανόνες.

Δεν είναι δύσκολο να δει κάποιος ότι $L = \{0^n 1^n |\ n \geq 0\}$, η οποία όπως γνωρίζουμε δεν είναι κανονική. Αυτό γιατί αν ξεκινήσουμε από την μικρότερη συμβολοσειρά που ανήκει στην γλώσσα (την κενή) και, εφαρμόζοντας τον δεύτερο κανόνα, αρχίσουμε να "κολλάμε" δεξιά και αριστερά αυτής άσσους και μηδενικά, θα πάρουμε συμβολοσειρές του τύπου $0^n 1^n$.

Αν τώρα θεωρήσουμε το σύμβολο S σαν μια μεταβλητή που αναπαριστά οποιαδήποτε συμβολοσειρά της L, τότε οι παραπάνω κανόνες μπορούν να πάρουν τη μορφή

1. $S \rightarrow \epsilon$
2. $S \rightarrow 0S1$

όπου το σύμβολο $\rightarrow$ σημαίνει "μπορεί να πάρει τη τιμή".

Μια *γραμματική ανεξάρτητη συμφραζόμενων (ΑΣ)* είναι μια τετράδα $G = (V, \Sigma, S, R)$, όπου V είναι ένα πεπερασμένο σύνολο *μεταβλητών*, Σ είναι το σύνολο των *τερματικών συμβόλων* με $V \cap \Sigma = \emptyset$, $S \in V$ είναι η *αρχική* μεταβλητή και R είναι ένα πεπερασμένο σύνολο *κανόνων* της μορφής $A \rightarrow \alpha$, όπου $A \in V$ και $\alpha \in (V \cup \Sigma)^*$.

Θεωρήστε τη γλώσσα P όλων των παλινδρομικών συμβολοσειρών στο $\{0,1\}^*$. Γνωρίζουμε ότι οι $\epsilon, 0, 1$ είναι παλινδρομικές συμβολοσειρές και ότι, αν η x ανήκει στην P, τότε και οι $0x0$, $1x1$ ανήκουν στην P. Οι συμβολοσειρές της P μπορούν να παραχθούν από τη γραμματική $G = (\{S\}, \{0,1\}, S, R)$, όπου το σύνολο των κανόνων R είναι το

$$S \to \epsilon \mid 0 \mid 1$$
$$S \to 0S0 \mid 1S1$$

Η γραμματική αποτελείται από μια μόνο μεταβλητή, την S, η οποία είναι και αρχική. Το σύνολο R αποτελείται από τους πέντε κανόνες που φαίνονται παραπάνω. Το σύμβολο "|" σ' έναν κανόνα μεταφράζεται "ή". Έτσι ο κανόνας $S \to 0S0 \mid 1S1$ αποτελεί συντομογραφία των $S \to 0S0$ ή $S \to 1S1$. Μια τυπική παραγωγή της G φαίνεται στη συνέχεια:

$$S \Rightarrow 0S0 \Rightarrow 01S10 \Rightarrow 011S110 \Rightarrow 0110110$$

Το πρώτο βήμα χρησιμοποιεί τον κανόνα $S \to 0S0$, το δεύτερο και τρίτο τον κανόνα $S \to 1S1$ και το τελευταίο το $S \to 0$.

Δώστε μια γραμματική ΑΣ για τη γλώσσα όλων των συμβολοσειρών που *δεν είναι* παλινδρομικές.

Άσκηση 7.1 Θέλουμε να δώσουμε μια γραμματική ΑΣ για τη γλώσσα όλων των *μη* παλινδρομικών συμβολοσειρών. Στις ασκήσεις αυτού του είδους προσπαθούμε πρώτα να βρούμε κάποια ιδιότητα που χαρακτηρίζει τις συμβολοσειρές της γλώσσας. Έτσι, μια συμβολοσειρά x δε θα είναι παλινδρομική, αν καθώς συγκρίνουμε σύμβολα από τα δύο άκρα, κάποια στιγμή δύο απ' αυτά είναι διαφορετικά. Αρχικά λοιπόν θα έχουμε τους κανόνες

$$S \to 0S0 \mid 1S1 \mid A$$

οι οποίοι, με τη μεταβλητή S, παράγουν ένα τμήμα που είναι παλινδρομικό και με την A, μια συμβολοσειρά στην οποία το πρώτο και τελευταίο σύμβολο είναι διαφορετικά. Άρα μπορούμε να δοκιμάσουμε τους κανόνες

$$A \to 0B1 \mid 1B0$$

όπου η μεταβλητή B χρησιμεύει για να παράγει οτιδήποτε, όπως φαίνεται παρακάτω

$$B \to 0B \mid 1B \mid \epsilon$$

Από την ανάλυση φαίνεται ότι αυτή η γραμματική είναι η ζητούμενη λύση. Γενικά όμως θα πρέπει να δείχνουμε ότι κάθε συμβολοσειρά της γλώσσας μπορεί να παραχθεί από τη γραμματική και αντίστροφα, ότι η γραμματική παράγει μόνο τις ζητούμενες συμβολοσειρές. Για παράδειγμα, η συμβολοσειρά 011100 έχει την ακόλουθη παραγωγή:

$$S \Rightarrow 0S0 \Rightarrow 0A0 \Rightarrow 01A00 \Rightarrow 01B00 \Rightarrow 011B00 \Rightarrow 0111B00 \Rightarrow 011100$$

Ας προσπαθήσουμε να βρούμε μια γραμματική για τη γλώσσα $L = \{x \in \{0,1\}^* \mid$ η x περιέχει περισσότερα 0 από 1$\}$.
Όπως και στο Παράδειγμα 7.1 θα προσπαθήσουμε να ορίσουμε την L αναδρομικά. Παρατηρήστε πρώτα ότι $0 \in L$. Αν τώρα x είναι μια συμβολοσειρά της L, τότε και οι $0x$, $x0$ θα ανήκουν στην L. Από αυτές τις παρατηρήσεις προκύπτουν οι κανόνες

$$S \to 0 \mid 0S \mid S0$$

Θα πρέπει τώρα να επιτρέψουμε και την παρουσία άσσων, αλλιώς η γραμματική θα αποτελείται μόνο από μηδενικά. Για να το κάνουμε αυτό θα χρησιμοποιήσουμε δύο συμβολοσειρές $x, y \in L$. Αν συνδυάσουμε αυτές τις δύο τότε θα έχουμε *τουλάχιστον* δύο περισσότερα 0 από 1, γιατί κάθε μια από τις x, y περιέχει τουλάχιστον ένα παραπάνω 0 από 1. Άρα μπορούμε να προσθέσουμε έναν άσσο και η συμβολοσειρά που προκύπτει να ανήκει πάλι στην L. Αυτός ο άσσος μπορεί να προστεθεί στα αριστερά, ανάμεσα ή στα δεξιά των x και y. Οι αντίστοιχοι κανόνες είναι

$$S \to 1SS \mid S1S \mid SS1$$

Άρα η ζητούμενη γραμματική G θα αποτελείται από τους παρακάτω κανόνες

$$S \to 0 \mid 0S \mid S0 \mid 1SS \mid S1S \mid SS1$$

δ(176 66).151

Ασκηση Αυτοαξιολόγησης 7.3

Δείξτε ότι η γραμματική με κανόνες

$$S \to AA$$
$$A \to AAA \mid bA \mid Ab \mid a$$

παράγει τη γλώσσα όλων των συμβολοσειρών $x \in \{a, b\}^*$, στις οποίες ο αριθμός των a είναι ζυγός και μεγαλύτερος του μηδενός.

σελ.161

Θεωρήστε ξανά τη γλώσσα $L_1 = \{x \in \{0,1\}^* |$ η x περιέχει περισσότερα 0 από 1$\}$ του Παραδείγματος 7.4. Μετονομάζοντας την αρχική μεταβλητή S σε A, η γραμματική G_1 δίνεται από τους παρακάτω κανόνες:

$$A \to 0 \mid 0A \mid A0 \mid 1AA \mid A1A \mid AA1$$

Αντίστοιχα, αν L_2 είναι η γλώσσα $\{x \in \{0,1\}^* |$ η x περιέχει περισσότερα 1 από 0$\}$, τότε μια γραμματική G_2 προκύπτει αντιστρέφοντας τους ρόλους των 0 και 1. Θέτοντας B ως την αρχική μεταβλητή, παίρνουμε τους κανόνες

$$B \to 1 \mid 1B \mid B1 \mid 0BB \mid B0B \mid BB0$$

Υποθέστε τώρα ότι μας ζητείται να δώσουμε μια γραμματική για τη γλώσσα L όλων των συμβολοσειρών $x \in \{0,1\}^*$ στις οποίες ο αριθμός των 0 *διαφέρει* από τον αριθμό των 1. Είναι φανερό ότι η L μπορεί να εκφραστεί ως η *ένωση* των L_1, L_2. Εισάγοντας λοιπόν μια νέα αρχική μεταβλητή S και τον κανόνα $S \to A \mid B$, μαζί με τους κανόνες των G_1 και G_2, η νέα γραμματική εκφράζει την ένωση των L_1, L_2. Η ίδια τεχνική μπορεί να εφαρμοστεί για να δείξουμε ότι η

Έστω $L = \{a^i b^j c^k \mid j = i + k\}$. Θα προσπαθήσουμε να εφαρμόσουμε τις ιδέες του Θεωρήματος 7.2 για να εκφράσουμε την L ως τη συνένωση απλούστερων γλωσσών. Μια πρώτη προσέγγιση είναι να θέσουμε $L = L_1 L_2 L_3$, όπου οι τρεις γλώσσες περιέχουν συμβολοσειρές από a, b και c, αντίστοιχα. Κάτι τέτοιο όμως οδηγεί σε λάθος. Η L_1 μπορεί να περιέχει το aa, η L_2 το b και η L_3 το c, όμως η L δεν περιέχει το $aabc$, αφού ο αριθμός των b πρέπει να είναι ίσος με το άθροισμα των a και c.

Υπάρχει όμως μια σχετικά εύκολη λύση. Η συμβολοσειρά $a^i b^j c^k$, όπου $j = i + k$ μπορεί να γραφεί

$$a^i b^{i+k} c^k = a^i b^i b^k c^k$$

Άρα η L είναι η συνένωση των L_1, L_2, όπου

$$L_1 = \{a^i b^i \mid i \geq 0\}$$
$$L_2 = \{b^k c^k \mid k \geq 0\}$$

Όμως έχουμε ήδη γραμματικές γι' αυτές τις γλώσσες. Η L_1 παράγεται με τους κανόνες

$$A \to aAb \mid \epsilon$$

και η L_2 με τους κανόνες

$$C \to bCc \mid \epsilon$$

Εισάγοντας μια νέα μεταβλητή S και συνενώνοντας τις αρχικές μεταβλητές των L_1, L_2 παίρνουμε τους κανόνες για την $L = L_1 L_2$:

$$S \to AC$$

$$A \to aAb \mid \epsilon$$
$$C \to bCc \mid \epsilon$$

Για παράδειγμα η συμβολοσειρά $abbbcc = (ab)(b^2c^2)$ μπορεί να παραχθεί ως εξής:

$$S \Rightarrow AC \Rightarrow aAbC \Rightarrow abC \Rightarrow abbCc \Rightarrow abbbCcc \Rightarrow abbbcc$$

Δίνεται η γλώσσα $L=\{a^n (bc)^{3n} \mid n \geq 0\}$, πχ η λέξη abcbcbc ανήκει στη γλώσσα.

i)Δώστε Γραμματική Ανεξάρτητη Συμφραζομένων που να παράγει την L.

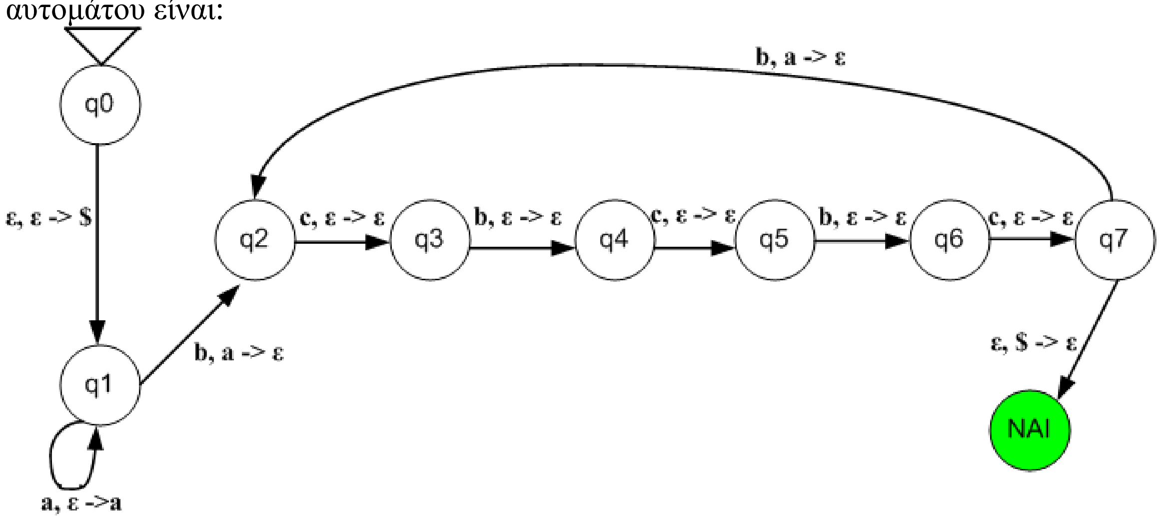
ii)Δώστε αυτόματο στοίβας για την L.

Λύση

i) Η γραμματική ΑΣ είναι: $S \to \varepsilon \mid aS(bc)^3$.

ii) Διαισθητικά, το αυτόματο αρχικά θα αποθηκεύει στην στοίβα του τα διαδοχικά a's που θα διαβάζει στο αρχικό τμήμα της λέξης στην εισόδου του. Μόλις δει το πρώτο b,

για κάθε 1 a που θα σβήνει από τη στοίβα, θα προσπερνά από την ταινία εισόδου 3 διαδοχικά bc'ς χωρίς να τροποποιεί την στοίβα. Θα τερματίζει σε ΝΑΙ μόνο όταν αδειάσει η στοίβα και στη ταινία εισόδου βρει άδειο κελί τέρμα δεξιά της εισόδου.
Το κάτω γράφημα (σύμφωνα με το συμβολισμό του M. Sipser στο βιβλίο σας) του αυτομάτου είναι:



Περιφραστικά οι άνω καταστάσεις του αυτομάτου στοίβας έχουν ως εξής:
Μετάβαση από q0 σε q1 = [σηματοδοτεί το πάτο της στοίβας με το «$»].
Μετάβαση από q1 σε q1 = [γράφει τα a'ς της εισόδου στη στοίβα].
Μετάβαση από q1 σε q2 = [διαβάζει το 1^{ο} b στην είσοδο και σβήνει 1 a από στοίβα].
Μετάβαση από q2 σε q7 = [διαβάζει το cbcbc στην είσοδο και δεν αλλοιώνει τη στοίβα].
Μετάβαση άνω από q7 σε q2 = [διαβάζει b στην είσοδο και σβήνει 1 a από στοίβα].
Μετάβαση από q7 σε ΝΑΙ = [τελειώνει η είσοδος όταν η στοίβα άδειασε και λέει ΝΑΙ].