Πανεπιστήμιο Αιγαίου

Θεωρία Υπολογισμού

4η Διάλεξη

Αυτόματα Στοίβας

Γλώσσες Ανεξάρτητες Συμφραζομένων (ΓΑΣ)

Λ. Άντλησης για γλώσσες μη ΑΣ

Αλέξιος Καπόρης

Άδειες Χρήσης

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 πό

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

ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

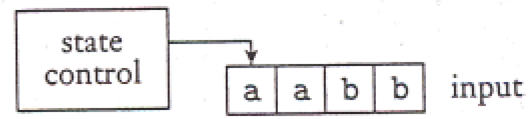


FIGURE 2.4
Schematic of a finite automaton

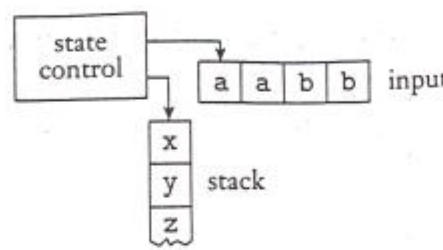


FIGURE 2.5
Schematic of a pushdown automaton

The following is the formal description of the PDA from page 102 that recognizes the language $\{0^n1^n \mid n \geq 0\}$. Let M_1 be $(Q, \Sigma, \Gamma, \delta, q_1, F)$, where

$Q = \{q_1, q_2, q_3, q_4\}$,

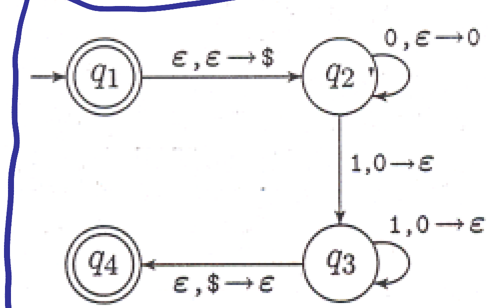
$\Sigma = \{0,1\}$,

$\Gamma = \{0, \$\}$,

$F = \{q_1, q_4\}$, and

δ is given by the following table, wherein blank entries signify $\emptyset$.

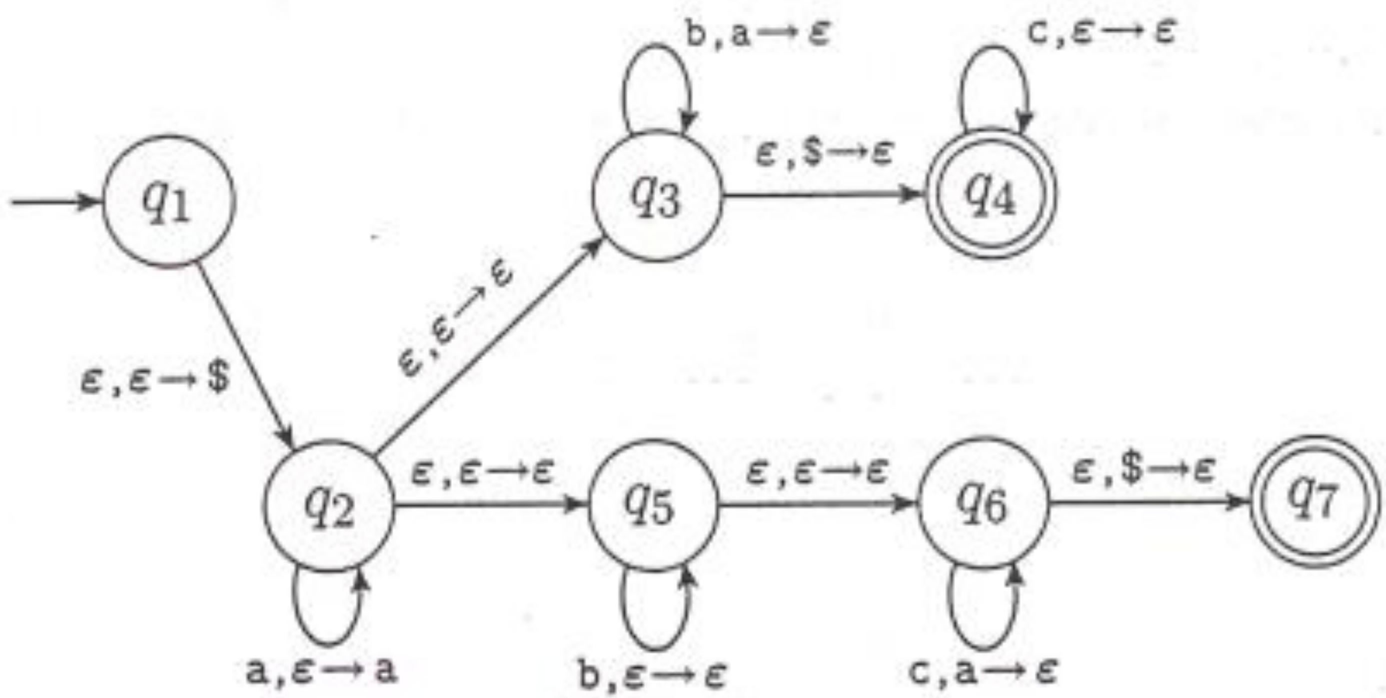
Input:	0			1			ε		
Stack:	0	\$	ε	0	\$	ε	0	\$	ε
q_1									$\{(q_2, \$)\}$
q_2			$\{(q_2, 0)\}$	$\{(q_3, \varepsilon)\}$					
q_3				$\{(q_3, \varepsilon)\}$				$\{(q_4, \varepsilon)\}$	
q_4									



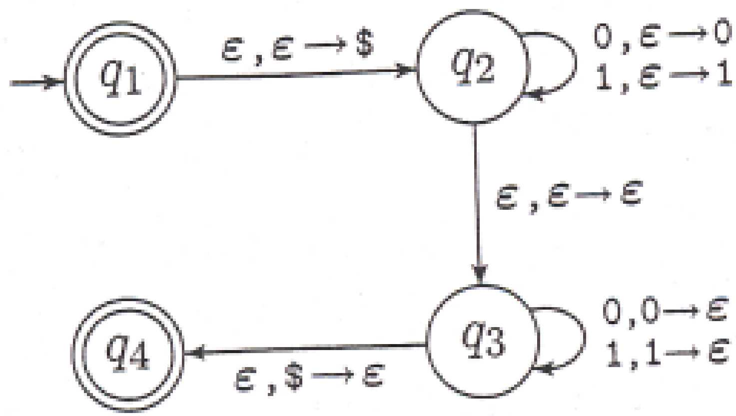
state to state. We write $\boxed{\text{“}a,b \rightarrow c\text{”}}$ to signify that when the machine is reading an a from the input it may replace the symbol b on the top of the stack with a c. Any of a, b, and c may be ε. If a is ε, the machine may make this transition without reading any symbol from the input. If b is ε, the machine may make this transition without reading and popping any symbol from the stack. If c is ε, the machine does not write any symbol on the stack when going along this transition.

This example illustrates a pushdown automaton that recognizes the language

$$\{a^i b^j c^k \mid i, j, k \geq 0 \text{ and } i = j \text{ or } i = k\}.$$



In this example we give a PDA M_3 recognizing the language $\{ww^{\mathcal{R}} | w \in \{0,1\}^*\}$. Recall that $w^{\mathcal{R}}$ means w written backwards. The informal description of the PDA follows.



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 $A \to \alpha$, 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 $S \to 0$.

Έστω $G = (V, \Sigma, S, R)$ η γραμματική όπου $V = \{S\}$, $\Sigma = \{+, -, /, *, (,), x, y\}$ και R αποτελείται από τους κανόνες

$$S \to S + S \mid S - S \mid S/S \mid S * S \mid (S) \mid x \mid y$$

Η γραμματική αυτή παράγει τη γλώσσα όλων των *σωστών* αλγεβρικών εκφράσεων. Ο αναδρομικός ορισμός που δίνεται από τους παραπάνω κανόνες δηλώνει ότι από δύο σωστές εκφράσεις, μια νέα έκφραση μπορεί να προκύψει είτε ενώνοντας τις δύο με ένα από τα σύμβολα $+$, $-$, $/$, $*$, ή τοποθετώντας την αρχική μέσα σε παρενθέσεις. Τα σύμβολα x, y παίζουν το ρόλο των μεταβλητών που μετέχουν σ' αυτές τις αλγεβρικές εκφράσεις. Δύο τυπικές παραγωγές για την έκφραση $x + x * y$ φαίνονται παρακάτω:

$$S \Rightarrow S + S \Rightarrow x + S \Rightarrow x + S * S \Rightarrow x + x * S \Rightarrow x + x * y$$

$$S \Rightarrow S * S \Rightarrow S + S * S \Rightarrow x + S * S \Rightarrow x + x * S \Rightarrow x + x * y$$

Η πρώτη απ' αυτές τις δύο φαίνεται περισσότερο "φυσική", αφού αρχίζει με την έκφραση $S + S$ δηλώνοντας ότι εκείνο που έχουμε στο νου μας είναι η πρόσθεση δύο εκφράσεων, της x και της $x * y$, αντιστοιχώντας στην έκφραση $x + (x * y)$. Η δεύτερη από την άλλη μεριά ερμηνεύει την έκφραση σαν γινόμενο. Αρχίζει με την παραγωγή $S * S$ και αντιστοιχεί στην έκφραση $(x + x) * y$. Όταν όμως έχουμε μια έκφραση της μορφής $x + x * y$, συνήθως την ερμηνεύουμε σαν $x + (x * y)$ και όχι σαν $(x + x) * y$, επειδή ο πολλαπλασιασμός $*$ έχει μεγαλύτερη προτεραιότητα από την πρόσθεση $+$.

Ας προσπαθήσουμε να βρούμε μια γραμματική για τη γλώσσα $L = \{x \in \{0,1\}^* |$ 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

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

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 ένωση δύο οποιωνδήποτε γλωσσών ΑΣ είναι πάλι ΑΣ. Το θεώρημα που ακολουθεί

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

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

$$A \rightarrow aAb \mid \epsilon \qquad C \rightarrow bCc \mid \epsilon$$

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:

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

Θεώρημα 7.2. *Οι γλώσσες ανεξάρτητες συμφραζόμενων είναι κλειστές ως προς τις πράξεις ένωση, συνένωση και αστέρι Kleene.*

Θεώρημα 7.2. *Οι γλώσσες ανεξάρτητες συμφραζόμενων είναι κλειστές ως προς τις πράξεις ένωση, συνένωση και αστέρι Kleene.*

Η τομή; Π.χ. για κανονικές:

Θεώρημα 7.2. *Οι γλώσσες ανεξάρτητες συμφραζόμενων είναι κλειστές ως προς τις πράξεις ένωση, συνένωση και αστέρι Kleene.*

Η τομή; Π.χ. για κανονικές:

Θεώρημα 4.1. *Αν L_1, L_2 είναι κανονικές γλώσσες πάνω στο Σ^*, τότε και οι $L_1 \cap L_2$, $L_1 - L_2$ είναι κανονικές.*

Θεώρημα 7.2. *Οι γλώσσες ανεξάρτητες συμφραζόμενων είναι κλειστές ως προς τις πράξεις ένωση, συνένωση και αστέρι Kleene.*

Η τομή; Π.χ. για κανονικές:

Θεώρημα 4.1. *Αν L_1, L_2 είναι κανονικές γλώσσες πάνω στο Σ^*, τότε και οι $L_1 \cap L_2$, $L_1 - L_2$ είναι κανονικές.*

Θεώρημα 9.2. *Οι γλώσσες ανεξάρτητες συμφραζόμενων δεν είναι κλειστές ως προς τις πράξεις της τομής και του συμπληρώματος.*

Θεώρημα 9.2. *Οι γλώσσες ανεξάρτητες συμφραζόμενων δεν είναι κλειστές ως προς τις πράξεις της τομής και του συμπληρώματος.*

Απόδειξη: Εκείνο που πρέπει να κάνουμε είναι να βρούμε δύο γλώσσες L_1 και L_2, οι οποίες είναι ΑΣ έτσι ώστε, αν πάρουμε την τομή τους $L = L_1 \cap L_2$, τότε η L δεν είναι ΑΣ.

Εφόσον ήδη γνωρίζουμε ότι η γλώσσα $\{0^n 1^n 2^n | \ n \geq 0\}$ δεν είναι ΑΣ, αρκεί να βρούμε δύο γλώσσες ΑΣ L_1 και L_2 τέτοιες ώστε $L_1 \cap L_2 = \{0^n 1^n 2^n | \ n \geq 0\}$. Αυτό όμως είναι εύκολο. Ας είναι $L_1 = \{0^n 1^n 2^m | \ n, m \geq 0\}$ και $L_2 = \{0^m 1^n 2^n | \ n, m \geq 0\}$, οι οποίες είναι ΑΣ. Είναι προφανές ότι η τομή των δύο γλωσσών αποτελείται από εκείνες τις συμβολοσειρές που έχουν τον ίδιο αριθμό από 0, 1 και 2 άρα δε μπορεί να είναι ΑΣ.