203: Διακριτά Μαθηματικά
Κεφάλαιο 7: Θεωρία Αυτομάτων και Τυπικών Γλωσσών

Σπυρίδων Τζίμας

Εαρινό Εξάμηνο 2025

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ & ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ
ΣΧΟΛΗ ΠΛΗΡΟΦΟΡΙΚΗΣ & ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ

Τυπικές Γλώσσες

Θεωρούμε ένα σύνολο Σ το οποίο καλούμε αλφάβητο. Τα στοιχεία του Σ τα καλούμε σύμβολα και τις ακολουθίες που σχηματίζονται αποκλειστικά από αυτά συμβολοσειρές. Γράφουμε τις συμβολοσειρές απλά παραθέτοντας τα σύμβολα που τις σχηματίζουν όπως γράφουμε μία λέξη μιας φυσικής γλώσσας (natural language).

Τις συμβολοσειρές που ανήκουν στο Σ^n τις καλούμε συμβολοσειρές μήκους n. Συμβολίζουμε με Σ^* το σύνολο των συμβολοσειρών με σύμβολα από το Σ ανεξαρτήτως μήκους, δηλαδή:

$$\Sigma^* = \Sigma^0 + \Sigma^1 + \Sigma^2 + \cdots$$

Οποιοδήποτε σύνολο $\Gamma \subseteq \Sigma^*$ το καλούμε τυπική γλώσσα (formal language).

Κανονικές Γλώσσες και Κανονικές Εκφράσεις

Θεωρούμε τους ακόλουθους τελεστές πάνω στα σύμβολα του αλφαβήτου. Το αποτέλεσμα της εφαρμογής του καθενός από αυτούς είναι μία έκφραση που περιγράφει μία ή περισσότερες συμβολοσειρές ταυτόχρονα.

Έστω σύμβολα a, b.

Συμβολίζουμε με ε την μοναδική συμβολοσειρά μήκους 0 ή αλλιώς την κενή συμβολοσειρά.

Γράφουμε	Ταιριάζουν	
ab	ab	1 αντίτυπο του a ακολουθούμενο από 1 αντίτυπο του b
$a \mid b$	a, b	είτε 1 αντίτυπο του a είτε 1 αντίτυπο του b
a^*	$\varepsilon, a, aa, aaa, \ldots$	οποιαδήποτε ακολουθία 0 ή περισσότερων αντιτύπων του a
a^+	$a, aa, aaa, \ldots$	οποιαδήποτε ακολουθία 1 ή περισσότερων αντιτύπων του a

Τα ορίσματα μπορεί να είναι όχι μόνο απλά σύμβολα αλλά και σύνθετες εκφράσεις.

Κανονικές Γλώσσες και Κανονικές Εκφράσεις

Μία κανονική έκφραση (regular expression (regexp)) είναι

- είτε ε,
- είτε ένα σύμβολο του αλφαβήτου,
- είτε μία έκφραση της μορφής $s_1 s_2$ ή $s_1 | s_2$ ή s_1^* ή s_1^+, όπου s_1, s_2 είναι κανονικές εκφράσεις.

Μέσα σε μία κανονική έκφραση, τηρούμε την ακόλουθη σειρά προτεραιότητας:

άστρο, συν > παράθεση > κάθετος

Μπορούμε να αλλάξουμε την σειρά εφαρμογής των τελεστών με παρενθέσεις.

Μια γλώσσα Γ καλείται κανονική γλώσσα (regular language) αν υπάρχει κάποια κανονική έκφραση που να περιγράφει όλες τις συμβολοσειρές της Γ και μόνο αυτές.

Κανονικές Εκφράσεις

Παράδειγμα: Γράψτε την κανονική έκφραση που περιγράφει τις συμβολοσειρές μήκους 8 με αλφάβητο το $\{0, 1\}$ (1 byte) που αρχίζουν σε 1 και τελειώνουν σε 0.

$$1(0 \mid 1)(0 \mid 1)(0 \mid 1)(0 \mid 1)(0 \mid 1)(0 \mid 1)0$$

Παράδειγμα: Γράψτε την κανονική έκφραση που περιγράφει τις συμβολοσειρές οποιουδήποτε μήκους με αλφάβητο το $\{0, 1\}$ που αρχίζουν σε 1 και τελειώνουν σε 0.

$$1(0 \mid 1)^*0$$

Παράδειγμα: Γράψτε την κανονική έκφραση που περιγράφει τις συμβολοσειρές μήκους τουλάχιστον 3 με αλφάβητο το $\{0, 1\}$ που αρχίζουν σε 1 και τελειώνουν σε 0.

$$1(0 \mid 1)^+0$$

Κανονικές Εκφράσεις

Παράδειγμα: Γράψτε την κανονική έκφραση που περιγράφει τις συμβολοσειρές *bard*, *bird* και *board* με αλφάβητο το λατινικό.

$$b(a \mid i \mid oa)rd$$

Παράδειγμα: Γράψτε την κανονική έκφραση που περιγράφει τις συμβολοσειρές *John*, *Johny* και *Johnny* με αλφάβητο το λατινικό.

$$John(\varepsilon \mid y \mid ny)$$

Παράδειγμα: Γράψτε την κανονική έκφραση που περιγράφει τις συμβολοσειρές *haha*, *hahaha*, *hahahaha*, κ.τ.λ. με αλφάβητο το λατινικό.

$$haha(ha)^* = ha(ha)^+$$

Αυτόματα

Τα αυτόματα είναι θεωρητικές μηχανές προπομποί των σύγχρονων υπολογιστών που

- ✔ δέχονται μια συμβολοσειρά ως είσοδο,
- ✔ εκτελούν κάποιον υπολογισμό με όρισμα την είσοδο, και
- ✔ επιστρέφουν μια συμβολοσειρά ως έξοδο.

Θεωρούμε ότι οι μηχανές αυτές έχουν μία κεφαλή που μπορεί να διαβάζει από / γράφει σε μία ταινία σύμβολα. Αν τοποθετήσουμε μία συμβολοσειρά εισόδου στην ταινία μιας τέτοιας μηχανής και την εκκινήσουμε, τότε, έως ότου τερματίσει, σε κάθε βήμα εκτελεί τις ακόλουθες λειτουργίες.

- Διαβάζει ένα σύμβολο από την ταινία.
- Αλλάζει την εσωτερική κατάσταση.
- Γράφει ένα σύμβολο στην ταινία.
- Γράφει ένα σύμβολο στην εσωτερική μνήμη.
- Μετακινεί την κεφαλή.

Ένα αυτόματο με λογική έξοδο λέμε ότι αποδέχεται μία συμβολοσειρά αν επιστρέφει Αλήθεια με αυτήν ως είσοδο, αλλιώς λέμε ότι την απορρίπτει. Λέμε ότι αναγνωρίζει μία γλώσσα αν αποδέχεται όλες τις συμβολοσειρές της γλώσσας και απορρίπτει όλες τις άλλες.

Αιτιοκρατικά Πεπερασμένα Αυτόματα

Το αιτιοκρατικό πεπερσμένο αυτόματο (deterministic finite automaton (DFA)) είναι μία κατηγορία αυτομάτου με τους ακόλουθους περιορισμούς:

- ✗ Δεν διαθέτει εσωτερική μνήμη και δεν γράφει στην ταινία.
- ✗ Σε κάθε βήμα, μετακινεί την κεφαλή υποχρεωτικά μία θέση δεξιά.
- ✗ Τερματίζει μόλις διαβάσει και το τελευταίο σύμβολο της συμβολοσειράς εισόδου.

Μαθηματικά, ένα DFA είναι μία πεντάδα $(Q, \Sigma, \delta, q_0, F)$, όπου:

- Q είναι το σύνολο όλων των δυνατών εσωτερικών του καταστάσεων,
- Σ είναι το αλφάβητο εισόδου,
- $\delta : Q \times \Sigma \to Q$ είναι η συνάρτηση μετάβασης από κατάσταση σε κατάσταση,
- q_0 είναι η αρχική του κατάσταση και
- F είναι το σύνολο των τερματικών του καταστάσεων.

Αποδεικνύεται πως τα DFAs είναι ακριβώς τα αυτόματα που αναγνωρίζουν όλες τις κανονικές γλώσσες και μόνο αυτές.