

ΘΥ 6: Συμβολοσειρές & γλώσσες

Επ. Καθ. Π. Κατσαρός

24/07/2014



ΑΡΙΣΤΟΤΕΛΕΙΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΕΣΣΑΛΟΝΙΚΗΣ

ΑΝΟΙΚΤΑ
ΑΚΑΔΗΜΑΙΚΑ
ΜΑΘΗΜΑΤΑ



Θεωρία Υπολογισμού

Ενότητα 6: Συμβολοσειρές & γλώσσες

Επ. Καθ. Π. Κατσαρός
Τμήμα Πληροφορικής



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης

ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΣΠΑ
2007-2013
Πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.
- Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδεια χρήσης, η άδεια χρήσης αναφέρεται ρητώς.



Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «Ανοικτά Ακαδημαϊκά Μαθήματα στο Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης» έχει χρηματοδοτήσει μόνο την 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.



- 1 Συμβολοσειρές
 - Αλφάβητο & συμβολοσειρές
 - Παράθεση συμβολοσειρών
 - Υποσυμβολοσειρά
 - Πολλαπλή παράθεση
 - Αντίστροφη συμβολοσειράς

- 2 Γλώσσες
 - Γλώσσες και ιδιότητες γλωσσών
 - Παράθεση γλωσσών
 - Kleene star γλώσσας

Αλφάβητο

Ορισμός 1 (Αλφάβητο)

Αλφάβητο Σ είναι ένα μη κενό, πεπερασμένο σύνολο συμβόλων.

- Παράδειγμα 1: $\{a,b,\dots,z,A,B,\dots,Z\}$
- Παράδειγμα 2: Δυαδικό αλφάβητο $\{0,1\}$
- Απλουστευτικά θεωρούμε ότι ένα αλφάβητο μπορεί να έχει ως σύμβολα μόνο γράμματα, αριθμούς και χαρακτήρες όπως οι $\#$ και $\$$.

Συμβολοσειρά

Ορισμός 2 (Συμβολοσειρά)

Συμβολοσειρά ενός αλφαβήτου είναι μια πεπερασμένη ακολουθία συμβόλων του.

- Αντί για ακολουθία με παρενθέσεις και κόμματα, απλά παραθέτουμε σύμβολα.
 - Παράδειγμα 1: Η fireplace είναι συμβολοσειρά του $\{a,b,\dots,z\}$
 - Παράδειγμα 2: Η 0111001 είναι συμβολοσειρά του $\{0,1\}$
- Η κενή συμβολοσειρά (ϵ ή ε) δεν περιέχει σύμβολα.
- Το σύνολο όλων των συμβολοσειρών (συμπεριλαμβανομένης της κενής) ενός αλφαβήτου Σ συμβολίζεται με Σ^* .

Μήκος συμβολοσειράς

Ορισμός 3 (Μήκος συμβολοσειράς)

Μήκος $|w|$ μιας συμβολοσειράς w είναι το μήκος της ως ακολουθία.

- Παράδειγμα: Το μήκος της συμβολοσειράς $acrd$ είναι 4
- Μια συμβολοσειρά $w \in \Sigma^*$ μπορεί να παρουσιαστεί και ως συνάρτηση $w : 1, \dots, |w| \mapsto \Sigma$, όπου η τιμή $w(j)$ είναι το σύμβολο που βρίσκεται στη θέση j .
Παράδειγμα: Αν $w = parrot$, τότε $w(1) = p, w(2) = a$ κλπ.

Παράθεση συμβολοσειρών

Ορισμός 4 (Παράθεση συμβολοσειρών)

Η παράθεση δύο συμβολοσειρών x και y γράφεται ως xy και αποτελείται από τα σύμβολα της x ακολουθούμενα από αυτά της y .

- Η παράθεση δύο συμβολοσειρών του ίδιου αλφαβήτου σχηματίζει μία νέα μεγαλύτερη συμβολοσειρά.
- Τυπικά γράφουμε $w = xy$ αν και μόνο αν
 - $|w| = |x| + |y|$ και
 - $w(j) = x(j)$ για $j = 1, \dots, |x|$ και $w(|x| + j) = y(j)$ για $j = 1, \dots, |y|$
- Παράδειγμα 1: $(01)(001) = 01001$
- Παράδειγμα 2: $(\text{water})(\text{proof}) = \text{waterproof}$

Ιδιότητες παράθεσης συμβολοσειρών

- Απορροφητικό στοιχείο
 $\forall w \in \Sigma^*, we = ew = w$
- Προσεταιριστικότητα
 $\forall w, x, y \in \Sigma^*, (wx)y = w(xy)$

Υποσυμβολοσειρά

Ορισμός 5 (Υποσυμβολοσειρά)

Η συμβολοσειρά v είναι υποσυμβολοσειρά μιας συμβολοσειράς w αν και μόνο αν υπάρχουν συμβολοσειρές x και y τέτοιες ώστε $w = xvy$.

- Κάθε συμβολοσειρά w είναι υποσυμβολοσειρά του εαυτού της, επειδή $w = ewe$.
- Η e είναι υποσυμβολοσειρά κάθε άλλης συμβολοσειράς, επειδή $w = wee$.
- Αν $w = xv$ για κάποιο x , τότε η v είναι κατάληξη της w .
Παράδειγμα: Η surfing είναι κατάληξη της kitesurfing
- Αν $w = vy$ για κάποια y , τότε η v είναι πρόθεμα της w .
Παράδειγμα: Η kite είναι πρόθεμα της kitesurfing

Πολλαπλή παράθεση

- Μια συμβολοσειρά μπορεί να περιλαμβάνει πολλές εμφανίσεις κάποιας υποσυμβολοσειράς.
Παράδειγμα: Η $ababab$ έχει τρεις εμφανίσεις της ab .
- Για κάθε συμβολοσειρά w και κάθε φυσικό αριθμό i , η συμβολοσειρά w^i ορίζεται ως

$$w^0 = e$$
$$w^{i+1} = w^i w \text{ για κάθε } i \geq 0$$

Παράδειγμα: $w^1 = w$ και $(do)^2 = dodo$

Αντίστροφη συμβολοσειράς

Ορισμός 6 (Αντίστροφη συμβολοσειράς)

Η αντίστροφη συμβολοσειρά w^R της w είναι η ακολουθία των συμβόλων της w «από το τέλος προς την αρχή».

Παράδειγμα: $\text{lemon}^R = \text{nomel}$

Τυπικός ορισμός με επαγωγή:

- 1 Αν w είναι συμβολοσειρά μήκους 0, τότε $w^R = w = e$.
- 2 Αν w είναι συμβολοσειρά μήκους $n + 1 > 0$ τότε $\exists a \in \Sigma : w = ua$ και $w^R = au^R$.

Αποδεικνύεται με επαγωγή ότι $(wx)^R = x^R w^R$

Γλώσσες

Ορισμός 7 (Γλώσσα)

Γλώσσα είναι κάθε σύνολο από συμβολοσειρές ενός αλφαβήτου Σ (δηλαδή κάθε υποσύνολο του Σ^*).

- Το Σ^* , το Σ και το \emptyset είναι γλώσσες.
- Μία πεπερασμένη γλώσσα μπορεί να ορισθεί παραθέτοντας όλες τις συμβολοσειρές που την απαρτίζουν.

Παράδειγμα: Η $\{\text{one, two, three}\}$ είναι γλώσσα με αλφάβητο το $\{a, b, \dots, z\}$

- Οι πιο πολλές γλώσσες από αυτές που συνήθως μελετάμε είναι *άπειρες*. Παραδείγματα:

$\{0, 01, 011, 0111, \dots\}$

$\{w \in \Sigma^* : \eta \ w \ \acute{\epsilon}\chi\epsilon\ \acute{\iota}\sigma\omicron\ \pi\acute{\lambda}\eta\theta\omicron\varsigma\ 1 \ \kappa\alpha\iota\ 0\}$

$\{w \in \Sigma^* : w = w^R\}$

- Ένας τρόπος για να ορίζουμε άπειρες γλώσσες:

$$L = \{w \in \Sigma^* : \eta \ w \ \acute{\epsilon}\chi\epsilon\ \text{την ιδιότητα } P\}$$

Άσκηση απόδειξης

Αν Σ είναι ένα αλφάβητο, τότε το Σ^* είναι ένα μετρήσιμα άπειρο σύνολο.

Απόδειξη.

- Ισχύει, επειδή μπορούμε να κατασκευάσουμε μία αμφιμονοσήμαντη αντιστοιχία $f: \mathbb{N} \rightarrow \Sigma^*$.
- Πρέπει πρώτα να διατάξουμε το αλφάβητο. Έστω $\Sigma = \{\alpha_1, \dots, \alpha_n\}$, όπου τα $\alpha_1, \dots, \alpha_n$ είναι όλα διαφορετικά.
- Τα στοιχεία του Σ^* μπορούν να απαριθμηθούν ως εξής:
 - για κάθε $k \geq 0$, όλες οι συμβολοσειρές μήκους k απαριθμούνται πριν από όλες τις συμβολοσειρές με μήκος $k + 1$.
 - οι n^k συμβολοσειρές μήκους k απαριθμούνται λεξικογραφικά, δηλαδή η $a_{i_1} \dots a_{i_k}$ προηγείται της $a_{j_1} \dots a_{j_k}$ υπό την προϋπόθεση ότι για κάποιο m , $0 \leq m \leq k - 1$, $i_l = j_l$ για $l = 1, \dots, m$ και $i_{m+1} < j_{m+1}$

Ιδιότητες γλωσσών

- Παράδειγμα 1: Αν $\Sigma = \{0, 1\}$ η λεξικογραφική σειρά θα ήταν η εξής:

$\epsilon, 0, 1, 00, 01, 10, 000, 001, 010, 011, \dots$

Σε αυτή τη σειρά οι μικρότερου μήκους συμβολοσειρές προηγούνται.

- Παράδειγμα 2: Αν $\Sigma = \{a, b, \dots, z\}$ η λεξικογραφική σειρά για τις ίσου μήκους συμβολοσειρές θα ήταν αυτή των λεξικών.
- Εφόσον οι γλώσσες είναι σύνολα μπορούν να συνδυαστούν με τις πράξεις συνόλων: ένωση, τομή και διαφορά.
- Όταν κάποιο αλφάβητο γίνεται κατανοητό από τα συμφραζόμενα τότε θα γράφουμε με \bar{A} το συμπλήρωμα του A , αντί για τη διαφορά $\Sigma^* - A$

Παράθεση γλωσσών

Ορισμός 8 (Παράθεση γλωσσών)

Η παράθεση των γλωσσών L_1 και L_2 είναι η γλώσσα $L = L_1L_2 = \{w \in \Sigma^* : \eta w = xy \text{ για κάποιο } x \in L_1 \text{ και } y \in L_2\}$.

- Παράδειγμα:

Αν $\Sigma = \{0,1\}$,

$L_1 = \{w \in \Sigma^* : w \text{ έχει άρτιο αριθμό } 0\}$ και

$L_2 = \{w \in \Sigma^* : w \text{ αρχίζει με } 0 \text{ και ακολουθούν μόνο } 1\}$

τότε

$L_1L_2 = \{w \in \Sigma^* : w \text{ έχει περιττό αριθμό } 0 \text{ και τελιώνει με διαδοχικά } 1\}$

Kleene star γλώσσας

Ορισμός 9 (Kleene star γλώσσας)

Η *Kleene star* μιας γλώσσας L συμβολίζεται με L^* και είναι το σύνολο των συμβολοσειρών που προκύπτουν από παράθεση 0 ή περισσότερων συμβολοσειρών της L .

$$L^* = \{w \in \Sigma^*: \eta \ w = w_1 \dots w_k \text{ για } k \geq 0 \text{ και } w_1 \dots w_k \in L \}$$

- Ισχύει $\emptyset^* = \{e\}$
- Συμβολίζουμε με L^+ την LL^*

Παράδειγμα:

Αν $L = \{01, 1, 100\}$ και $w = 1100011$

τότε $w \in L^*$ αφού κάθε μια από τις παρατιθέμενες υποσυμβολοσειρές της $(1)(100)(01)(1)$ ανήκει στην L

Τέλος ενότητας

Επεξεργασία: Εμμανουέλα Στάχτιαρη
Θεσσαλονίκη, 24/07/2014