ΠΛΗ30

ΕΝΟΤΗΤΑ 3: ΚΑΝΟΝΙΚΕΣ ΓΛΩΣΣΕΣ

Μάθημα 3.1:
Κανονικές Εκφράσεις

Δημήτρης Ψούνης

www.psounis.gr

ΠΕΡΙΕΧΟΜΕΝΑ

Α. Σκοπός του Μαθήματος

Οι στόχοι του μαθήματος είναι:

Επίπεδο Α
- Κατασκευή Κανονικών Εκφράσεων
- Ορισμός Κανονικής Έκφρασης και Κανονικής Γλώσσας
- Ορισμός Πράξεων Γλωσσών

Επίπεδο Β
- Τυπικός Ορισμός Κανονικής Γλώσσας

Επίπεδο Γ
- (-)

Β. Θεωρία

1.Εισαγωγικοί Ορισμοί

1.Αλφάβητο

Ορισμός:
Αλφάβητο είναι οποιοδήποτε πεπερασμένο σύνολο συμβόλων. Συμβολίζεται με Σ

Παραδείγματα:
- Σ={0,1} το δυαδικό αλφάβητο
- Σ={a,b}
- Σ={Α,Β,Γ,…,Ω} το αλφαβητο των ελληνικών κεφαλαίων γραμμάτων

Ορισμός:
Έστω Σ ένα αλφάβητο. Το σύνολο όλων των συμβολοσειρών που μπορούμε να παράγουμε από σύμβολα του Σ, συμβολίζεται με **Σ***.
Το σύνολο Σ* καλείται αστέρι Kleene του Σ και συμβολίζει την διάταξη 0 ή περισσότερων συμβόλων του Σ

Παράδειγμα
Έστω Σ={0,1} το δυαδικό αλφάβητο
Τότε Σ*={ε,0,1,00,01,10,11,000,001,010,011,100,101,110,111,…}

Ορισμός: Το ε είναι η συμβολοσειρά μήκους 0 και καλείται κενή συμβολοσειρά

Β. Θεωρία
1.Εισαγωγικοί Ορισμοί
2.Γλώσσα

Ορισμός:
Γλώσσα ενός αλφαβήτου Σ είναι οποιοδήποτε υποσύνολο του Σ*. Συνήθως συμβολίζεται με L.

Παραδείγματα γλωσσών του Σ={0,1}:

- L_1={ w | w αρχίζει με 00}
- L_2={ w | w περιέχει το 11}
- L_3={ w | w τελειώνει με 01}
- L_4={ w | w έχει μήκος τουλάχιστον 2}
- L_5={ w | w έχει άρτιο πλήθος 1}
- L_6={ w | Η w είναι παλινδρομική}
- L_7={ w | Ο δυαδικός αριθμός που αντιστοιχεί στην w είναι πρώτος αριθμός}
- ….

➢ Μπορούμε να κατασκευάσουμε άπειρες γλώσσες ενός αλφαβήτου.

Β. Θεωρία

1.Εισαγωγικοί Ορισμοί

3.Πράξεις Γλωσσών

Ορισμός:
Έστω L, L_1, L_2 γλώσσες του αλφαβήτου Σ. Ορίζονται οι γλώσσες:

- Ένωση Γλωσσών: $L_1 \cup L_2 = \{w|w \in L_1 \text{ ή } w \in L_2\}$
- Τομή Γλωσσών: $L_1 \cap L_2 = \{w|w \in L_1 \text{ και } w \in L_2\}$
- Παράθεση (ή Συνένωση) Γλωσσών: $L_1L_2 = \{xy|x \in L_1 \text{ και } y \in L_2\}$
- Συμπλήρωμα Γλωσσας: $\overline{L} = \{w|w \notin L\}$
- Αστέρι Kleene Γλωσσας: $L^* = \{w|\text{Η } w$ είναι παράθεση 0 ή περισσοτέρων συμβολοσειρών της L.

Παράδειγμα στο Σ={α,β}
Αν L1={w|w αρχίζει με α} και L2={w|w τελειώνει με β}
Τότε $L_1 \cap L_2$={w| w αρχίζει με α και τελειώνει με β}

Β. Θεωρία

2.Κανονικές Εκφράσεις

1. Συντακτικό Κανονικών Εκφράσεων

- Μια <u>κανονική έκφραση</u> είναι ένας εύκολος τρόπος περιγραφής των συμβολοσειρών που ανήκουν σε μία <u>κανονική γλώσσα</u>.
- Οι <u>κανονικές γλώσσες</u> είναι οι απλούστερες γλώσσες που μπορούν να κατασκευασθούν.
- <u>Παράδειγμα:</u> Ποια η κανονική έκφραση της γλώσσας $L=\{w\in\{0,1\}^*|w$ αρχίζει με $1\}$;
  - Η κανονική έκφραση είναι <u>1(0+1)*</u>
    - Η κανονική έκφραση είναι μια συμβολοσειρά που διαβάζεται από αριστερά προς τα δεξιά.
    - Ο πρώτος 1 σημαίνει ότι ξεκινά με 1.
    - Ο όρος (0+1)* διαβάζεται 0 ή 1 (λόγω του 0+1) επαναλαμβάνεται 0 ή περισσότερες φορές.
  - Αν εφαρμόσουμε το αστέρι Kleene διαδοχικά 0,1,2,… φορές παίρνουμε:
    - 1
    - 1(0+1)=10 ή 11
    - 1(0+1)(0+1)=100 ή 101 ή 110 ή 111
    - ….

Β. Θεωρία

2.Κανονικές Εκφράσεις

1. Συντακτικό Κανονικών Εκφράσεων

Πρακτικά:

Μια κανονική έκφραση κατασκευάζεται με τα εξής στοιχεία:

1. Τα σύμβολα του αλφαβήτου
2. Το + που διαβάζεται «ή διαζευκτικό»
3. Το * που είναι το αστέρι Kleene. Διαβάζεται «0 ή περισσότερες φορές».
4. Παρενθέσεις που υποδεικνύουν την προτεραιότητα των πράξεων
5. Υπονοείται και η πράξη της παράθεσης που είναι όταν έχουμε δύο διαδοχικές παραστάσεις και σημαίνει ότι παραθέτουμε (βάζουμε διαδοχικά) την πρώτη και την δεύτερη παράσταση.

Η προτεραιότητα των συμβόλων είναι πρώτα το αστέρι Kleene, έπειτα η παράθεση και έπειτα το +, εφόσον αυτή δεν καθορίζεται με παρενθέσεις.

Π.χ. η κανονική έκφραση 11*+(00)* ορίζει την γλώσσα που περιέχει συμβολοσειρές που:

- Ή έχουν τουλάχιστον έναν άσσο (και κανένα μηδενικό)
- Ή συμβολοσειρές που έχουν άρτια μηδενικά (και κανέναν άσσο)

Β. Θεωρία

2.Κανονικές Εκφράσεις

2. Παραδείγματα Κανονικών Εκφράσεων

- ΑΣΚΗΣΗ: Κατασκευάστε Κανονικές Εκφράσεις για τις Γλώσσες του {0,1}:
  - L_1={ w | w τελειώνει με 1 }
  - L_2={ w | w αρχίζει με 00 }
  - L_3={ w | w περιέχει το 01 }
  - L_4={ w | w έχει μήκος 2 }
  - L_5={ w | w έχει μήκος τουλάχιστον 2 }
  - L_6={ w | w έχει μήκος το πολύ 2 }
  - L_7={ w | w έχει άρτιο μήκος }
  - L_8={ w | w έχει περιττό μήκος }
  - L_9={ w | w έχει άρτιο μήκος ή αρχίζει με 00}
  - L_{10}={ w | w δεν αρχίζει με 01}
  - L_{11}={ w | w δεν περιέχει το 01}
  - L_{12}={ w | w περιέχει άρτια 0}

Β. Θεωρία

2.Κανονικές Εκφράσεις

3. Τυπικός Ορισμός Κανονικής Έκφρασης

- Κάθε κανονική έκφραση αντιστοιχεί σε μία γλώσσα. Η κατασκευή της γλώσσας που αντιστοιχεί στην έκφραση μπορεί να γίνει με τον τυπικό ορισμό:

Ορισμός:

- ∅ είναι η κανονική έκφραση που αντιστοιχεί στην κενή γλώσσα.
- ε είναι η κανονική έκφραση που αντιστοιχεί στην γλώσσα {ε}
- Για κάθε σύμβολο $\sigma \in \Sigma$, σ είναι η κανονική έκφραση που αντιστοιχεί στην γλωσσα {σ}
- Αν r και s είναι εκφράσεις που αντιστοιχούν στις γλώσσες L_r και L_s, τότε και οι $(rs), (r+s)$ και r^* είναι οι κανονικές εκφράσεις που αντιστοιχούν στις κανονικές γλώσσες $\mathrm{L_r L_s}$, $\mathrm{L_r + L_s}$, L_r^*
- Τίποτα δεν είναι κανονική έκφραση αν δεν παράγεται από κάποιον από τους παραπάνω κανόνες.

Β. Θεωρία

2.Κανονικές Εκφράσεις

4. Ορισμός Κανονικής Γλώσσας

Ορισμός:

- Μία γλώσσα θα λέγεται **κανονική γλώσσα** αν και μόνο αν
  - Υπάρχει κανονική έκφραση που την περιγράφει.

- Συνεπώς όλες οι γλώσσες της προηγούμενης άσκησης είναι κανονικές.
- Υπάρχουν και άλλοι τύποι γλωσσών που θα δούμε σε επόμενες ενότητες:
  - Γλώσσες Ανεξάρτητες συμφραζομένων
  - Αποφασίσιμες Γλώσσες
  - Αποδεκτές Γλώσσες
- Κάθε οικογένεια γλωσσών σχετίζεται με το πόσο δύσκολο είναι να υπολογιστούν τα μέλη της. Έτσι κάθε μία συμβολίζει και ένα επίπεδο δυσκολίας του υπολογισμού.
  - Οι καν. γλώσσες υπολογίζονται από Πεπερασμένο Αυτόματο (Ενοτ.3)
  - Οι γλώσσες χωρίς συμφραζόμενα από Αυτόματο Στοίβας (Ενότητα 4)
  - Οι αποφασίσιμες γλώσσες από Μηχανή Turing (Ενότητα 5)
  - Οι αποδεκτές γλώσσες ΔΕΝ υπολογίζονται (Ενότητα 5)

Β. Θεωρία

2.Κανονικές Εκφράσεις

4. Ορισμός Κανονικής Γλώσσας

- ΑΣΚΗΣΗ: Ορίστε με περιγραφικό τρόπο τις γλώσσες των κανονικών εκφράσεων:
  1. (0+1)*11(0+1)*
  2. 0(0+1)*10
  3. 00(0+1)*11(0+1)*11
  4. 0(0+1)*0 + 1(0+1)*1
  5. 1(0+1)*0 + 0(0+1)*1
  6. 0*(10*10*)*
  7. 0(0+1)*+(0+1)*1
  8. 1(00+01+10+11)*
  9. (0+10*1)*
  10. 0*(10*10*10*)*

Β. Θεωρία

2.Κανονικές Εκφράσεις

5. Κάθε Πεπερασμένη Γλώσσα είναι Κανονική

Θεώρημα:
Κάθε πεπερασμένη γλωσσα είναι κανονική
Απόδειξη: Πράγματι περιγράφεται από την κανονική έκφραση που με + θα ενώνει όλες τις συμβολοσειρές της γλώσσας

Παράδειγμα:

Έστω L={ε,0,1,00,01,10,11}

Η L είναι κανονική γιατί περιγράφεται από την κανονική έκφραση:

ε+0+1+00+01+10+11

Γ. Ασκήσεις
Άσκηση Κατανόησης 1

Δίδονται οι γλώσσες του αλφαβήτου {0,1}:
L_1={w|w αρχίζει με 0}
L_2={w|w τελειώνει με 1}
Εξετάστε ποιες από τις ακόλουθες συμβολοσειρές ανήκουν στις γλώσσες:

$$L_1 \cup L_2,\quad L_1 \cap L_2,\quad L_1L_2,\quad L_2L_1,\quad \overline{L_1},\quad \overline{L_2},\quad L_1^*,\quad L_2^*$$

- w_1=0011
- w_2=0010
- w_3=1111
- w_4=1011

Γ. Ασκήσεις

Εφαρμογή 1

Κατασκευάστε κανονικές εκφράσεις για τις γλώσσες:

- $L_1=\{w\in\{0,1\}^*|$ η w ξεκινά με το 00 και τελειώνει με 10$\}$
- $L_2=\{w\in\{0,1\}^*|$ η w ξεκινά με το 11, περιέχει το 00 και τελειώνει με 10$\}$
- $L_3=\{w\in\{a,b\}^*|$ η w περιέχει το aabb$\}$
- $L_4=\{w\in\{a,b\}^*|$ η w περιέχει τρία συνεχόμενα a$\}$
- $L_5=\{w\in\{a,b\}^*|$ η w περιέχει άρτια a ή περιττά b$\}$

Γ. Ασκήσεις
Εφαρμογή 2

Κατασκευάστε κανονικές εκφράσεις για τις γλώσσες:

- $L_1=\{w\in\{0,1\}^*|$ η w ξεκινά με 01$\}$
- $L_2=\{w\in\{0,1\}^*|$ η w περιέχει το 01$\}$
- $L_3=\{w\in\{0,1\}^*|$ η w τελειώνει με 01$\}$
- $L_4=\{w\in\{0,1\}^*|$ η w δεν ξεκινά με 01$\}$
- $L_5=\{w\in\{0,1\}^*|$ η w δεν περιέχει το 01$\}$
- $L_6=\{w\in\{0,1\}^*|$ η w δεν τελειώνει με 01$\}$

Γ. Ασκήσεις
Εφαρμογή 3

Δώστε τις γλώσσες με αλφάβητο {0,1} που αντιστοιχούν στις παρακάτω κανονικές εκφράσεις:

1. $L = 0^*1(0^*10^*1)^*0^*$

2. $L = 1^* + 1^*01^* + 1^*01^*01^*$

3. $L = (0 + 1)^*\, 11 + (0 + 1)^*\, 10 + (0 + 1)^*\, 01 + 0 + 1 + \varepsilon$

4. $L = 1(0 + 1)^* + 0(0 + 1)^*$

5. $L = 1^*(01^*01^*01^*)^*$

Γ. Ασκήσεις
Εφαρμογή 4

1. Δώστε κανονική έκφραση για τη γλώσσα με όλες τις λέξεις του Σ= {a,b} που δεν τελειώνουν σε ab ή ba.

2. Περιγράψτε με λόγια την γλώσσα με κανονική έκφραση b*a(b*ab*a)*b*.

3. Είναι κανονική στο Σ= {a,b} γλώσσα η $L=\{a^i b^i \mid 0 \leq i \leq 3\}$

4. Είναι κανονική στο Σ= {0,1} η γλώσσα $L' = \{(00111)^n \mid n \geq 0\}$ όπου n οποιοσδήποτε φυσικός αριθμός;

5. Είναι κανονική γλώσσα στο Σ= {a,b} η $N=\{(a+b)^i \mid i>2\}$