Πεπερασμένη περιγραφή (αναπαράσταση) γλωσσών

Μπορούμε να σκεφτούμε πολλές περιγραφές γλωσσών. Π.χ. μπορούμε να ορίσουμε ως εξής μια γλώσσα:

$$L = \{w : w \in \{0,1\}^* \text{ και περιέχει ίσο αριθμό από } 0 \text{ και } 1\}$$

Γενικότερα οι περιγραφές είναι της μορφής:

$$L = \{w : w \text{ ικανοποιεί την ιδιότητα } T\}, \text{ για κάποιο } T$$

Η ιδιότητα T πρέπει να έχει πεπερασμένη περιγραφή.

Γενικά υπάρχουν πολλά είδη περιγραφών. Εδώ θα ξεκινήσουμε με μια απλή και φυσική περιγραφή και στη συνέχεια θα την εμπλουτίσουμε.

Κανονικές εκφράσεις (ΚΕ / Regular expressions)

1. $\emptyset, \sigma \in \Sigma$
2. Αν, α και β είναι ΚΕ, τότε $(\alpha \cup \beta)$ είναι ΚΕ
3. Αν, α και β είναι ΚΕ, τότε $(\alpha\beta)$ είναι ΚΕ
4. Αν, α είναι ΚΕ, τότε (α^*) είναι ΚΕ
5. Τίποτα άλλο δεν είναι ΚΕ

Γλώσσες και κανονικές εκφράσεις

Σε κάθε ΚΕ αντιστοιχούμε μία γλώσσα ως εξής: Αν α είναι μία ΚΕ, τότε $\mathcal{L}(\alpha)$ θα συμβολίζει τη γλώσσα.

1. $\mathcal{L}(\emptyset) = \emptyset, \mathcal{L}(\sigma) = \{\sigma\}$
2. $\mathcal{L}((\alpha \cup \beta)) = \mathcal{L}(\alpha) \cup \mathcal{L}(\beta)$
3. $\mathcal{L}((\alpha\beta)) = \mathcal{L}(\alpha)\mathcal{L}(\beta)$
4. $\mathcal{L}(\alpha^*) = \mathcal{L}(\alpha)^*$

Μια γλώσσα L λεγεται κανονική αν και μόνο αν υπάρχει αντίστοιχη κανονική έκφραση ρ, δηλαδη $L = \mathcal{L}(\rho)$.

Παραδείγματα κανονικών εκφράσεων

Ειναι οι επομενες κανονικες εκφρασεις του αλφαβητου $\{a, b\}$;

- $\emptyset$, a, b: Ναι
- $a \cup b$: Οχι γιατι λειπουν οι παρενθεσεις. Συνηθως ομως παραλειπουμε παρενθεσεις αν αυτο δεν προκαλει συγχυση.
- $((a \cup b)(a^*))$: Ναι
- $(a)^*$: Οχι
- $((a \cup (a^*))^*)$: Ναι

Παραδειγματα γλωσσων κανονικων εκφρασεων

Ποια η γλωσσα της κανονικης εκφρασης $(a \cup b)^*$;

Απαντηση:

$$
\begin{aligned}
\mathcal{L}((a \cup b)^*) &= \mathcal{L}((a \cup b))^* && [\text{κανονας 4}] \\
&= (\mathcal{L}(a) \cup \mathcal{L}(b))^* && [\text{κανονας 2}] \\
&= (\{a\} \cup \{b\})^* && [\text{κανονας 1}] \\
&= \{a, b\}^*
\end{aligned}
$$

Δηλαδη η γλωσσα που περιεχει ολες τις συμβολοσειρες.

Παραδειγματα γλωσσων κανονικων εκφρασεων

Ποια η γλωσσα της κανονικης εκφρασης $((a \cup ba)^*)$;

Απαντηση:

$$
\begin{aligned}
\mathcal{L}(((a \cup (ba))^*)) &= \mathcal{L}((a \cup (ba)))^* && [\text{κανονας 4}] \\
&= (\mathcal{L}(a) \cup \mathcal{L}((ba)))^* && [\text{κανονας 2}] \\
&= (\mathcal{L}(a) \cup \mathcal{L}(b)\mathcal{L}(a)))^* && [\text{κανονας 3}] \\
&= (\{a\} \cup \{ba\})^* && [\text{κανονας 1}] \\
&= \{a, ba\}^*
\end{aligned}
$$

Δηλαδη η γλωσσα που περιεχει τις συμβολοσειρες στις οποιες καθε b ακολουθειται παντα απο a.

Κανονικές εκφράσεις (ΚΕ / Regular expressions)

1. $\emptyset, \sigma \in \Sigma$
2. Αν, α και β είναι ΚΕ, τότε $(\alpha \cup \beta)$ είναι ΚΕ
3. Αν, α και β είναι ΚΕ, τότε $(\alpha\beta)$ είναι ΚΕ
4. Αν, α είναι ΚΕ, τότε (α^*) είναι ΚΕ
5. Τίποτα άλλο δεν είναι ΚΕ

Γλώσσες και κανονικές εκφράσεις

Σε κάθε ΚΕ αντιστοιχούμε μία γλώσσα ως εξής: Αν α είναι μία ΚΕ, τότε $\mathcal{L}(\alpha)$ θα συμβολίζει τη γλώσσα.

1. $\mathcal{L}(\emptyset) = \emptyset, \mathcal{L}(\sigma) = \{\sigma\}$
2. $\mathcal{L}((\alpha \cup \beta)) = \mathcal{L}(\alpha) \cup \mathcal{L}(\beta)$
3. $\mathcal{L}((\alpha\beta)) = \mathcal{L}(\alpha)\mathcal{L}(\beta)$
4. $\mathcal{L}(\alpha^*) = \mathcal{L}(\alpha)^*$

Μια γλώσσα L λεγεται κανονική αν και μόνο αν υπάρχει αντίστοιχη κανονική έκφραση ρ, δηλαδη $L = \mathcal{L}(\rho)$.

Παραδείγματα γλωσσών κανονικών εκφράσεων

Δώστε μια κανονική έκφραση για τη γλώσσα

$$L = \{w : w \in \{a, b\}^* \text{ και } w \text{ περιεχει περιττο αριθμο απο } a\}.$$

Απαντηση:$b^*a(b^*ab^*a)^*b^*$

οπου παραλειπονται αυτονοητες παρανθεσεις. Ποια ειναι η κανονικη έκφραση αν βαλουμε τις καταλληλες παρενθεσεις;

Παρατηρηση: Μια γλωσσα μπορει να εχει πολλες κανονικες εκφρασεις.

Σχέσεις Κλειστότητας

Κλειστότητα: : Λεμε οτι ενα συνολο γλωσσων A κλεινεται απο μια δυαδικη πραξη $\oplus$ (π.χ. ενωση, τομη, παραθεση) αν και μονο αν για καθε δυο γλωσσες L_1, L_2 του A, η γλωσσα $L_1 \oplus L_2$ ανηκει επισης στο A.

Αναλογα οριζουμε την κλειστοτητα για μοναδιαιες πραξεις (π.χ. Kleene star, συμπληρωμα συνολου).

Θεωρημα: Οι κανονικες γλωσσες κλεινονται απο την **ενωση**, **παραθεση**, και **Kleene star**.

Ουσιαστικα, οι κανονικες γλωσσες ενος αλφαβητου Σ οριζονται ακριβως σαν το συνολο των γλωσσων που περιεχει τις στοιχειωδεις γλωσσες $\emptyset$, και $\{\sigma\}$ για καθε συμβολο $\sigma \in \Sigma$ και κλεινεται απο την ενωση, παραθεση, και Kleene star.

Ερωτηση: Κλεινονται οι κανονικες γλωσσες ως προς την τομη; Το συμπληρωμα;

Γραμματικές και Αυτόματα

Οι κανονικες εκφρασεις αποτελουν ενα ειδος γραμματικης: Γενικα **γραμματικη** μιας γλωσσας ειναι ενα συστημα που περιγραφει πως μπορουμε να **παραγουμε** τις συμβολοσειρες μιας γλωσσας.

Ενα αλλος τροπος για να περιγραψουμε μια γλωσσα ειναι τα αυτοματα: Γενικα, **αυτοματο** μιας γλωσσας ειναι ενας μηχανισμος (αλγοριθμος) που μας επιτρεπει να μπορουμε να **αναγνωριζουμε** με συστηματικο τροπο αν μια συμβολοσειρα ανηκει στις συμβολοσειρες μιας γλωσσας ή οχι.

Στις περισσοτερες περιπτωσεις, θα μελεταμε καποιο ειδος γραμματικων σε συνδυασμο με το ειδος αυτοματων που αναγνωριζουν την ιδια γλωσσα.