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

Είναι οι επόμενες κανονικές εκφράσεις του αλφαβήτου $\{a, b\}$;

- $\emptyset$, a, b: Ναι
- $a \cup b$: Όχι γιατί λείπουν οι παρενθέσεις. Συνήθως όμως παραλείπουμε παρενθέσεις αν αυτό δεν προκαλεί σύγχυση.
- $((a \cup b)(a^*))$: Ναι
- $(a)^*$: Οχι
- $((a \cup (a^*))^*)$: Ναι

Παραδείγματα γλωσσών κανονικών εκφράσεων

Ποιά η γλώσσα της κανονικής έκφρασης $(a \cup b)^*$;

Απάντηση:

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

Δηλαδή η γλώσσα που περιέχει όλες τις συμβολοσειρές.

Παραδείγματα γλωσσών κανονικών εκφράσεων

Ποιά η γλώσσα της κανονικής έκφρασης $((a \cup ba)^*)$;

Απάντηση:

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

Δηλαδή η γλώσσα που περιέχει τις συμβολοσειρές στις οποίες κάθε b ακολουθείται πάντα από 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 λέγεται κανονική αν και μόνο αν υπάρχει αντίστοιχη κανονική έκφραση ρ, δηλαδή $L = \mathcal{L}(\rho)$.

Παραδείγματα γλωσσών κανονικών εκφράσεων

Δώστε μια κανονική έκφραση για τη γλώσσα

$$L = \{w : w \in \{a, b\}^* \text{ και } w \text{ περιέχει περιττό αριθμό από } a\}.$$

Απάντηση:$b^*a(b^*ab^*a)^*b^*$

όπου παραλείπονται αυτονόητες παρενθέσεις. Ποια είναι η κανονική έκφραση αν βάλουμε τις κατάλληλες παρενθέσεις;

Παρατήρηση: Μια γλώσσα μπορεί να έχει πολλές κανονικές εκφράσεις.

Σχέσεις Κλειστότητας

Κλειστότητα: : Λέμε οτι ένα σύνολο γλωσσών A κλείνεται απο μια δ-υαδική πράξη $\oplus$ (π.χ. ένωση, τομή, παράθεση) αν και μόνο αν για κάθε δύο γλώσσες L_1, L_2 του A, η γλώσσα $L_1 \oplus L_2$ ανήκει επίσης στο A.

Ανάλογα ορίζουμε την κλειστότητα για μοναδιαίες πράξεις (π.χ. Kleene star, συμπλήρωμα συνόλου).

Θεωρημα: Οι κανονικές γλώσσες κλείνονται απο την **ένωση**, **παράθεση**, και **Kleene star**.

Ουσιαστικά, οι κανονικές γλώσσες ενός αλφαβήτου Σ ορίζονται ακριβώς σαν το σύνολο των γλωσσών που περιέχει τις στοιχειώδεις γλώσσες $\emptyset$, και $\{\sigma\}$ για κάθε σύμβολο $\sigma \in \Sigma$ και κλείνεται απο την ένωση, παράθεση, και Kleene star.

Ερώτηση: Κλείνονται οι κανονικές γλώσσες ως προς την τομή; Το συμπλήρωμα;

Γραμματικές και Αυτόματα

Οι κανονικές εκφράσεις αποτελούν ενα είδος γραμματικης: Γενικά **γραμματική** μιας γλώσσας είναι ενα σύστημα που περιγράφει πως μπορούμε να **παράγουμε** τις συμβολοσειρές μιας γλώσσας.

Ενα άλλος τρόπος για να περιγράψουμε μια γλώσσα είναι τα αυτόματα: Γενικά, **αυτόματο** μιας γλώσσας είναι ένας μηχανισμός (αλγόριθμος) που μας επιτρέπει να μπορούμε να **αναγνωρίζουμε** με συστηματικό τρόπο αν μια συμβολοσειρά ανήκει στις συμβολοσειρές μιας γλώσσας ή όχι.

Στις περισσότερες περιπτώσεις, θα μελετάμε κάποιο είδος γραμματικών σε συνδυασμό με το είδος αυτομάτων που αναγνωρίζουν την ίδια γλώσσα.