Κανονικές Εκφράσεις και Κανονικές Γλώσσες

Δημήτρης Φωτάκης

Τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων

Πανεπιστήμιο Αιγαίου

Ανακεφαλαίωση

- Προβλήματα απόφασης κωδικοποιούνται σε γλώσσες.
- Γλώσσα: σύνολο συμβολοσειρών σε πεπερασμένο αλφάβητο.
  - Πράξεις συμβολοσειρών: παράθεση, επανάληψη, αντίστροφη.
  - Πράξεις γλωσσών: ένωση, τομή, συμπλήρωμα, παράθεση, Kleene star.
- Αρίθμηση στοιχείων συνόλου S: διατύπωση 1-1 και επί συνάρτησης από S σε N (φυσικούς).
  - Σ^* (σύνολο συμβολοσειρών): αριθμήσιμο.
  - 2^{Σ^*} (σύνολο γλωσσών): μη-αριθμήσιμο (διαγωνιοποίηση).
  - Υπάρχουν γλώσσες χωρίς πεπερασμένη αναπαράσταση.

Αναπαράσταση Γλωσσών

- Πως παράγουμε συμβ/ρές εφαρμόζοντας πράξεις: «μορφολογική» ή «παραγωγική» αναπαράσταση.
  - $L_1 = \{01, 11\}^* \cup (\{1\}\{00, 1\}^*)$
  - (Πεπερασμένη) αναπαράσταση: συμβ/ρά σε αλφάβητο.
  - Κάθε σχήμα πεπερασμένης αναπαράστασης αδυνατεί να περιγράψει πολλές γλώσσες!
- Ιδιότητες που αναγνωρίζονται «εύκολα»: «αλγοριθμική».
  - $L_2 = \{w \in \Sigma^* : w \text{ αρχίζει από } 1\}$
  - (Υπολογιστικές) μηχανές αναγνώρισης.
- Περιγραφή με απλές αναπαραστάσεις.
- Αναγνώριση από απλές μηχανές.

Κανονικές Εκφράσεις

- Κανονική έκφραση αλφάβητου Σ :
  1. Το $\emptyset$ και κάθε $\sigma \in \Sigma$ είναι ΚΕ.
  2. Αν α και β είναι ΚΕ, τότε (α $\cup$ β) είναι ΚΕ.
  3. Αν α και β είναι ΚΕ, τότε (α β) είναι ΚΕ.
  4. Αν α είναι ΚΕ, τότε α^* είναι ΚΕ.
  5. Τίποτα άλλο δεν είναι ΚΕ.
- Παραδείγματα ΚΕ του Σ = {0, 1}: (παραλείπουμε παρενθέσεις)
  - ε γιατί εκφράζεται σαν $\emptyset^*$
  - $(0 \cup 1)$, $(0 \cup 1)1^*$, $((0 \cup 1)1(0 \cup 1)0)^*$
  - $(10 \cup 01)^* \cup (0001 \cup 1000)^*$, $(1 \cup 1^*)^*$
  - $0^*10^*010^*(10^* \cup \emptyset^*)$

Κανονικές Γλώσσες

- Γλώσσες που αναπαριστώνται από ΚΕ.
- Αν α μια ΚΕ, $\mathcal{L}(\alpha)$ είναι η αντίστοιχη κανονική 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. $\mathcal{L}(\alpha^*) = \mathcal{L}(\alpha)^*$
- Κάθε κανονική γλώσσα αναπαρίσταται από άπειρες κανονικές εκφράσεις.
- Ενδιαφέρουσες ασκήσεις:
  - Γλώσσα οριζόμενη με ιδιότητα $\to$ Κανονική έκφραση.
  - Κανονική έκφραση $\to$ Ιδιότητα.

Παραδείγματα

- Ποια γλώσσα αντιστοιχεί στην ΚΕ: $(a \cup b)^*$

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

- Ποια γλώσσα αντιστοιχεί στην ΚΕ: $((a \cup b)^*(ba))$

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

Παραδείγματα

- Ποια γλώσσα αντιστοιχεί στην ΚΕ: $(a \cup (ba))^*$

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

  - Κάθε *b* ακολουθείται από *a* (ή δεν περιέχει δύο συνεχόμενα *b*).
- Γλώσσα $0^* \cup 0^*(1 \cup 11)(00^*(1 \cup 11))^*0^*$
  - Δεν περιέχει 111.

Παραδείγματα

- ΚΕ για $L = \{w \in \{a, b\}^* : w \text{ περιέχει } ba \text{ και τελειώνει σε } b\}$
  $(a \cup b)^*ba(a \cup b)^*b$
- ΚΕ για $L = \{w \in \{0, 1\}^* : w \text{ περιέχει ζυγό αριθμό } 0\}$
  $(1^*01^*0)^*1^*$ ή $(1 \cup 01^*0)^*$
- $L = \{w \in \{0, 1\}^* : w \text{ τελειώνει με 1 και δεν περιέχει } 00\}$
  $(1 \cup 01)(1 \cup 01)^*$
- ΚΕ για $L = \{w \in \{0, 1\}^* : w \text{ περιέχει μία εμφάνιση του } 00\}$
  $(1 \cup 01)^*00(1 \cup 10)^*$
- ΚΕ για $L = \{w \in \{0, 1\}^* : w \text{ δεν περιέχει } 110\}$
  $(0 \cup 10)^*1^*$

Παρατηρήσεις

- Κάθε πεπερασμένη γλώσσα είναι κανονική.
  - Απόδειξη (εύκολη) με επαγωγή στον #συμβ/ρών.
- Κανονική γλώσσα που αναπαρίσταται από ΚΕ χωρίς * ;
  - Είναι πεπερασμένη.
  - Ο τελεστής * μοναδικός που δημιουργεί άπειρο πλήθος συμβ/ρών.

Κλειστότητα

- Σύνολο (γλωσσών) $\mathcal{C}$ είναι κλειστό ως προς $\oplus$ (π.χ. ένωση, παράθεση, τομή) αν $\forall L_1, L_2 \in \mathcal{C}, L_1 \oplus L_2 \in \mathcal{C}$
  - Παρόμοια κλειστό ως προς * : $\forall L \in \mathcal{C}, L^* \in \mathcal{C}$
- Σύνολο κανονικών γλωσσών κλειστό ως προς ένωση, παράθεση, και Kleene star.
  - Απόδειξη: εξ' ορισμού!
  - ΚΓ στο Σ ορίζονται σαν το (ελάχιστο) σύνολο γλωσσών που περιέχει $\emptyset$ και $\{\sigma\}$ για κάθε $\sigma \in \Sigma$ και είναι κλειστό ως προς ένωση, παράθεση, και Kleene star.
- Σύνολο κανονικών γλωσσών είναι κλειστό και ως προς τομή, διαφορά, και συμπλήρωμα.

Γραμματικές και Αυτόματα

- Κανονικές εκφράσεις αποτελούν είδος γραμματικής.
  - Γραμματική: σύστημα που περιγράφει πως παράγουμε συμβ/ρές γλώσσας.
- Πεπερασμένα αυτόματα αποτελούν (υπολογιστικές) μηχανές.
  - (Υπολογιστική) μηχανή: αλγόριθμος αναγνώρισης συμβ/ρών που ανήκουν σε γλώσσα.
- Μελέτη γραμματικών και είδους μηχανών που αναγνωρίζει αντίστοιχες γλώσσες.
  - Κανονικές Γλώσσες = Γλώσσες που αναγνωρίζονται από Πεπερασμένα Αυτόματα.