ΘΕΩΡΙΑ ΥΠΟΛΟΓΙΣΜΩΝ ΚΑΙ ΑΥΤΟΜΑΤΩΝ

Ενότητα 1: Εισαγωγή

Ρεφανίδης Ιωάννης

Τμήμα Εφαρμοσμένης 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 άδειας 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 Πανεπιστήμιο Μακεδονίας» έχει χρηματοδοτήσει μόνο τη 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.

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Εισαγωγή

Θεωρία Υπολογισμού

- Τι είναι «Υπολογισμός»
- Τι μπορεί να κάνει ένας υπολογιστής;
- Τι δεν μπορεί να κάνει ένας υπολογιστής;
- Ποια προβλήματα μπορεί να λύσει ο υπολογιστής αποδοτικά;

Αποφασισιμότητα (decidability)

- Μελέτη του τι μπορούν να κάνουν οι υπολογιστές;
  - Τα προβλήματα που μπορούν να λύσουν οι υπολογιστές ονομάζονται αποφασίσιμα (decidable).
- Μη αποφασίσιμο πρόβλημα:
  - Να ελεγχθεί εάν δύο διαφορετικά προγράμματα (σε μορφή πηγαίου κώδικα) παράγουν το ίδιο ακριβώς αποτέλεσμα.

Πολυπλοκότητα (complexity)

- Τι μπορούν να κάνουν οι υπολογιστές αποδοτικά; (efficiently)
- Αντιμετωπίσιμα προβλήματα (tractable problems)
- Μη αντιμετωπίσιμα (δυσεπίλυτα) προβλήματα (intractable problems)

Αφηρημένα μοντέλα υπολογισμών

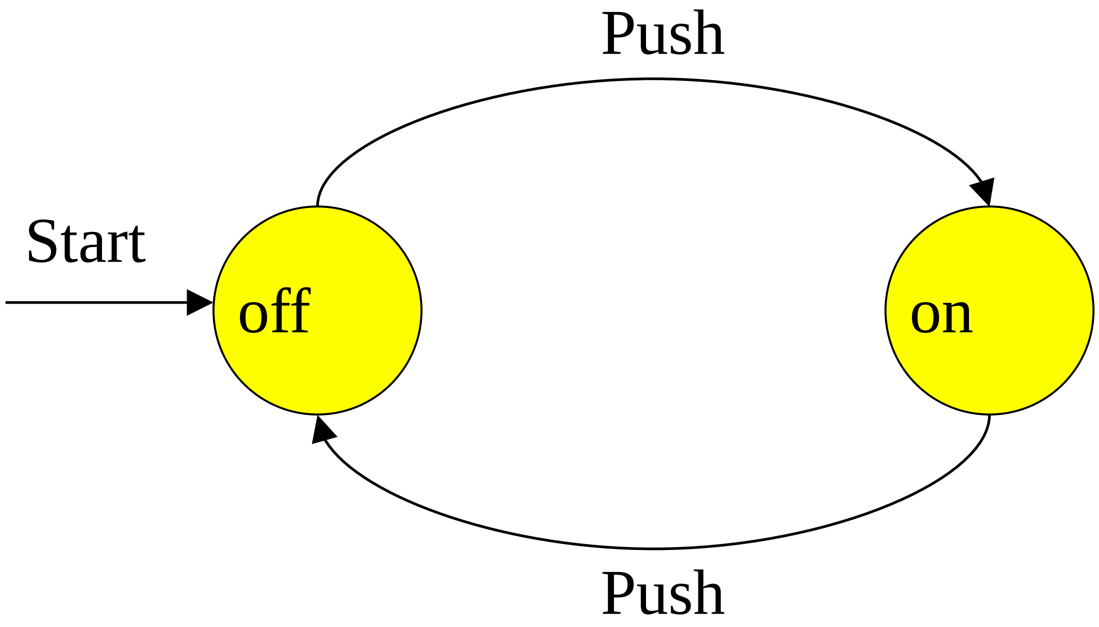
- Πεπερασμένα αυτόματα
  – Finite Automata, FA
- Αυτόματα στοίβας
  – PushDown Automata, PDA
- Μηχανή Turing
  – Turing Machine, TM

Τι είναι τα αυτόματα

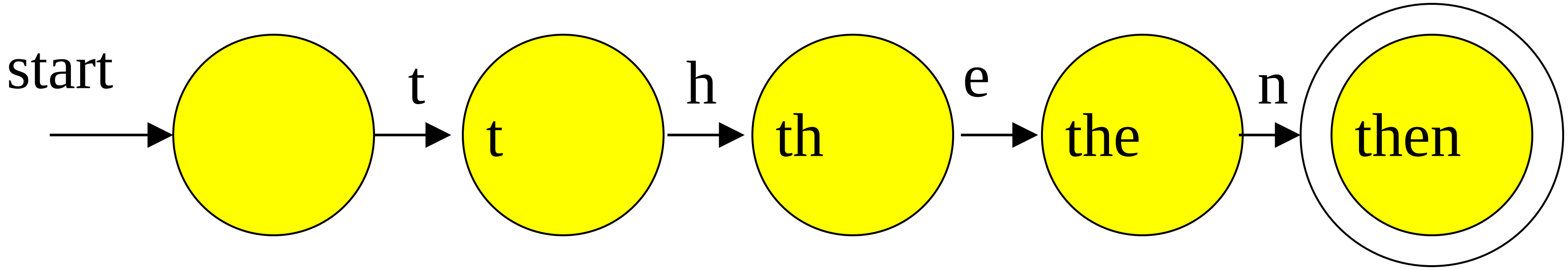
- Χρήσιμο μοντέλο μελέτης / ανάπτυξης για πολλά είδη hardware και software.
  - Λογισμικό σχεδίασης και ελέγχου ψηφιακών κυκλωμάτων
  - Λεξικοί αναλυτές (π.χ. για τη μετάφραση - compilation – προγραμμάτων)
  - Αναζήτηση συγκεκριμένων προτύπων φράσεων εντός κειμένου (π.χ. στο WEB)
  - Λογισμικό ελέγχου συστημάτων οποιουδήποτε τύπου που έχουν πεπερασμένο αριθμό εσωτερικών καταστάσεων, όπως π.χ. πρωτόκολλα τηλεπικοινωνίας, πρωτόκολλα ασφαλούς ανταλαγής δεδομένων κλπ.

Ένα απλό αυτόματο

- Διακόπτης on/off (switch)



Ένα λίγο πιο πολύπλοκο αυτόματο



Κεντρικό πρόβλημα

- Για ποιες ακολουθίες χαρακτήρων εισόδου (strings) ένα αυτόματο καταλήγει σε μια τελική κατάσταση;
- Για ένα αυτόματο M, το σύνολο εκείνων των ακολουθιών εισόδου που οδηγούν σε τελικές καταστάσεις ονομάζεται γλώσσα του αυτομάτου, L(M).

Αλφάβητα και Γλώσσες

Αλφάβητα (Alphabets)

- Ένα αλφάβητο είναι ένα πεπερασμένο, μη κενό σύνολο συμβόλων. Συνήθως χρησιμοποιούμε το γράμμα Σ για ένα αλφάβητο. Παραδείγματα:
  - $\Sigma = \{0,1\}$
  - $\Sigma = \{a,b,c, ..., z\}$

Λέξεις (Strings ή Words)

- Μία λέξη είναι μια πεπερασμένη ακολουθία συμβόλων επιλεγμένων από κάποιο αλφάβητο.
  - Σ = {0,1}
    - w=01101
- Μήκος λέξης είναι το πλήθος των συμβόλων της:
  - |w| = |01101| = 5
- Ως κενή λέξη ορίζεται η λέξη με μηδενικό μήκος και συμβολίζεται με *ε.*
  - |*ε*| = *0*

(Συμβάσεις σημειογραφίας)

- Συνηθίζεται να χρησιμοποιούμε:
  - τα πεζά γράμματα στην αρχή του αλφάβητου για να δηλώνουμε απλά σύμβολα.
    - a, b, c, ...
  - τα πεζά γράμματα στο τέλος του αλφάβητου για να δηλώνουμε λέξεις.
    - w, x, y, z, ...

Δυνάμεις ενός αλφάβητου

- Σ^k: Το σύνολο όλων των λέξεων μήκους k, αποτελούμενες από σύμβολα του Σ.
- Παράδειγμα: Έστω Σ={0,1}, τότε:
  - $\Sigma^0=\{\varepsilon\}$
  - $\Sigma^1=\{0, 1\}$
  - $\Sigma^2=\{00,01,10,11\}$
- Ειδικές περιπτώσεις:
- $\Sigma^*=\Sigma^0 \cup \Sigma^1 \cup \Sigma^2 \cup \Sigma^3 \cup \ldots$
- $\Sigma^+= \Sigma^1 \cup \Sigma^2 \cup \Sigma^3 \cup \ldots$

Παράθεση λέξεων (concatenation)

- Έστω x και y δύο λέξεις. Ορίζουμε ως xy τη λέξη που προκύπτει από την παράθεση των δύο επιμέρους λέξεων.
- Παράδειγμα:
  - x=01101
  - y=110
  - xy=01101110
  - yx=11001101

Ιδιότητες παράθεσης

- $|xy|=|x|+|y|$
- $x(yz)=(xy)z$
- Αν $w=xyz$, τότε:
  - η x λέγεται πρόθεμα (prefix) της w
  - η z λέγεται κατάληξη (suffix) της w
  - η y λέγεται υπολέξη της w
- Ορίζουμε ως w^k την παράθεση $w^{k-1}w$
  - Εάν $w=01$ τότε $w^3=010101$

Ανάστροφη λέξης

- Η ανάστροφη μιας λέξης w συμβολίζεται με w^R και είναι η λέξη w διαβασμένη από το τέλος προς την αρχή.
  - Εάν w=0011010 τότε w^R=0101100
- Ιδιότητες:
  - $(xy)^R=y^Rx^R$
    - Έστω x=abc και y=de, τότε:
      xy=abcde
      $(xy)^R$=edcba
      x^R=cba
      y^R=ed
      y^Rx^R=edcba
  - $(w^R)^R=w$

Γλώσσες

- Μια γλώσσα L ενός αλφαβήτου Σ ορίζεται ως ένα υποσύνολο του Σ^*.
  - $L \subseteq \Sigma^*$
- Μια γλώσσα μπορεί να έχει άπειρες λέξεις ή και καμία.
- Παραδείγματα γλωσσσών για το αλφάβητο Σ={0,1}:
  - Το σύνολο των λέξεων που αποτελούνται από n μηδενικά ακολουθούμενα από n μονάδες:
    {ε, 01, 0011, 000111, ...}
  - Το σύνολο των λέξεων με ίσους αριθμούς μηδενικών και μονάδων:
    {ε, 01, 10, 0011, 0101, 0110, ...}
  - Το σύνολο των δυαδικών αριθμών που αντιστοιχούν σε πρώτους αριθμούς:
    {10, 11, 101, 111, 1011, ...}

Περιγραφή γλωσσών ως σύνολα

- Ένας γενικός τρόπος ορισμού μιας γλώσσας είναι μέσω της περιγραφής των ιδιοτήτων των λέξεων της, με συμβολισμό συνόλων:
- {w | ιδιότητες της λέξης w}
  - {w | η w αποτελείται από ίσο αριθμό 0 και 1}
  - {w | η w είναι η δυαδική μορφή ενός πρώτου αριθμού}
  - {w | η w είναι ένα συντακτικά σωστό πρόγραμμα της C}
  - $\{0^n1^n \mid n \geq 1\}$
  - $\{0^i1^j \mid 0 \leq i \leq j\}$
- Ο παραπάνω συμβολισμός συνηθίζεται στις περιπτώσεις γλωσσών με άπειρο πλήθος λέξεων.

Πράξεις με γλώσσες (1/3)

- Εφόσον οι γλώσσες είναι σύνολα, μπορούν να συνδυαστούν με πράξεις επί συνόλων:
  - Ένωση, $L_1 \cup L_2$
    - Το σύνολο όλων των λέξεων που ανήκουν είτε στην L_1 είτε στην L_2.
    - $L_1 \cup L_2 = \{ w \mid w \in L_1 \vee w \in L_2 \}$
  - Τομή, $L_1 \cap L_2$
    - Το σύνολο των λέξεων που ανήκουν και στην L_1 και στην L_2.
    - $L_1 \cap L_2 = \{ w \mid w \in L_1 \wedge w \in L_2 \}$

Πράξεις με γλώσσες (2/3)

– Διαφορά, $L_1 - L_2$

- Το σύνολο των λέξεων που ανήκουν στην L_1 και δεν ανήκουν στην στην L_2.
- $L_1 - L_2 = \{ w \mid w \in L_1 \wedge w \notin L_2 \}$

– Συμπλήρωμα, $\overline{L_1} = \Sigma^* - L_1$

- Το σύνολο των λέξεων του Σ* που δεν ανήκουν στην L_1.
- $\overline{L_1} = \{ w \mid w \in \Sigma^* \wedge w \notin L_1 \}$

Πράξεις με γλώσσες (3/3)

– Παράθεση, L_1L_2
  - Το σύνολο των λέξεων που σχηματίζονται από την παράθεση μιας λέξης από την L_1 και μιας λέξης από την L_2.
  - $L_1L_2 = \{ w \mid w=xy, x \in L_1 \wedge y \in L_2 \}$

– Κλειστότητα ή Kleene star, L*
  - Το σύνολο των λέξεων που σχηματίζονται από την παράθεση οποιουδήποτε πλήθους λέξεων της L.
  - $L^*=\{w\in\Sigma^*: w=w_1w_2...w_k, k\geq 0, w_1, w_2, ..., w_k \in L\}$
  - $(L^*)^*=L^*$

Προβλήματα

- Στη θεωρία αυτομάτων ένα πρόβλημα είναι η ερώτηση εάν μια λέξη ανήκει σε μια γλώσσα ή όχι. Ειδικότερα:
- **Δοθείσης μιας λέξης w $\in$ Σ*, αποφάσισε εάν η w ανήκει στη γλώσσα L.**
- Όλα τα προβλήματα μπορούν να αναχθούν σε ισοδύναμα (απλούστερα) προβλήματα της παραπάνω μορφής με ίδια χαρακτηριστικά υπολογισιμότητας και πολυπλοκότητας.

Κανονικές εκφράσεις (regular expressions)

Τυποποιημένοι τρόποι περιγραφής γλωσσών

- Κανονικές εκφράσεις
  – περιορισμένες δυνατότητες
- Γραμματικές ανεξάρτητες από τα συμφραζόμενα.
  – ισχυρές
- Γραμματικές εξαρτημένες από τα συμφραζόμενα.
  – ισχυρότερες

Κανονικές εκφράσεις (regular expressions)

- Οι κανονικές εκφράσεις ενός αλφαβήτου Σ είναι οι συμβολοσειρές του αλφαβήτου $\Sigma \cup \{\ )\ ,\ (\ ,\ \varepsilon\ ,\ +\ ,\ *\ \}$
  για τις οποίες ισχύουν τα παρακάτω:
  - $\emptyset$, ε και κάθε στοιχείο του Σ είναι κανονική έκφραση.
  - Αν α και β είναι κανονικές εκφράσεις, και η (αβ) είναι επίσης.
  - Αν α και β είναι κανονικές εκφράσεις, και η α+β είναι επίσης.
  - Αν α είναι κανονική έκφραση, και η α* είναι επίσης
  - Τίποτα δεν είναι κανονική έκφραση εκτός από τα παραπάνω.

Παραδείγματα

- (a+b)*a : Περιγράφει λέξεις που τελειώνουν σε a.
- aa*+bb* : Περιγράφει λέξεις που αποτελούνται είτε μόνο από a (ένα ή περισσότερα) είτε μόνο από b (ένα ή περισσότερα).
- c*(a+bc*)*: Περιγράφει λέξεις στις οποίες δεν εμφανίζεται η υπολέξη ac.
- 0*+0*(1+11)(00*(1+11))*0*: Περιγράφει το σύνολο των λέξεων από 0 και 1, οι οποίες δεν περιέχουν την υπολέξη 111.

Τέλος Ενότητας

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