ΘΕΩΡΙΑ ΥΠΟΛΟΓΙΣΜΩΝ ΚΑΙ ΑΥΤΟΜΑΤΩΝ

Ενότητα 5: Μη κανονικές γλώσσες

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

Μη κανονικές γλώσσες

Μη κανονικές γλώσσες

- Είναι οι γλώσσες εκείνες για τις οποίες δεν υπάρχει πεπερασμένο αυτόματο (ντετερμινιστικό ή μη) που να τις δέχεται.
- Δεν υπάρχει γενική μέθοδος απόδειξης ότι μία γλώσσα δεν είναι κανονική.
- Για να αποδείξουμε ότι μια γλώσσα δεν είναι κανονική, χρησιμοποιούμε κάποιες βασικές ιδιότητες που ισχύουν για τις κανονικές γλώσσες.

Ιδιότητα 1

- Καθώς διαβάζουμε μια λέξη (οσοδήποτε μεγάλου μήκους) από αριστερά προς τα δεξιά, το ποσό της μνήμης που χρειάζεται για να αποφασιστεί στο τέλος αν η λέξη ανήκει ή όχι στη γλώσσα πρέπει να είναι φραγμένο και εξαρτώμενο από τη γλώσσα (δηλ.: να μην εξαρτάται από τη συγκεκριμένη λέξη).
  - π.χ. κάτι τέτοιο δεν ισχύει για τις λέξεις της γλώσσας $\{a^n b^n: n>0\}$

Ιδιότητα 2

- Οι κανονικές γλώσσες με άπειρο αριθμό λέξεων (προφανώς μη φραγμένου μήκους) έχουν άπειρα υποσύνολα με κάποια απλή επαναληπτική δομή που προκύπτει από άστρο στην αντίστοιχη κανονική έκφραση ή από κύκλο στο διάγραμμα καταστάσεων ενός πεπερασμένου αυτόματου.
  - Η γλώσσα $\{a^n$: ο n είναι πρώτος αριθμός$\}$ δεν είναι κανονική.

Θεώρημα άντλησης

- Έστω L μια άπειρη κανονική γλώσσα. Τότε υπάρχουν λέξεις x,y,z τέτοιες ώστε $y \neq \varepsilon$ και $xy^{n}z \in L$ για κάθε n ≥ 0.
  – Με άλλα λόγια, υπάρχουν ορισμένα σημεία μέσα σε ορισμένες λέξεις (και σίγουρα στις "πολύ μεγάλες" λέξεις) στα οποία μπορεί να εισαχθεί επανειλημένα μια υπολέξη, χωρίς αυτό να επηρεάσει το γεγονός ότι η αρχική λέξη γίνεται δεκτή.

Παράδειγμα 1

- Η γλώσσα L={$a^n b^n$: n≥0}δεν είναι κανονική.
  - Προσπαθούμε να φτιάξουμε λέξεις της μορφής xy^nz που να ανήκουν στη γλώσσα για κάθε *n*.
  - Υπάρχουν τρεις περιπτώσεις:
    - Το *y* να αποτελείται μόνο από *a*.
    - Το *y* να αποτελείται μόνο από *b*.
    - Το *y* να αποτελείται από μερικά *a* και στη συνέχεια από μερικά *b*.
  - Και στις τρεις περιπτώσεις μπορεί εύκολα να φανεί ότι δεν είναι δυνατόν λέξεις αυτής της μορφής να ανήκουν στη γλώσσα *L* για κάθε *n*, είτε γιατί δεν θα είναι ίσος ο αριθμός των *a* και των *b* (πρώτη και δεύτερη περίπτωση), είτε γιατί θα υπάρχουν 'ανακατωμένα' *a* και *b* (τρίτη περίπτωση).

Παράδειγμα 2

- Η γλώσσα $L=\{a^n$: Ο n είναι πρώτος αριθμός$\}$ δεν είναι κανονική.
  - Σύμφωνα με το θεώρημα, θα πρέπει να υπάρχουν $x=a^p$, $y=a^q$, $z=a^r$, $p,r \geq 0$, $q>0$, τέτοια ώστε κάθε λέξη της μορφής xy^nz να ανήκει στη γλώσσα.
  - Δηλαδή θα πρέπει η λέξη a^{p+nq+r} να ανήκει στη γλώσσα για κάθε $n\geq0$, δηλαδή ο αριθμός $p+nq+r$ πρέπει να είναι πρώτος για κάθε $n\geq0$.
  - Όμως, έστω $n=p+2q+r+2$, τότε $p+nq+r=(q+1)(p+2q+r)$

Γραμματικές χωρίς συμφραζόμενα

Γραμματικές χωρίς συμφραζόμενα

- Μια γραμματική χωρίς συμφραζόμενα *G* ορίζεται ως μια τετράδα (*V*, *Σ*, *R*, *S*) όπου:
  - *V* είναι ένα **αλφάβητο**
  - *Σ* είναι το σύνολο των **τερματικών συμβόλων**, $\Sigma \subseteq V$.
  - *R*, το **σύνολο των κανόνων**, είναι μια **σχέση** από το (*V*-*Σ*) στο *V**.
  - *S*, το **αρχικό σύμβολο**, είναι στοιχείο του *V*-*Σ*.

Εφαρμογή κανόνων

- Τα στοιχεία του συνόλου V-Σ ονομάζονται **μη-τερματικά**.
- Για κάθε $A \in V$-Σ και $u \in V^*$, εφόσον ισχύει $R(A,u)$ γράφουμε $A \rightarrow u$.
- Για δύο συμβολοσειρες $v,w \in V^*$, λέμε ότι η w **παράγεται** από την v σε ένα βήμα, και συμβολίζουμε με $v \Rightarrow w$ εάν:
  - $v=xAy, \quad x,y \in V^*$
  - $w=xuy$
  - Υπάρχει ο κανόνας $A \rightarrow u$ στο σύνολο κανόνων.
- Οι γραμματικές λέγονται "χωρίς συμφραζόμενα" επειδή στο αριστερό μέρος των κανόνων τους έχουν ένα μη-τερματικό σύμβολο και μόνο.

Παραγωγή των λέξεων της γλώσσας

- Συμβολίζουμε με $v \overset{*}{\Rightarrow} w$ το γεγονός ότι η λέξη *w* **παράγεται** από τη λέξη *v* με εφαρμογή πολλών κανόνων από το σύνολο κανόνων *R*.
- Με βάση τα παραπάνω, η γλώσσα *L*(*G*) που **παράγεται** από μια γραμματική χωρίς συμφραζόμενα *G* ορίζεται ως εξής:
  - $L(G)=\{w \in \Sigma^*: S \underset{*}{\Rightarrow} w\}$

Παράδειγμα 1 (1/2)

- Έστω η γραμματική χωρίς συμφραζόμενα $G = (V, \Sigma, R, S)$ όπου:
  - $V=\{S,a,b\}$
  - $\Sigma=\{a,b\}$
  - Η σχέση R αποτελείται από τους κανόνες $S\rightarrow aSb$ και $S\rightarrow \varepsilon$.
- Μια παραγωγή της γλώσσας αυτής είναι η:
  - $S\Rightarrow aSb\Rightarrow aaSbb\Rightarrow aabb$
- Είναι εύκολο να δει κανείς ότι $L(G)=\{a^n b^n: n\geq 0\}$

Παράδειγμα 1 (2/2)

- Βλέπουμε από το παράδειγμα ότι κάποιες γλώσσες χωρίς συμφραζόμενα δεν είναι κανονικές.
- Ωστόσο ισχύει ότι όλες οι κανονικές γλώσσες μπορούν να προκύψουν από γραμματικές χωρίς συμφραζόμενα.

Παράδειγμα 2

- Έστω η γραμματική $G=(V,\Sigma,R,S)$ με:
  - $V=\{S,\Pi,P,O,E,A\}\cup\Sigma$
  - Σ={ο, είναι, Νίκος, καλός, φοιτητής}
  - $R=\{S\to\Pi P\Pi$, $\Pi\to AEO$, $\Pi\to EO$, $\Pi\to AO$, $\Pi\to O$, $A\to$ο, $P\to$είναι, $O\to$Νίκος, $E\to$καλός, $O\to$φοιτητής}
- Έγκυρες "λέξεις" που παράγονται είναι οι:
  - ο Νίκος είναι καλός φοιτητής
  - ο Νίκος είναι φοιτητής
  - ο καλός Νίκος είναι φοιτητής
  - κλπ

Π=Προσδιορισμός
Ρ=Ρήμα
Ο=Ουσιαστικό
Ε=Επίθετο
Α=Άρθρο

Παράδειγμα 3 (1/2)

- Έστω η γραμματική $G=(V,\Sigma,R,E)$ με:
  - $V=\{E\}\cup\Sigma$
  - $\Sigma=\{\ x,\ y,\ +,\ \cdot\ ,\ (,\ )\ \ \}$
  - $R=\{E\rightarrow(E),\ E\rightarrow E+E,\ E\rightarrow E\cdot E,\ E\rightarrow x,\ E\rightarrow y\}$
- Έγκυρες "λέξεις" που παράγονται είναι οι:
  - x+y
  - x+x·y
  - (x·y+x·x) ·(x+y· (x+y·y))
  - κλπ

Ε=Έκφραση

Παράδειγμα 3 (2/2)

- $(x \cdot y + x \cdot x) \cdot (x + y \cdot (x + y \cdot y))$

  $E \Rightarrow E \cdot E$

  $\Rightarrow (E) \cdot E$

  $\Rightarrow (E) \cdot (E)$

  $\Rightarrow (E+E) \cdot (E)$

  $\Rightarrow (E \cdot E+E) \cdot (E)$

  $\Rightarrow (E \cdot E+E \cdot E) \cdot (E)$

  $\Rightarrow (E \cdot E+E \cdot E) \cdot (E+E)$

  $\Rightarrow (E \cdot E+E \cdot E) \cdot (E+E \cdot E)$

  $\Rightarrow (E \cdot E+E \cdot E) \cdot (E+E \cdot (E))$

  $\Rightarrow (E \cdot E+E \cdot E) \cdot (E+E \cdot (E+E))$

  $\Rightarrow (E \cdot E+E \cdot E) \cdot (E+E \cdot (E+E \cdot E))$

  ……….

  ……….

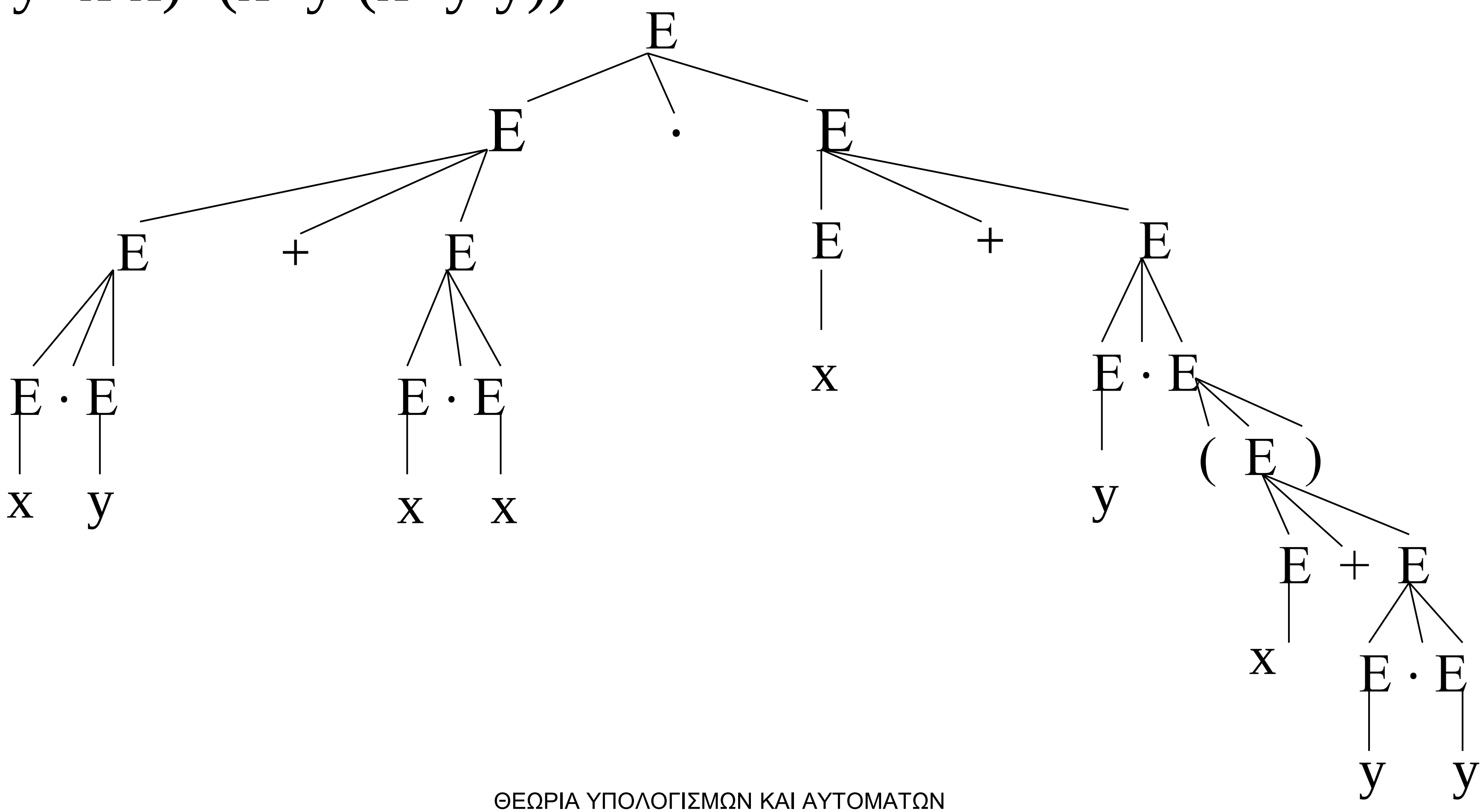
  $(x \cdot y + x \cdot x) \cdot (x + y \cdot (x + y \cdot y))$

Συντακτικό δέντρο (Parse Tree)

- Πρόκειται για μια διαγραμματική αναπαράσταση, με μορφή δέντρου, του τρόπου που παράγεται μια λέξη εφαρμόζοντας τους κανόνες της γραμματικής.
- Κάθε κόμβος του αντιστοιχεί σε ένα σύμβολο.
- Τα παιδιά ενός κόμβου προκύπτουν από την εφαρμογή ενός κανόνα στο σύμβολο του κόμβου.
- Τα φύλλα του δέντρου είναι τερματικά σύμβολα.
- Η λέξη προκύπτει παραθέτοντας όλα τα φύλλα του δέντρου, διαβάζοντάς τα από αριστερά προς τα δεξιά.

Παράδειγμα συντακτικού δέντρου (για την παραγωγή της διαφάνειας 19)

$(x \cdot y + x \cdot x) \cdot (x + y \cdot (x + y \cdot y))$



Εναλλακτικές Παραγωγές

- Είναι δυνατόν μια λέξη να παράγεται με περισσότερους από έναν τρόπους από τους κανόνες μιας γραμματικής (και άρα να έχει εναλλακτικά συντακτικά δέντρα).
- Για παράδειγμα, η λέξη x+x·y μπορεί να παραχθεί από τη γραμματική του παραδείγματος 3 ως εξής:
  - $E \Rightarrow E+E \Rightarrow E+E \cdot E \Rightarrow \ldots \Rightarrow x+x \cdot y$
- αλλά και ως εξής:
  - $E \Rightarrow E \cdot E \Rightarrow E+E \cdot E \Rightarrow \ldots \Rightarrow x+x \cdot y$
- Οι γραμματικές που επιτρέπουν εναλλακτικές παραγωγές λέγονται διφορούμενες (ambiguous).

Αριστερότερη Παραγωγή (leftmost derivation)

- Ως αριστερότερη παραγωγή μιας λέξης ορίζεται αυτή η οποία σε κάθε βήμα αντικαθιστά το αριστερότερο μη-τερματικό σύμβολο της λέξης.
- Στο παράδειγμα της λέξης x+x·y, αριστερότερη παραγωγή είναι η:
  - $E \Rightarrow E \cdot E \Rightarrow E+E \cdot E \Rightarrow x+E \cdot E \Rightarrow x+x \cdot E \Rightarrow x+x \cdot y$
- Παρόμοια ορίζεται η δεξιότερη παραγωγή (rightmost derivation).
  - Για κάθε λέξη, εφόσον υπάρχει μια παραγωγή, τότε υπάρχουν σίγουρα μια αριστερότερη και μια δεξιότερη παραγωγή.

Κανονικές Γραμματικές

- Μια γραμματική λέγεται **κανονική**, εάν στο δεξιό μέρος κάθε κανόνα της υπάρχει το πολύ ένα μη-τερματικό σύμβολο, και αυτό είναι **πάντα στην τελευταία θέση**.
- Τα σύνολα των γλωσσών των κανονικών γραμματικών και των γλωσσών των κανονικών εκφράσεων (ή των πεπερασμένων αυτομάτων) ταυτίζονται.

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