



ΗΥ - 280 Θεωρία Υπολογισμού

Διάλεξη 8

Κωνσταντίνος Βάρσος

Τμήμα Επιστήμης Υπολογιστών – Πανεπιστήμιο Κρήτης

Χειμερινό εξάμηνο 2025 - 26

Γλώσσες και Αυτόματα

Τι είδαμε στην Διάλεξη 7

- Γραμματικές.
- Δεξιογραμμικές και Αριστερογραμμικές.
- Ασυμφραστικές γραμματικές και γλώσσες.
- Μετατροπή DFA σε γραμματική και το αντίστροφο.

Γλώσσες και Γραμματικές χωρίς Συμφραζόμενα

Η παραγωγή μιας συμβολοσειρά από μια γραμματική μας δίνει αρκετή ευελιξία στο ποιες μεταβλητές να διαλέξουμε. Συχνά εφαρμόζουμε περιορισμούς ως προς αυτές τις επιλογές.

■ **αριστερότατη** παραγωγή, αντικαθιστούμε την αριστερότερη μεταβλητή που εμφανίζεται στην τρέχουσα κατάσταση της παραγωγής.

$$(\alpha_1 A \alpha_2 \Rightarrow \alpha_1 \gamma \alpha_2, \text{ για } A \rightarrow \gamma \in R) \iff \alpha_1 \in \Sigma^*$$

Συμβολίζουμε με $\Rightarrow_\ell$. Επαναλαμβάνουμε μέχρι να τερματιστεί η παραγωγή.

■ **δεξιότατη** παραγωγή, αντικαθιστούμε την δεξιότερη μεταβλητή που εμφανίζεται στην τρέχουσα κατάσταση της παραγωγής.

$$(\alpha_1 A \alpha_2 \Rightarrow \alpha_1 \gamma \alpha_2, \text{ για } A \rightarrow \gamma \in R) \iff \alpha_2 \in \Sigma^*$$

Συμβολίζουμε με $\Rightarrow_r$. Επαναλαμβάνουμε μέχρι να τερματιστεί η παραγωγή.

Γλώσσες και Γραμματικές χωρίς Συμφραζόμενα

π.χ., Έστω η γραμματική με τις ακόλουθες παραγωγές,

$$\begin{aligned} S &\rightarrow X \mid S + S \mid S * S \mid (S), \\ X &\rightarrow a \mid b \mid Xa \mid Xb \mid X0 \mid X1. \end{aligned}$$

Θα προσπαθήσουμε να παράξουμε την συμβολοσειρά $w = a0 + ab * (a + b1)$.

Γενική παραγωγή,

$$\begin{aligned} S &\Rightarrow S + S \Rightarrow S + S * S \Rightarrow S + S * (S) \Rightarrow X + S * (S) \Rightarrow X0 + S * (S) \\ &\Rightarrow X0 + S * (S) \Rightarrow a0 + S * (S) \Rightarrow X0 + X * (S) \Rightarrow a0 + Xb * (S) \Rightarrow a0 + ab * (S) \\ &\Rightarrow a0 + ab * (S + S) \Rightarrow a0 + ab * (X + S) \Rightarrow a0 + ab * (X + X) \\ &\Rightarrow a0 + ab * (a + X) \Rightarrow a0 + ab * (a + X1) \Rightarrow a0 + ab * (a + b1). \end{aligned}$$

Γλώσσες και Γραμματικές χωρίς Συμφραζόμενα

π.χ., (συνέχεια) Έστω η γραμματική με τις ακόλουθες παραγωγές,

$$\begin{aligned} S &\rightarrow X \mid S + S \mid S * S \mid (S), \\ X &\rightarrow a \mid b \mid Xa \mid Xb \mid X0 \mid X1. \end{aligned}$$

Θα προσπαθήσουμε να παράξουμε την συμβολοσειρά $w = a0 + ab * (a + b1)$.

Αριστερότατη παραγωγή,

$$\begin{aligned} S &\xRightarrow{\ell} S + S \xRightarrow{\ell} X + S \xRightarrow{\ell} X0 + S \xRightarrow{\ell} a0 + S \xRightarrow{\ell} a0 + S * S \\ &\xRightarrow{\ell} a0 + X * S \xRightarrow{\ell} a0 + Xb * S \xRightarrow{\ell} a0 + ab * S \xRightarrow{\ell} a0 + ab * (S) \xRightarrow{\ell} a0 + ab * (S + S) \\ &\xRightarrow{\ell} a0 + ab * (X + S) \xRightarrow{\ell} a0 + ab * (a + S) \xRightarrow{\ell} a0 + ab * (a + X) \\ &\xRightarrow{\ell} a0 + ab * (a + X1) \xRightarrow{\ell} a0 + ab * (a + b1). \end{aligned}$$

Γλώσσες και Γραμματικές χωρίς Συμφραζόμενα

π.χ., (συνέχεια) Έστω η γραμματική με τις ακόλουθες παραγωγές,

$$\begin{aligned} S &\rightarrow X \mid S + S \mid S * S \mid (S), \\ X &\rightarrow a \mid b \mid Xa \mid Xb \mid X0 \mid X1. \end{aligned}$$

Θα προσπαθήσουμε να παράξουμε την συμβολοσειρά $w = a0 + ab * (a + b1)$.

Δεξιότερη παραγωγή,

$$\begin{aligned} S &\Rightarrow_r S + S \Rightarrow_r S + S * S \Rightarrow_r S + S * (S) \Rightarrow_r S + S * (S + S) \Rightarrow_r S + S * (S + X) \\ &\Rightarrow_r S + S * (S + X1) \Rightarrow_r S + S * (S + b1) \Rightarrow_r S + S * (a + b1) \Rightarrow_r S + X * (a + b1) \\ &\Rightarrow_r S + Xb * (a + b1) \Rightarrow_r S + ab * (a + b1) \Rightarrow_r X + ab * (a + b1) \\ &\Rightarrow_r X0 + ab * (a + b1) \Rightarrow_r a0 + ab * (a + b1). \end{aligned}$$

□

Γραμματικές – Συντακτικό δέντρο

Ένα **συντακτικό δέντρο** (σ.δ., parse tree) απεικονίζει τον τρόπο παραγωγής μιας συμβολοσειράς από μια δοσμένη γραμματική $G = (V, \Sigma, R, S)$.

Τα χαρακτηριστικά ενός συντακτικού δέντρου είναι τα εξής:

- Η ρίζα επιγράφεται με το αρχικό σύμβολο S .
- Κάθε (εσωτερικός) κόμβος επιγράφεται με ένα σύμβολο του V .
- Κάθε φύλλο επιγράφεται με στοιχείο του Σ ή ε .
- Με παράθεση των επιγραφών των φύλλων από αριστερά προς τα δεξιά παίρνουμε την παραγόμενη συμβολοσειρά του συντακτικού δέντρου.

■ Αν ένας κόμβος έχει ετικέτα $A \in V$ και τα παιδιά του έχουν ετικέτες $w_1, \dots, w_k \in V \cup \Sigma \in \{\varepsilon\}$, τότε $A \rightarrow w_1, \dots, w_k \in R$.

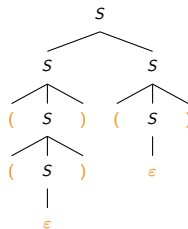
■ Αν ένα φύλλο έχει ετικέτα ε τότε είναι το μοναδικό παιδί.

Γραμματικές – Συντακτικό δέντρο

π.χ., Έστω $G = (V, \Sigma, R, S)$ με $V = \{S\}$, $\Sigma = \{ (,) \}$ και κανόνες παραγωγής

$$R : S \rightarrow SS \mid (S) \mid \varepsilon$$

Το συντακτικό δέντρο είναι,



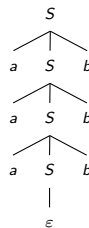
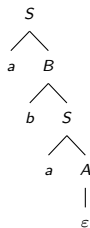
και η παραγόμενη συμβολοσειρά είναι $((()))()$.



Γραμματικές – Συντακτικό δέντρο

■ Έστω $G = (V, \Sigma, R, S)$ μια CFG, τότε $S \xRightarrow{*} w$ ανν υπάρχει συντακτικό δέντρο με φύλλωμα το w .

π.χ.,



Γραμματικές – Συντακτικό δέντρο

■ Συντακτικά δέντρα, παραγωγές, αριστερότερες παραγωγές, δεξιότερες παραγωγές είναι ισοδύναμες ως προς την παραγωγή της γλώσσας μιας CFG γραμματικής.

Θεώρημα. Έστω CFG $G = (V, \Sigma, R, S)$, $A \in V$ και $w \in \Sigma^*$. Τότε

$$A \xRightarrow{*} w \iff \exists \text{ σ.δ. με ρίζα } A \text{ το οποίο παράγει την } w.$$

Απόδειξη. Με επαγωγή ως προς το μήκος της w .

[Βάση επαγωγής] Έστω ότι $A \xRightarrow{*} w$ παραγωγή μήκους 0. Τότε το A είναι σ.δ. με ρίζα A το οποίο παράγει A .

[Επαγωγική υπόθεση] Έστω ότι η δήλωση ισχύει για όλες τις παραγωγές μήκους έως και $k - 1$.

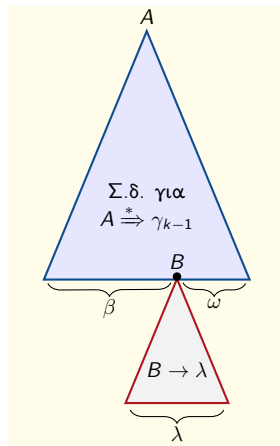
Έστω η παραγωγή α από A σε k βήματα, $A = \gamma_1 \Rightarrow \gamma_2 \Rightarrow \dots \Rightarrow \gamma_{k-1} \Rightarrow \gamma_k = \alpha$

Γραμματικές – Συντακτικό δέντρο

Απόδειξη. (Σχηματικά)

Συντακτικό δέντρο για την παραγωγή,

$$A \xRightarrow{*} \beta\lambda\omega = \alpha.$$



Γραμματικές – Συντακτικό δέντρο

Απόδειξη. (συνέχεια) Έστω ότι η παραγωγή α από A σε k βήματα,

$$A = \gamma_1 \Rightarrow \gamma_2 \Rightarrow \dots \Rightarrow \gamma_{k-1} \Rightarrow \gamma_k = \alpha$$

Στο τελευταίο βήμα πρέπει να εφαρμόσουμε ενός κανόνα παραγωγής.

Άρα, $\gamma_{k-1} = \beta B \omega$ και $\alpha = \beta \lambda \omega$ όπου

$$\beta, \omega \in (V \cup \Sigma)^*, \quad B \in V, \quad B \rightarrow \lambda.$$

Επεκτείνουμε το σ.δ. από τα $k - 1$ βήματα ως εξής,

Αν $\lambda = X_1 \dots X_n$ με $X_1, \dots, X_n \in V \cup \Sigma$ προσθέτουμε τα παιδιά $X_1, \dots, X_n$ στον κόμβο B .

□

Γραμματικές – Συντακτικό δέντρο

Θεώρημα. Έστω CFG $G = (V, \Sigma, R, S)$, $A \in V$ και $w \in \Sigma^*$. Τότε

$\exists$ σ.δ. με ρίζα A το οποίο παράγει την $w \iff \exists$ αριστερότατη παραγωγή τ.ω. $A \xRightarrow[\ell]{*} w$

Απόδειξη. Με επαγωγή ως προς το ύψος του δέντρου.

[Βάση επαγωγής] Το σ.δ. έχει ύψος $\mu = 0$. Τότε η A είναι η αριστερότατη παραγωγή σε μηδέν βήματα.

[Επαγωγική υπόθεση] Έστω ότι η δήλωση ισχύει για όλα τα σ.δ. ύψους έως και $\mu - 1$.

Παίρνουμε την ρίζα και το k -οστό παιδί. Αυτό αντιστοιχεί στον κανόνα παραγωγής,

$$A \rightarrow X_1 \dots X_k.$$

Γραμματικές – Συντακτικό δέντρο

Απόδειξη. (συνέχεια) Με επαγωγή ως προς το ύψος του δέντρου.

■ Αν X φύλλο, τότε η παραγωγή του υποδέντρου με ρίζα το X_i είναι $w_i = X_i$. Τετριμμένα παίρνουμε, $X_i \xRightarrow[\ell]{*} w_i$.

■ Αν X δεν είναι φύλλο, τότε η w_i παράγεται από το υποδέντρο με ρίζα το X_i βάθους το πολύ $\mu - 1$. Έπειτα, από επαγωγική υπόθεση παίρνουμε $X_i \xRightarrow[\ell]{*} w_i$.

Τότε το ακόλουθο είναι μια αριστερότατη παραγωγή του α από το A .

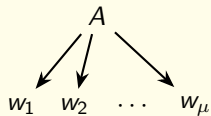
$$A \Rightarrow X_1 X_2 \dots X_k \xRightarrow[\ell]{*} w_1 X_2 \dots X_k \xRightarrow[\ell]{*} w_1 w_2 X_3 \dots X_k \xRightarrow[\ell]{*} \dots \xRightarrow[\ell]{*} w_1 \dots w_k.$$



Γραμματικές – Συντακτικό δέντρο

Απόδειξη. (Σχηματικά)

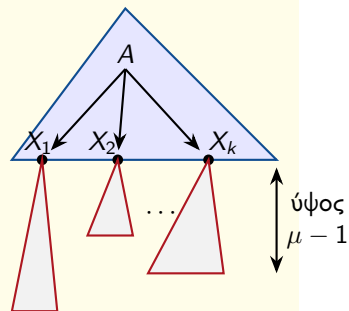
Βάση



$$\alpha = w_1 w_2 \cdots w_\mu$$

$$(A, \alpha) \equiv (A \rightarrow \alpha) \in R$$

Επαγωγή



Γραμματικές – Συντακτικό δέντρο

- Το προηγούμενο Θεώρημα μπορεί να διατυπωθεί και για δεξιότατες παραγωγές (η απόδειξη είναι παρόμοια, χρειάζεται μερικές αλλαγές)
- Τα προηγούμενα 2 Θεωρήματα εδραιώνουν τις εξής ισοδυναμίες,

$\exists$ σ.δ. με ρίζα A το οποίο παράγει την w

$$\iff A \xRightarrow{*} w$$

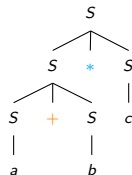
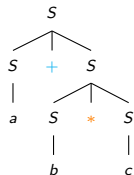
$$\iff A \xRightarrow[\ell]{*} w \quad [\text{Εξ ορισμού}]$$

$$(\iff A \xRightarrow[r]{*} w \quad [\text{Εξ ορισμού}])$$

Γραμματικές – Συντακτικό δέντρο

Ορισμός. Μια γραμματική G λέγεται **διφορούμενη** (αμφίσημη, ambiguous) αν υπάρχουν δύο συντακτικά δέντρα που έχουν ως φύλλωμα το ίδιο $w \in L(G)$.

π.χ.,



■ Μια CFG G είναι διφορούμενη αν μια συμβολοσειρά $w \in L(G)$ έχει τουλάχιστον δύο αριστερότατες (ή δεξιότατες) παραγωγές.

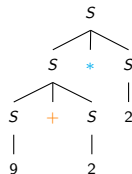
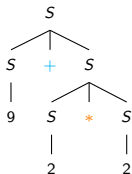
■ Η αμφισημία είναι μια ιδιότητα της γραμματικής. Ωστόσο, υπάρχουν και γλώσσες που είναι εγγενώς αμφίσημες.

Γραμματικές – Συντακτικό δέντρο

π.χ., Έστω CFG $G = (V = \{S\}, \Sigma = \{0, 1, \dots, 9, +, *\}, R, S)$ με κανόνες παραγωγής

$$R: S \rightarrow S + S \mid S * S \mid 0 \mid 1 \mid \dots \mid 9$$

Θα παράξουμε το $9 + 2 * 2$ χρήσει σ.δ.



□

■ Εφόσον υπάρχουν 2 σ.δ. ένας compiler δεν ξέρει αν θα πρέπει να χρησιμοποιήσει το 13 ή το 22.

■ Εδώ η αμφισημία μπορεί να λυθεί, π.χ., με την χρήση κανόνων προτεραιότητας πράξεων.

Γραμματικές – Συντακτικό δέντρο

π.χ., Έστω CFGς $G_1 = (V_1 = \{S\}, \Sigma = \{(\,,\,)\}, R_1, S)$, $G_2 = (V_2 = \{A, B\}, \Sigma = \{(\,,\,)\}, R_2, B)$ με κανόνες παραγωγής και $w = ()()()$.

$$R_1 : \{S \rightarrow SS \mid (S) \mid ()\}, \quad R_2 : \{B \rightarrow (AB \mid \varepsilon, \quad A \rightarrow) \mid ((AA)\}.$$

Γραμματικές – Συντακτικό δέντρο

π.χ., Έστω CFGs $G_1 = (V_1 = \{S\}, \Sigma = \{(\,,\,)\}, R_1, S)$, $G_2 = (V_2 = \{A, B\}, \Sigma = \{(\,,\,)\}, R_2, B)$ με κανόνες παραγωγής και $w = ()()()$.

$$R_1 : \{S \rightarrow SS \mid (S) \mid ()\}, \quad R_2 : \{B \rightarrow (AB \mid \varepsilon, \quad A \rightarrow) \mid ((AA)\}.$$

Η G_1 είναι διφορούμενη καθώς υπάρχουν 2 αριστερότατες παραγωγές για την $()()()$.

$$(i) \quad S \xRightarrow{\ell} SS \xRightarrow{\ell} ()S \xRightarrow{\ell} ()SS \xRightarrow{\ell} ()()S \xRightarrow{\ell} ()()()$$

$$(ii) \quad S \xRightarrow{\ell} SS \xRightarrow{\ell} SSS \xRightarrow{\ell} ()SS \xRightarrow{\ell} ()()S \xRightarrow{\ell} ()()().$$

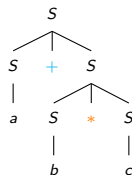
Η G_2 δεν είναι διφορούμενη καθώς υπάρχει ακριβώς ένας κανόνας σε κάθε στάδιο της παραγωγής.

$$B \xRightarrow{\ell} (AB \xRightarrow{\ell} ()B \xRightarrow{\ell} ()(AB \xRightarrow{\ell} ()()B \xRightarrow{\ell} ()()(AB \xRightarrow{\ell} ()()()B \xRightarrow{\ell} ()()().$$

□

Γραμματικές – Συντακτικό δέντρο

■ Η αμφισημία μπορεί να λυθεί εισάγοντας επιπλέον κανόνες παραγωγής.



Η γραμματική είναι η $G = (V = \{S\}, \Sigma = \{a, +, *\}, R, S)$ με κανόνες παραγωγής
 $R : S \rightarrow S + S \mid S * S \mid a$.

■ Επεκτείνουμε τους κανόνες ώστε να περιλαμβάνουν και αφαίρεση και διαίρεση. Η πρόσθεση και η αφαίρεση βρίσκονται στο ίδιο επίπεδο προτεραιότητας, ενώ ο πολλαπλασιασμός και η διαίρεση βρίσκονται σε διαφορετικό, ανώτερο επίπεδο.

Γραμματικές – Συντακτικό δέντρο

■ Από την γραμματική είναι η $G = (V = \{S\}, \Sigma = \{a, b, c, +, *\}, R, S)$ με κανόνες παραγωγής

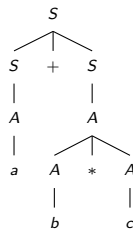
$$R : S \rightarrow S + S \mid S * S \mid a \mid b \mid c.$$

■ Στην $G = (V = \{S\}, \Sigma = \{a, b, c, +, *\}, R, S)$ με κανόνες παραγωγής

$$R : S \rightarrow S + S \mid S - S \mid A, \quad A \rightarrow A * A \mid A / A \mid (S) \mid a \mid b \mid c.$$

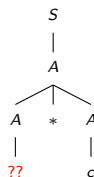
Γραμματικές – Συντακτικό δέντρο

■ Το συντακτικό δέντρο για την $w = a + b * c$ είναι



Γραμματικές – Συντακτικό δέντρο

■ Ας προσπαθήσουμε να εκφράσουμε την $w = a + b * c$ με τέτοιο τρόπο ώστε το c να πολλαπλασιάζεται στο άθροισμα.



! Δεν έχουμε κανόνα παραγωγής $A \rightarrow A + A$

■ Αυτό σημαίνει ότι η G έπαψε να είναι διφορούμενη; Δυστυχώς όχι. Εξετάστε το σ.δ. της $w = a - b - c$.

Γραμματικές – Συντακτικό δέντρο

Μερικά σχόλια

- Δεν είναι πάντοτε εύκολο να άρουμε την αμφισημία από μια CFG.
- Δεν είναι πάντοτε εφικτό να άρουμε την αμφισημία από μια CFG. Κάποιες CFL παράγονται αποκλειστικά από αμφίσημες CFG. [τις αναφέρουμε ως εγγενώς αμφίσημες, π.χ., $\{0^i 1^j 2^k \mid i = j \vee j = k\}$]
- Δεν υπάρχει αλγόριθμος που να καθορίζει αν μια CFG είναι αμφίσημη.
- Κοινές μορφές αμφισημίας μπορούν να αναγνωριστούν από parser generators, όπως YACC, bison κτλ.
- Παρότι δεν μπορούμε να αποφύγουμε πάντα την αμφισημία, μπορούμε να αποφύγουμε την περίπτωση των απείρως πολλών σ.δ. για μια συμβολοσειρά, Κανονική μορφή Chomsky.

Κανονική μορφή Chomsky

■ Η ιδέα της Κανονικής μορφής Chomsky (CNF) είναι να περιορίσουμε το μέγεθος του δεξιού μέλους κάθε παραγωγής, έτσι ώστε μια μεταβλητή να αντικαθίσταται από δύο μόνο άλλες μεταβλητές το πολύ σε κάθε βήμα.

π.χ., Η γραμματική

$$S \rightarrow SaS \mid b$$

μετατρέπεται αρχικά σε

$$S \rightarrow SAS \mid b, \quad A \rightarrow a$$

και, εν τέλει, σε:

$$S \rightarrow SB \mid b, \quad A \rightarrow a, \quad B \rightarrow AS.$$

Κανονική μορφή Chomsky

Ορισμός. Μια CFG λέγεται ότι είναι σε CNF αν ισχύουν τα εξής:

1 Κάθε κανόνας έχει τη μορφή

■ $A \rightarrow BC$, ή

■ $A \rightarrow a$,

όπου A, B, C είναι μεταβλητές και a είναι τερματικό.

2 Η αρχική μεταβλητή δεν εμφανίζεται στο δεξί μέλος κανενός κανόνα.

3 Ο κανόνας $S \rightarrow \varepsilon$ μπορεί να υπάρχει [ανάλογα με το αν η γλώσσα περιέχει το ε ή όχι].

Κανονική μορφή Chomsky

Μερικά σχόλια

- Δοθείσας γραμματικής μπορούν να υπάρξουν παραπάνω της μιάς CNF γραμματική.
- Η CNF γραμματική παράγει την ίδια γλώσσα με την CFG.
- Οι CNF χρησιμοποιούνται σε πολλούς αλγορίθμους ως προπαρασκευαστικό βήμα
- Για παραγόμενη συμβολοσειρά w μήκους n απαιτούνται $2n - 1$ παραγωγές στην CNF.
- Κάθε CNF χωρίς ε στην γλώσσα της έχει ισοδύναμη CNF.

Μετατροπή CFG σε CNF

■ Έστω $G = (V, \Sigma, R, S)$ μια CFG.

Αλγόριθμος μετατροπής της G σε CNF.

1. Αφαίρεση του S από το δεξί μέλος των κανόνων.
 - Αν η S εμφανίζεται στο δεξί μέλος κάποιου κανόνα, προσθέτουμε μια νέα αρχική μεταβλητή S_0 και τον κανόνα $S_0 \rightarrow S$.

Μετατροπή CFG σε CNF

2. Αφαίρεση των ϵ -παραγωγών.

- Επιλέγουμε μια ϵ -παραγωγή $A \rightarrow \epsilon$ και την αφαιρούμε [όπου η A δεν είναι η αρχική μεταβλητή].
- Για κάθε εμφάνιση της A στο δεξί μέλος κάποιου κανόνα, προσθέστε έναν κανόνα που διαγράφει αυτή την εμφάνιση της A . Αν η A εμφανίζεται πολλές φορές στο δεξί μέλος ενός κανόνα, τότε μπορεί να προστεθούν πολλοί κανόνες.
- Αν έχουμε τον κανόνα $R \rightarrow A$, τότε προσθέστε τον κανόνα $R \rightarrow \epsilon$, εκτός αν ο κανόνας $R \rightarrow \epsilon$ έχει αφαιρεθεί προηγουμένως.
- Επαναλαμβάνουμε μέχρι να μην υπάρχουν άλλες ϵ -παραγωγές. [εκτός ίσως από αυτούς που αφορούν την αρχική μεταβλητή]

Μετατροπή CFG σε CNF

π.χ., Έστω ότι μια γραμματική έχει τους ακόλουθους κανόνες:

$$A \rightarrow \varepsilon, \quad B \rightarrow uAv, \quad C \rightarrow xAyAz$$

Τότε η γραμματική που σχηματίζεται αφαιρώντας τον κανόνα $A \rightarrow \varepsilon$ θα έχει το αντίστοιχο σύνολο κανόνων:

$$B \rightarrow uAv \rightsquigarrow B \rightarrow uv, \quad [\text{προσθήκη κανόνα}]$$

$$C \rightarrow xAyAz \rightsquigarrow \begin{cases} C \rightarrow xyAz, & [\text{προσθήκη κανόνα με διαγραφή 1ου } A] \\ C \rightarrow xAyz, & [\text{προσθήκη κανόνα με διαγραφή 2ου } A] \\ C \rightarrow xyz, & [\text{προσθήκη κανόνα με διαγραφή και των δύο } A]. \end{cases}$$

Μετατροπή CFG σε CNF

3. Αφαιρούμε τους μοναδιαίους κανόνες.

■ Αφαιρούμε έναν κανόνα $A \rightarrow B$ και για όλους τους κανόνες της μορφής $B \rightarrow u$ προσθέτουμε τον κανόνα $A \rightarrow u$, εκτός αν ο κανόνας $A \rightarrow u$ είναι μοναδιαίος κανόνας που έχει ήδη αφαιρεθεί. Επαναλαμβάνουμε μέχρι να μην υπάρχουν άλλοι μοναδιαίοι κανόνες.

π.χ., Έστω ότι μια γραμματική έχει τους ακόλουθους κανόνες:

$$A \rightarrow B, \quad B \rightarrow u$$

■ Τότε η γραμματική που σχηματίζεται αφαιρώντας τον κανόνα $A \rightarrow B$ θα έχει το αντίστοιχο σύνολο κανόνων:

$$B \rightarrow u, \quad A \rightarrow u \text{ [προσθήκη κανόνα]}$$

Μετατροπή CFG σε CNF

4. Συντόμευση του δεξιού μέλους.

■ Για κάθε κανόνα της μορφής $A \rightarrow u_1 u_2 \dots u_k$ για $k \geq 3$, όπου $u_i \in V \cup \Sigma$, αντικαθιστούμε τον κανόνα με:

$$A \rightarrow u_1 A_1$$

$$A_1 \rightarrow u_2 A_2$$

$$A_2 \rightarrow u_3 A_3$$

$$\vdots$$

$$A_{k-2} \rightarrow u_{k-1} u_k$$

όπου τα A_i είναι οι νέες μεταβλητές που προστίθενται στη γραμματική.

Μετατροπή CFG σε CNF

5. Αντικατάσταση ορισμένων τερματικών.

■ Αν υπάρχει κανόνας της μορφής $A \rightarrow uv$ όπου τουλάχιστον ένα από τα u ή v είναι τερματικό σύμβολο (π.χ. u), τότε αντικαθιστούμε τον κανόνα $A \rightarrow uv$ με:

$$A \rightarrow Uv$$

$$U \rightarrow u$$

όπου το U είναι μια νέα μεταβλητή που προστίθεται στη γραμματική. Επαναλαμβάνουμε μέχρι να μην υπάρχουν τέτοιοι κανόνες.

Μετατροπή CFG σε CNF

π.χ., Έστω η CFG με αρχική μεταβλητή S ,

$$S \rightarrow ASB, \quad A \rightarrow aASA \mid a \mid \varepsilon, \quad B \rightarrow SbS \mid A \mid bb.$$

Θα την μετατρέψουμε σε CNF.

1. Καθώς η αρχική μεταβλητή S εμφανίζεται και στο δεξιό μέλος των παραγωγών, προσθέτουμε νέα αρχική μεταβλητή S_0 ,

$$S_0 \rightarrow S$$

$$S \rightarrow ASB$$

$$A \rightarrow aASA \mid a \mid \varepsilon$$

$$B \rightarrow SbS \mid A \mid bb$$

Μετατροπή CFG σε CNF

π.χ., (συνέχεια)

2. Διαγράφουμε την ε -παραγωγή $A \rightarrow \varepsilon$,

$$S_0 \rightarrow S$$

$$S \rightarrow ASB \mid \mathbf{SB}$$

$$A \rightarrow aASA \mid a \mid \mathbf{aSA} \mid \mathbf{aAS} \mid \mathbf{aS}$$

$$B \rightarrow SbS \mid A \mid bb\varepsilon$$

3. Διαγράφουμε την ε -παραγωγή $B \rightarrow \varepsilon$,

$$S_0 \rightarrow S$$

$$S \rightarrow ASB \mid SB \mid \mathbf{AS} \mid \mathbf{S}$$

$$A \rightarrow aASA \mid a \mid aSA \mid aAS \mid aS$$

$$B \rightarrow SbS \mid A \mid bb$$

Μετατροπή CFG σε CNF

π.χ., (συνέχεια)

4. Διαγράφουμε τον μοναδιαίο κανόνα $S \rightarrow S$,

$$S_0 \rightarrow S$$

$$S \rightarrow ASB \mid SB \mid AS$$

$$A \rightarrow aASA \mid a \mid aSA \mid aAS \mid aS$$

$$B \rightarrow SbS \mid A \mid bb$$

5. Διαγράφουμε τον μοναδιαίο κανόνα $B \rightarrow A$,

$$S_0 \rightarrow S$$

$$S \rightarrow ASB \mid SB \mid AS$$

$$A \rightarrow aASA \mid a \mid aSA \mid aAS \mid aS$$

$$B \rightarrow SbS \mid \mathbf{aASA} \mid \mathbf{a} \mid \mathbf{aSA} \mid \mathbf{aAS} \mid \mathbf{aS} \mid bb$$

Μετατροπή CFG σε CNF

π.χ., (συνέχεια)

6. Διαγράφουμε τον μοναδιαίο κανόνα $S_0 \rightarrow S$,

$$S_0 \rightarrow \mathbf{ASB} \mid \mathbf{SB} \mid \mathbf{AS}$$

$$S \rightarrow ASB \mid SB \mid AS$$

$$A \rightarrow aASA \mid a \mid aSA \mid aAS \mid aS$$

$$B \rightarrow SbS \mid A \mid bb$$

7. Προσθέτουμε την μεταβλητή X και τον κανόνα $X \rightarrow AS$,

$$S_0 \rightarrow \mathbf{XB} \mid SB \mid AS$$

$$S \rightarrow \mathbf{XB} \mid SB \mid AS$$

$$A \rightarrow \mathbf{aXA} \mid a \mid aSA \mid \mathbf{aX} \mid aS$$

$$B \rightarrow SbS \mid \mathbf{aXA} \mid a \mid aSA \mid \mathbf{aX} \mid aS \mid bb$$

$$\mathbf{X} \rightarrow \mathbf{AS}$$

Μετατροπή CFG σε CNF

π.χ., (συνέχεια)

8. Προσθέτουμε την μεταβλητή Y και τον κανόνα $Y \rightarrow aX$,

$$S_0 \rightarrow XB \mid SB \mid AS$$

$$S \rightarrow XB \mid SB \mid AS$$

$$A \rightarrow \mathbf{YA} \mid a \mid aSA \mid aX \mid aS$$

$$B \rightarrow SbS \mid \mathbf{YA} \mid a \mid aSA \mid aX \mid aS \mid bb$$

$$X \rightarrow AS$$

$$\mathbf{Y} \rightarrow \mathbf{aX}$$

Μετατροπή CFG σε CNF

π.χ., (συνέχεια)

9. Προσθέτουμε την μεταβλητή Z και τον κανόνα $Z \rightarrow aS$,

$$S_0 \rightarrow XB \mid SB \mid AS$$

$$S \rightarrow XB \mid SB \mid AS$$

$$A \rightarrow YA \mid a \mid ZA \mid aX \mid aS$$

$$B \rightarrow SbS \mid YA \mid a \mid \mathbf{ZA} \mid aX \mid aS \mid bb$$

$$X \rightarrow AS$$

$$Y \rightarrow aX$$

$$\mathbf{Z} \rightarrow \mathbf{aS}$$

Μετατροπή CFG σε CNF

π.χ., (συνέχεια)

10. Προσθέτουμε την μεταβλητή W και τον κανόνα $W \rightarrow Sb$,

$$S_0 \rightarrow XB \mid SB \mid AS$$

$$S \rightarrow XB \mid SB \mid AS$$

$$A \rightarrow YA \mid a \mid ZA \mid aX \mid aS$$

$$B \rightarrow \mathbf{WS} \mid YA \mid a \mid ZA \mid aX \mid aS \mid bb$$

$$X \rightarrow AS$$

$$Y \rightarrow aX$$

$$Z \rightarrow aS$$

$$\mathbf{W} \rightarrow \mathbf{Sb}$$

Μετατροπή CFG σε CNF

π.χ., (συνέχεια)

11. Προσθέτουμε τις μεταβλητές V , U και τους κανόνες $V \rightarrow a$, $U \rightarrow b$,

$$S_0 \rightarrow XB \mid SB \mid AS$$

$$S \rightarrow XB \mid SB \mid AS$$

$$A \rightarrow YA \mid a \mid ZA \mid \mathbf{VX} \mid \mathbf{VS}$$

$$B \rightarrow WS \mid YA \mid a \mid ZA \mid \mathbf{VX} \mid \mathbf{VS} \mid \mathbf{UU}$$

$$X \rightarrow AS$$

$$Y \rightarrow \mathbf{VX}$$

$$Z \rightarrow \mathbf{VS}$$

$$W \rightarrow \mathbf{SU}$$

$$\mathbf{V} \rightarrow \mathbf{a}$$

$$\mathbf{U} \rightarrow \mathbf{b}$$

Μετατροπή CFG σε CNF

π.χ., (συνέχεια), Καταλήγουμε ότι η Κανονική μορφή Chomsky της γραμματικής είναι:

$$S_0 \rightarrow XB \mid SB \mid AS$$

$$S \rightarrow XB \mid SB \mid AS$$

$$A \rightarrow YA \mid ZA \mid VX \mid VS \mid a$$

$$B \rightarrow WS \mid YA \mid ZA \mid VX \mid VS \mid UU \mid a$$

$$X \rightarrow AS$$

$$Y \rightarrow VX$$

$$Z \rightarrow VS$$

$$W \rightarrow SU$$

$$V \rightarrow a$$

$$U \rightarrow b$$



Κανονική μορφή Chomsky

Μερικά σχόλια για τις CNF.

■ **Απλότητα αποδείξεων:** Υπάρχουν πολλές αποδείξεις γύρω από τις γραμματικές χωρίς συμφραζόμενα. Όσο πιο περιορισμένο είναι το σύνολο των γραμματικών με το οποίο ασχολούμαστε, τόσο πιο απλές είναι οι αποδείξεις. Επομένως, οι κανονικές μορφές μπορούν να είναι χρήσιμες εδώ.

■ **Διευκόλυνση της ανάλυσης:** Όπως θα δούμε, υπάρχουν αυτόματα που μπορούν να χρησιμοποιηθούν για την ανάλυση συμβολοσειρών με οποιαδήποτε γραμματική, αυτό όμως συχνά δεν είναι βολικό. Οι κανονικές μορφές μπορούν να μας δώσουν περισσότερη δομή για να δουλέψουμε, με αποτέλεσμα την ευκολότερη ανάπτυξη αλγορίθμων ανάλυσης.

Γραμματικές

Μερικά σχόλια.

■ **Παραγωγή Γλωσσών:** Οι γραμματικές παρέχουν έναν τυπικό τρόπο περιγραφής της σύνταξης των γλωσσών. Καθορίζουν πώς μπορούν να δημιουργηθούν οι συμβολοσειρές σε μια γλώσσα, κάτι που είναι απαραίτητο για τον ορισμό των γλωσσών προγραμματισμού και των φυσικών γλωσσών.

■ **Θεωρία Αυτομάτων:** Οι γραμματικές σχετίζονται στενά με τα αυτόματα, τα οποία είναι αφηρημένες μηχανές που χρησιμοποιούνται για την αναγνώριση γλωσσών.

■ **Σχεδιασμός Μεταγλωττιστών:** Στο σχεδιασμό μεταγλωττιστών, οι γραμματικές χρησιμοποιούνται για την ανάλυση του πηγαίου κώδικα. Η γραμματική μιας γλώσσας προγραμματισμού βοηθά στην κατασκευή του συντακτικού δέντρου, το οποίο είναι ένα κρίσιμο βήμα για τη μετάφραση του κώδικα υψηλού επιπέδου σε κώδικα.

Γραμματικές

Μερικά σχόλια.

■ **Τυπική Επαλήθευση:** Οι γραμματικές χρησιμοποιούνται στην τυπική επαλήθευση (ορθότητα) για να διασφαλιστεί ότι τα συστήματα λειτουργούν όπως αναμένεται. Ορίζοντας τους κανόνες μιας γλώσσας, οι γραμματικές βοηθούν στην επαλήθευση ότι οι έξοδοι του συστήματος είναι σωστές.

■ **Κανονικές Μορφές:** Οι κανονικές μορφές, όπως η CNF, απλοποιούν τις γραμματικές και τις καθιστούν ευκολότερες στη χρήση. Αυτές οι μορφές είναι χρήσιμες στην απόδειξη ιδιοτήτων για γλώσσες και αυτόματα, όπως η ύπαρξη συγκεκριμένων τύπων αυτομάτων για κάθε CFG.

Κανονικές Γραμματικές vs CFG

Κανονικές Γραμματικές	CFG
1. δημιουργούν κανονικές γλώσσες	δημιουργούν CFL
2. τυπικά αντικαθιστούν ένα μη τερματικό με ένα τερματικό που ακολουθείται από το πολύ ένα μη-τερματικό, σχηματίζοντας γραμμικούς κανόνες δεξιά ή αριστερά	επιτρέπουν την αντικατάσταση ενός μόνο μη-τερματικού από μια ακολουθία τερματικών και μη-τερματικών χωρίς περιορισμούς στη θέση των μη-τερματικών.
3. περιγράφει γραμμικές δομές	περιγράφει εμφωλιασμένες δομές
4. δεν απαιτεί επιπλέον μνήμη	απαιτεί επιπλέον μνήμη
5. FA	[Διάλεξη 9]

■ Οι κανονικές γραμματικές είναι κατάλληλες για απλή αντιστοίχιση προτύπων, όπως ο διαχωρισμός λέξεων σε κείμενο.

■ Οι CFG είναι κατάλληλες για πιο σύνθετες δομές γλώσσας, όπως η ανάλυση της σύνταξης των γλωσσών προγραμματισμού ή η επικύρωση εμφωλιασμένων εκφράσεων.