Parsing

Example: Transformation into CNF

Kurt Eberle

k.eberle@lingenio.de

January 09, 2020

Outline

Outline

Chomsky Normal Form

Grammar $G = (V, \Sigma, P, S)$ is in CNF iff G has the form (i.e. the G rules have the form:)

- $A \rightarrow BC$
- $A \rightarrow a$
- ($S \rightarrow \epsilon$; in this case S doesn't appear on the right hand side of a rule)

for $A, B, C \in V, a \in \Sigma$.

Outline

Grammar

Grammar $G = (V, \Sigma, P, S)$ where

- V= {S, A, B, C, D, E ,F, G, H, I}
- Σ= {c, d, e, f, g, h, i}
- and P:

$S \rightarrow ABC$	$D \rightarrow \varepsilon$
$S \rightarrow \varepsilon$	$E \rightarrow deE$
$A \rightarrow DEF$	$E \rightarrow \varepsilon$
$B \rightarrow GHI$	$F \rightarrow efF$
$B \rightarrow \varepsilon$	$F \rightarrow \varepsilon$
$C \rightarrow cC$	$G \rightarrow g$
$C \rightarrow \varepsilon$	$H \rightarrow gh$
$D \rightarrow cdD$	$I \rightarrow hi$

Outline

1. New unary rules

- $\Sigma = \{c, d, e, f, g, h, i\}$
- unary rules

$X_c \to c$ $X_g \to g$
$X_d \to d$ $X_h \to h$
$X_e \to e$ $X_i \to i$
$X_f \to f$

- replacement

$S \to ABC$ $D \to \varepsilon$
$S \to \varepsilon$ $E \to X_d X_e E$
$A \to DEF$ $E \to \varepsilon$
$B \to GHI$ $F \to X_e X_f F$
$B \to \varepsilon$ $F \to \varepsilon$
$C \to X_c C$ $G \to X_g$
$C \to \varepsilon$ $H \to X_g X_h$
$D \to X_c X_d D$ $I \to X_h X_i$

2. New pair of symbols rules

- new pair rules

$X_{AB} \to AB$ $X_{cd} \to X_c X_d$
$X_{DE} \to DE$ $X_{de} \to X_d X_e$
$X_{GH} \to GH$ $X_{ef} \to X_e X_f$

- replacement

$S \to X_{AB} C$ $D \to \varepsilon$
$S \to \varepsilon$ $E \to X_{de} E$
$A \to X_{DE} F$ $E \to \varepsilon$
$B \to X_{GH} I$ $F \to X_{ef} F$
$B \to \varepsilon$ $F \to \varepsilon$
$C \to X_c C$ $G \to X_g$
$C \to \varepsilon$ $H \to X_g X_h$
$D \to X_{cd} D$ $I \to X_h X_i$

3. S-ε rule

- new rules
  $S' \to S \quad S' \to \varepsilon$
- result: ($+$ new rules from steps 1 and 2)

$S' \to S$	$D \to X_{cd} D$
$S' \to \varepsilon$	$D \to \varepsilon$
$S \to X_{AB} C$	$E \to X_{de} E$
$S \to \varepsilon$	$E \to \varepsilon$
$A \to X_{DE} F$	$F \to X_{ef} F$
$B \to X_{GH} I$	$F \to \varepsilon$
$B \to \varepsilon$	$G \to X_g$
$C \to X_c C$	$H \to X_g X_h$
$C \to \varepsilon$	$I \to X_h X_i$

4. A-ε rules (first pass)

- reject rules $S \to \varepsilon$ $D \to \varepsilon$ $B \to \varepsilon$ $E \to \varepsilon$ $C \to \varepsilon$ $F \to \varepsilon$
- add rules (for B, C, D, E, F) $X_{AB} \to A$ $X_{DE} \to D$ $C \to X_c$ $X_{DE} \to \varepsilon$ $S \to X_{AB}$ $E \to X_{de}$ $X_{DE} \to E$ $F \to X_{ef}$ $D \to X_{cd}$ $A \to X_{DE}$
- result: (+ terminal rules)

$X_{AB} \to AB$ $S' \to S$
$X_{AB} \to A$ $S' \to \varepsilon$
$X_{DE} \to DE$ $S \to X_{AB}C$
$X_{DE} \to D$ $S \to X_{AB}$
$X_{DE} \to E$ $A \to X_{DE}F$
$X_{DE} \to \varepsilon$ $A \to X_{DE}$
$X_{GH} \to GH$ $B \to X_{GH}I$

$C \to X_c C$
$C \to X_c$
$D \to X_{cd} D$
$D \to X_{cd}$
$E \to X_{de} E$
$E \to X_{de}$

$F \to X_{ef} F$
$F \to X_{ef}$
$G \to X_g$
$H \to X_g X_h$
$I \to X_h X_i$

4. A-ε rules (second pass)

- reject rules $X_{DE} \to \varepsilon$
- add rules (for X_{DE}) $A \to F$ $A \to \varepsilon$
- result: (+ terminal rules)

$X_{AB} \to AB$
$X_{AB} \to A$
$X_{DE} \to DE$
$X_{DE} \to D$
$X_{DE} \to E$
$X_{GH} \to GH$

$S' \to S$
$S' \to \varepsilon$
$S \to X_{AB} C$
$S \to X_{AB}$
$A \to X_{DE} F$
$A \to F$
$A \to X_{DE}$
$A \to \varepsilon$

$B \to X_{GH} I$
$C \to X_c C$
$C \to X_c$
$D \to X_{cd} D$
$D \to X_{cd}$
$E \to X_{de} E$
$E \to X_{de}$

$F \to X_{ef} F$
$F \to X_{ef}$
$G \to X_g$
$H \to X_g X_h$
$I \to X_h X_i$

4. A-ε rules (third pass)

- reject rules $A \to \varepsilon$
- add rules (for A) $X_{AB} \to B$ $X_{AB} \to \varepsilon$
- result: (+ terminal rules)

$X_{AB} \to AB$
$X_{AB} \to A$
$X_{AB} \to B$
$X_{AB} \to \varepsilon$
$X_{DE} \to DE$
$X_{DE} \to D$
$X_{DE} \to E$
$X_{GH} \to GH$

$S' \to S$
$S' \to \varepsilon$
$S \to X_{AB} C$
$S \to X_{AB}$
$A \to X_{DE} F$
$A \to F$
$A \to X_{DE}$

$B \to X_{GH} I$
$C \to X_c C$
$C \to X_c$
$D \to X_{cd} D$
$D \to X_{cd}$
$E \to X_{de} E$
$E \to X_{de}$

$F \to X_{ef} F$
$F \to X_{ef}$
$G \to X_g$
$H \to X_g X_h$
$I \to X_h X_i$

4. A-ε rules (fourth pass)

- reject rules $X_{AB} \to \varepsilon$
- add rules (for X_{AB}) $S \to C$ $S \to \varepsilon$
- result: (+ terminal rules)

$S' \to S$

$X_{AB} \to AB$	$S' \to \varepsilon$	$B \to X_{GH} I$	
$X_{AB} \to A$	$S \to X_{AB} C$	$C \to X_c C$	$F \to X_{ef} F$
$X_{AB} \to B$	$S \to C$	$C \to X_c$	$F \to X_{ef}$
$X_{DE} \to DE$	$S \to X_{AB}$	$D \to X_{cd} D$	$G \to X_g$
$X_{DE} \to D$	$S \to \varepsilon$	$D \to X_{cd}$	$H \to X_g X_h$
$X_{DE} \to E$	$A \to X_{DE} F$	$E \to X_{de} E$	$I \to X_h X_i$
$X_{GH} \to GH$	$A \to F$	$E \to X_{de}$	
	$A \to X_{DE}$		

4. A-ε rules (fifth pass)

- reject rules $S \to \varepsilon$
- add rules (for S) $\emptyset$
- result: (+ terminal rules)

$X_{AB} \to AB$
$X_{AB} \to A$
$X_{AB} \to B$
$X_{DE} \to DE$
$X_{DE} \to D$
$X_{DE} \to E$
$X_{GH} \to GH$

$S' \to S$
$S' \to \varepsilon$
$S \to X_{AB} C$
$S \to C$
$S \to X_{AB}$
$A \to X_{DE} F$
$A \to F$
$A \to X_{DE}$

$B \to X_{GH} I$
$C \to X_c C$
$C \to X_c$
$D \to X_{cd} D$
$D \to X_{cd}$
$E \to X_{de} E$
$E \to X_{de}$

$F \to X_{ef} F$
$F \to X_{ef}$
$G \to X_g$
$H \to X_g X_h$
$I \to X_h X_i$

5. Chain rules

- replace rules

$X_{AB} \to A$	$X_{AB} \to X_{DE}F$, $X_{AB} \to X_e X_f$, $X_{AB} \to X_{ef}F$, $X_{AB} \to DE$
$X_{AB} \to B$	$X_{AB} \to X_{GH}I$
$X_{DE} \to D$	$X_{DE} \to X_c X_d$, $X_{DE} \to X_{cd}D$
$X_{DE} \to E$	$X_{DE} \to X_d X_e$, $X_{DE} \to X_{de}E$
$S' \to S$	$S' \to X_{AB}C$, $S' \to X_c C$, $S' \to c$, $S' \to AB$
$S \to C$	$S \to X_c C$, $S \to c$
$S \to X_{AB}$	$S \to AB$
$A \to F$	$A \to X_e X_f$, $A \to X_{ef}F$
$A \to X_{DE}$	$A \to DE$
$C \to X_c$	$C \to c$
$D \to X_{cd}$	$D \to X_c X_d$
$E \to X_{de}$	$E \to X_d X_e$
$F \to X_{ef}$	$F \to X_e X_f$
$G \to X_g$	$G \to g$

Result

$X_{AB} \to AB$
$X_{AB} \to X_{DE}F \mid X_e X_f \mid X_{ef} F \mid DE$
$X_{AB} \to X_{GH} I$
$X_{DE} \to DE$
$X_{DE} \to X_c X_d \mid X_{cd} D$
$X_{DE} \to X_d X_e \mid X_{de} E$
$X_{GH} \to GH$
$S' \to X_{AB} C \mid X_c C \mid c \mid AB$
$S' \to \varepsilon$

$A \to X_{DE} F$
$A \to X_e X_f \mid X_{ef} F$
$A \to DE$
$B \to X_{GH} I$
$C \to X_c C$
$C \to c$
$D \to X_{cd} D$
$D \to X_c X_d$
$E \to X_{de} E$
$E \to X_d X_e$

$F \to X_{ef} F$
$F \to X_e X_f$
$G \to g$
$H \to X_g X_h$
$I \to X_h X_i$