ECS120 Introduction to the Theory of Computation
Fall Quarter 2007

Discussion Notes
Wednesday, October 31, 2007

Announcements

Homework 4 is not yet graded, it will be finished this weekend. Homework (from 5 on) will always be due on Mondays at the start of class. Homework in the Kemper 1131 homework box will be picked up at 1pm sharp! (Remember late homework is not accepted.)

Chomsky Normal Form

A context-free grammar is in **Chomsky normal form** if every rule is of the form:

$$\begin{aligned} S &\rightarrow \epsilon \\ A &\rightarrow BC \\ A &\rightarrow a \end{aligned}$$

where a is any terminal, S is the start variable, A is any variable, and B, C are any variables except the start variable.

Theorem 2.9 states that "Any context-free language is generated by a context-free grammar in Chomsky normal form" (page 107).

Let u, v are strings including variables and terminals. To convert a grammar into CNF, use the following steps:

1. Add a new start variable S_0 and the rule $S_0 \rightarrow S$.
2. Eliminate all ϵ rules. For every rule of the form $A \rightarrow \epsilon$ where $A \neq S$:
   (a) Remove rule $A \rightarrow \epsilon$.
   (b) For every rule containing the variable A:
      - For every occurrence of A in the rule where $R \rightarrow uAv$:
         - Add rule $R \rightarrow uv$.

           Special Cases:
           * The rule $R \rightarrow A$ adds rule $R \rightarrow \epsilon$. Notice that $R \rightarrow A$ may be rewritten as $R \rightarrow \epsilon\, A\, \epsilon$. Therefore $u, v = \epsilon$ and the new rule $R \rightarrow u\,v$ is actually $R \rightarrow \epsilon\,\epsilon$. This is equivalent to $R \rightarrow \epsilon$.

- $*$ The rule has multiple occurrences of A. For example $R \to x\,A\,y\,A\,z$. Consider each occurrence of A separately.

$$R \to \underbrace{x}_{u}\;A\;\underbrace{y\,A\,z}_{v} \quad \text{versus} \quad R \to \underbrace{x\,A\,y}_{u}\;A\;\underbrace{z}_{v}$$

  The first split sets $u = x$ and $v = yAz$, adding the new rule $R \to xyAz$. The second split sets $u = xAy$ and $v = z$, adding the new rule $R \to xAyz$.

  Now we have two new rules we have to consider. Since $R \to xyAz$ and $R \to xAyz$ both have occurences of A, we must repeat this step. Both cases result in adding the new rule $R \to xyz$.

  Therefore the rule $R \to x\,A\,y\,A\,z$ becomes $R \to x\,A\,y\,A\,z \mid x\,y\,A\,z \mid x\,A\,y\,z \mid x\,y\,z$.

3. Eliminate all unit rules. For every rule of the form $A \to B$:

   (a) Remove rule $A \to B$.

   (b) For every occurence of $B \to u$, add the rule $A \to u$.

4. Convert all rules into proper form.

   (a) For every rule of the form $A \to u_1\,u_2\,\cdots\,u_k$ where $k \geq 3$ and each u_i is a variable or terminal:

   - Remove the rule $A \to u_1\,u_2\,\cdots\,u_k$.
   - Add rule $A \to u_1\,A_1$.
   - For $i = 1 \ldots k-3$, add rule $A_i \to u_{i+1}\,A_{i+1}$.
   - Add rule $A_{k-2} \to u_{k-1}\,u_k$.

   For example rule $A \to BCD$ is replaced by rules $A \to BA_1$, $A_1 \to CD$.

   (b) For every rule of the form $A \to u_1\,u_2$:

   - For every terminal u_i:
     - Replace u_i in the rule with variable U_i.
     - Add the rule $U_i \to u_i$.

   For example rule $A \to bB$ becomes $A \to U_1B$ and we add rule $U_1 \to b$.

There is an extensive example in the book on page 108.