Homework 6

Posted: Friday, Nov 6, 2015 – 11:59pm
Due: Tuesday, Nov 17, 2015 – 3pm (HFH 2108) or 3:30pm (in class)

Task 1 – Simplifying Context-Free Grammars (10 points)

a) Remove all unit productions from the following CFG with start variable $S$:

$S \rightarrow aAB \mid b$
$A \rightarrow B \mid aC \mid a$
$B \rightarrow S \mid bBaC \mid C \mid a$
$C \rightarrow aC \mid bb \mid A$

b) Remove all $\lambda$, unit- and useless productions (in this order) from the following CFG with start variable $S$:

$S \rightarrow aAB \mid b$
$A \rightarrow B \mid aCB \mid \lambda \mid a$
$B \rightarrow bb \mid C \mid \lambda$
$C \rightarrow aC \mid bC$

c) Explain (1) why removing unit productions from a CFG without $\lambda$-productions does not add new $\lambda$-productions, and (2) why removing useless productions from a CFG without $\lambda$- and unit-productions does not add any $\lambda$- and unit-productions.

d) Can removing $\lambda$-productions from a CFG $G$ with $\lambda \notin L(G)$ create new unit productions?

e) Can removing unit-productions from a CFG without $\lambda$-productions create new useless productions?

Task 2 – Chomsky Normal Form (6 points)

Transform each of the two following CFGs into an equivalent CFG in Chomsky Normal Form.

a) $S \rightarrow aSa \mid bSb$
$S \rightarrow aa \mid bb$

b) $S \rightarrow aAB \mid b \mid bCa$
$A \rightarrow aB \mid bC$
$B \rightarrow aB \mid bAbB \mid a$
$C \rightarrow aC \mid bCAa \mid \lambda$. 
Task 3 – The CYK Algorithm (8 points)

Consider the following CFG $G$ with start variable $S$:

\[
S \rightarrow AB \mid AC \\
C \rightarrow SD \\
D \rightarrow BB \\
A \rightarrow a \\
B \rightarrow b
\]

Note that $G$ is already in Chomsky Normal Form.

a) Run the CYK algorithm with input $w = aabbb$. Is $w$ in $L(G)$?

b) Run the CYK algorithm with input $w' = aaabbb$. Is $w'$ in $L(G)$?

**Hint:** For both a) and b), complete the triangular table describing the complete execution of the CYK algorithm and use it to conclude whether the string $w$ is in $L(G)$ or not. In particular, give all sets $V_{ij}$ for $i \leq j$.

As explained in class, one can modify the CYK algorithm to build a derivation tree for its input $w$, whenever $w \in L(G)$: When a variable $A$ is added to $V_{ij}$ because of a production $A \rightarrow BC$ with $B \in V_{i,k}$ and $C \in V_{k+1,j}$, then one stores this production together with $k$ and $A$ in the table entry.

c) Use the modified CYK algorithm to give a derivation tree for $w = aaabbbbb$.

Task 4 – Length of Derivations (6 points)

We want to study the relationship between the structural assumptions made on a CFG $G = (V, T, S, P)$ and the length of a derivation $S \Rightarrow^* w$ for a string in $w \in L(G)$.

a) Assume that $G$ does not have any unit- and $\lambda$-productions. Show that every derivation of $w \in L(G)$ consists of at most $2|w|$ steps.

b) Assume that $G$ is in Greibach Normal Form. What is the exact number of steps (as a function of $|w|$) needed to derive a string $w \in L(G)$?

c) Assume that $G$ is in Chomsky Normal Form. What is the exact number of steps (as a function of $|w|$) needed to derive a string $w \in L(G)$?

**Hint:** For b) and c), it may be worth it to initially find some examples of concrete CFGs and of strings belonging to the respective languages, and take an educated guess. Then, prove your guess correct for any grammar of the given form.