

Assignment 2

(due Tue, 29/11/2011)

Compiler Design I (Kompilorteknik I) 2011

1 Context-free grammars

Give the definition of a context free grammar over the alphabet $\Sigma = \{a, b\}$ that describes all strings that have a different number of 'a's and 'b's.

2 Parsing and semantic actions

The following context-free grammar can parse all the lowercase roman numerals from 1-99. The terminal symbols are $\{c, l, x, v, i\}$ and the initial symbol is S . If you are unfamiliar with roman numerals, please have a look at http://en.wikipedia.org/wiki/Roman_numerals.

$$\begin{aligned} S &\rightarrow \mathbf{x}TU \mid \mathbf{l}X \mid X \\ T &\rightarrow \mathbf{c} \mid \mathbf{l} \\ X &\rightarrow \mathbf{x}X \mid U \\ U &\rightarrow \mathbf{i}Y \mid \mathbf{v}I \mid I \\ Y &\rightarrow \mathbf{x} \mid \mathbf{v} \\ I &\rightarrow \mathbf{i}I \mid \epsilon \end{aligned}$$

1. Draw a parse tree for 42: “**xlii**”
2. Is this grammar ambiguous?
3. Write semantic actions for each of the 14 rules in the grammar (remember $X \rightarrow A|B$ is short for $X \rightarrow A$ and $X \rightarrow B$) to calculate the decimal value of the input string. You can associate a synthesized attribute `val` to each of the non-terminals to store their value. The final value should be returned in `S.val`.

3 LL(1) Parsers

In the following context-free grammar, the symbols **0**, **1**, **2** and **3** are terminals and S is the initial symbol.

$$\begin{aligned} S &\rightarrow \mathbf{0} \mid \mathbf{1} S \mathbf{2} S \mathbf{3} \mid \mathbf{1} A \mathbf{3} \\ A &\rightarrow S \mid A S \end{aligned}$$

1. Explain briefly why this grammar is not LL(1).
2. Convert this grammar to an equivalent that is LL(1).
3. For the grammar of the previous subtask, construct the complete LL(1) parsing table.
4. Show all the steps required to parse the input string: **1 1 0 2 0 3 0 1 0 3 3**

4 LR(1) Parsers

In the following context-free grammar, the symbols $(, \mathbf{a},)$ and $,$ are terminals. and S is the initial symbol.

- (1) $S \rightarrow (L)$
- (2) $S \rightarrow \mathbf{a}$
- (3) $L \rightarrow L , S$
- (4) $L \rightarrow S$

Because $,$ is a symbol of the language we are going to use $|$ as a separator between the core of the LR(1) items and the lookahead symbols. Lookaheads with the same core can be separated as usual with $/$.

1. Calculate the closure of the LR(1) item $[S \rightarrow (\cdot L) | \$]$.
2. Construct the full LR(1) DFA, showing all items in each state.
3. Construct the LR(1) parsing table using the DFA. For the reduce actions, please use the provided enumeration of the productions in the grammar.
4. Show all the steps required to parse the input string: $((\mathbf{a} , \mathbf{a}) , \mathbf{a} , \mathbf{a})$

Instructions

There are two ways to submit this assignment:

1. Submit a physical copy of your answers in my mailbox (Aronis Stavros, 59) on the 4th floor of building 1, opposite the 'fika' room.
2. Send an email with an electronic copy of your answers to `stavros.aronis@it.uu.se`.

Good luck!