CS 434 Meeting 30— 4/26/02

Announcements

1. Colloquium speaker: David Detlefs on Garbage collection.

The Correctness of LR(0) parsing

1. The correctness of the LR(0) parsers we have discussed rests on the theorem:

   Theorem $[N\rightarrow\beta_1.\beta_2] \in \Delta(\pi_0, \gamma)$ iff $[N\rightarrow\beta_1.\beta_2]$ is valid for γ.

2. Rather than completely prove the theorem, I would like to prove it in one direction (the easy one) and let you convince yourselves of the other direction. In particular, I will prove that:

   > $[N \rightarrow \beta_1.\beta_2] \in \Delta(\pi_0, \gamma)$ only if $[N\rightarrow\beta_1.\beta_2]$ is valid for γ.

3. The proof of this lemma is simplified by identifying a way to partition the set of LR(0) items associated with a state of the LR(0) machine into tow parts.

   Kernel items We say that an LR(0) item is a kernel item if either:

   (a) it is the item $[S'\rightarrow.S]$, or
   (b) it is of the form $[N\rightarrow\beta_1.\beta_2]$ for $\beta_1 \neq \epsilon$.

   All other items are called non-kernel items.

   Note that the set of items associated with a state in the LR(0) machine is just the closure of the set of kernel items associated with the state. In particular, all items in goto(π,x) are kernel items.

4. Given this partition, we can break the proof of this half of the theorem into two lemmas:

   > Lemma 1: Given a set π of LR(0) items valid for some $\gamma \in (V_n \cup V_t)^*$, all items in closure($\pi$) are valid for γ.
   >
   > Lemma 2: For kernel items, $[N \rightarrow \beta_1.\beta_2] \in \Delta(\pi_0, \gamma)$ only if $[N\rightarrow\beta_1.\beta_2]$ is valid for γ.

About this closure stuff...

1. Recall the definition for the closure of a set of LR(0) items:

   closure Given a set π of LR(0) items for a grammar G with productions P, we define closure(π) to be the smallest set of LR(0) items such that:

   (a) closure(π) $\supseteq \pi$
   (b) if $[N_1\rightarrow\beta_1.N_2\beta_2] \in$ closure(π) and $N_2\rightarrow\beta_3 \in P$ then $[N_2\rightarrow.\beta_3] \in$ closure(π)

   and the algorithm given for computing the closure of a given set of LR(0) items:

   - An algorithm to compute closure(π)
     (a) set π' equal to π.
     (b) while there is some $[N\rightarrow\beta_1.M\beta_2] \in \pi'$ such that $M\rightarrow\beta_3 \in P$ and $[M\rightarrow.\beta_3] \notin \pi'$ add $[M\rightarrow.\beta_3]$ to π'.

2. How can we prove that the sets described by the definition and produced by the algorithm are the same?

   By identifying the invariant of the loop. Namely:

   > At the beginning (and end) of each iteration of step (b), π' is a subset of the closure of π.

   - This is clearly true before the first iteration.

- If it is true before any subsequent iteration, then the item [$M \rightarrow .\beta_3$] that the iteration will add must also be a member of $closure(\pi)$ by the definition of $closure$.

These facts establish that π' will always be a subset of $closure(\pi)$. If the loop terminates, then we know that π' must contain $closure(\pi)$ and therefore $\pi' = closure(\pi)$. Luckily, the loop must terminate since there are only a finite number of LR(0) items that step (b) could add to π'.

3. We can prove that each item in closure(π) is valid for γ by induction on the number of iterations of the loop in the algorithm to compute the closure executed before the item was included in π'.

basis If an item was added in 0 steps it must be an element of π and is therefore valid for γ by assumption.

induction Assume that any item added in less than n steps is valid for γ and let $[M \rightarrow .\beta_3]$ be the item added at step n. The addition of this item implies that some item of the form $[N \rightarrow \beta_1 .M\beta_2]$ must have been added at an earlier step. By our inductive assumption this item that was added earlier must be valid for γ. Accordingly, there must be some derivation:

$$S' \overset{*}{\underset{\text{rm}}{\Rightarrow}} \alpha N \omega \underset{\text{rm}}{\Rightarrow} \alpha \beta_1 M \beta_2 \omega$$

with $\gamma = \alpha\beta_1$. Assuming the grammar has no useless non-terminals, it must be possible to derive some string of terminals, ω' from β_2. Thus, there is a derivation:

$$S' \overset{*}{\underset{\text{rm}}{\Rightarrow}} \alpha N \omega \overset{*}{\underset{\text{rm}}{\Rightarrow}} \alpha \beta_1 M \omega' \omega \underset{\text{rm}}{\Rightarrow} \alpha \beta_1 \beta_3 \omega' \omega$$

The existence of this derivation implies that the item $[M \rightarrow .\beta_3]$ is valid for γ.

4. Now, assuming Lemma 1, we can prove Lemma 2 by induction on the length of γ. The basis step is so simple that we will look at the induction step first:

induction Assume that we know that the theorem holds for all strings of length n and consider some string γx such that γ is of length n and x is a single symbol.

Suppose that $[N \rightarrow \beta_1 x.\beta_2]$ is an item in $\Delta(\pi_0, \gamma x)$. The fact that this item is in this set implies that the item $[N \rightarrow \beta_1 .x\beta_2]$ must be in $\Delta(\pi_0, \gamma)$. This, together with our inductive assumption implies that $[N \rightarrow \beta_1 .x\beta_2]$ must be valid for γ. Therefore, there exists a derivation:

$$S' \overset{*}{\underset{\text{rm}}{\Rightarrow}} \alpha N \omega \underset{\text{rm}}{\Rightarrow} \alpha \beta_1 x \beta_2 \omega$$

with $\alpha\beta_1 = \gamma$. This, however implies that $[N \rightarrow \beta_1 x.\beta_2]$ is indeed valid for γx.

basis Similarly, when we consider strings of length 0, the only kernel item in $\Delta(\pi_0, \epsilon)$ is $[S' \rightarrow .S]$. The derivation $S' \overset{*}{\underset{\text{rm}}{\Rightarrow}} S' \underset{\text{rm}}{\Rightarrow} S$ shows that this item is valid for ϵ.