← • HOME • UP • →

Top/Bottom-up parsing/LR(0) items

9.3. LR(0) items

LR(0) items

An *LR(0) item* is a production of the grammar with exactly one dot on the right-hand side. For example, production $T \rightarrow T * F$ leads to four LR(0) items:

$$T \rightarrow \cdot\ T * F$$

$$T \rightarrow T \cdot\ * F$$

$$T \rightarrow T * \cdot\ F$$

$$T \rightarrow T * F \cdot$$

What is to the left of the dot has just been read, and the parser is ready to read the remainder, after the dot.

Two LR(0) items that come from the same production but have the dot in different places are considered different LR(0) items.

Closures

Suppose that S is a set of LR(0) items. The following rules tell how to build *closure(S)*, the *closure* of S. You must add LR(0) items to S until there are no more to add.

1. All members of S are in the closure(S).
2. Suppose closure(S) contains item $A \rightarrow \alpha \cdot B\beta$, where B is a nonterminal. Find all productions $B \rightarrow \gamma_1, \ldots, B \rightarrow \gamma_n$ with B on the left-hand side. Add LR(0) items $B \rightarrow \cdot\gamma_1, \ldots\ B \rightarrow \cdot\gamma_n$ to closure(S).

For example, let's take the closure of set $\{E \rightarrow E + \cdot\ T\}$.

1. Since there is an item with a dot immediately before nonterminal T, we add $T \rightarrow \cdot\ F$ and $T \rightarrow \cdot\ T * F$.

   The set now contains the following LR(0) items.

   $$E \rightarrow E + \cdot\ T$$

   $$T \rightarrow \cdot\ F$$

   $$T \rightarrow \cdot\ T * F$$

2. Now there is an item in the set with a dot immediately followed by F. So we add items $F \rightarrow \cdot\ \mathbf{n}$ and $F \rightarrow \cdot\ \mathbf{(}\ E\ \mathbf{)}$.

The set now contains the following items.

$$E \rightarrow E + \cdot\ T$$
$$T \rightarrow \cdot\ F$$
$$T \rightarrow \cdot\ T * F$$
$$F \rightarrow \cdot\ \mathbf{n}$$
$$F \rightarrow \cdot\ \mathbf{(}\ E\ \mathbf{)}$$

3. No more LR(0) items need to be added, so the closure is finished.

What is the point of the closure? LR(0) item $E \rightarrow E + \cdot\ T$ indicates that the parser has just finished reading an expression followed by a + sign. In fact, E + are the top two symbols on the stack.

Now, the parser is looking to see if there is a T next. (It does not *predict* that there is a T next. It is just considering that as a possibility.)

But that means it should be looking for something that is the right-hand side of a production for T. So we add items for T with the dot at the beginning.

← • HOME • UP • →

Top/Bottom-up parsing/LR(0) items