FIRST and FOLLOW

Lecture 8
Section 4.4

Robb T. Koether

Hampden-Sydney College

Mon, Feb 9, 2009

Outline

Left Factoring

- A problem occurs when two productions for the same nonterminal begin with the same token.
- We cannot decide which production to use.
- This is not necessarily a problem since we could process the part they have in common, then make a decision based on what follows.

Left Factoring

- Consider the grammar

$$A \quad \to \quad \alpha\beta \mid \alpha\gamma$$

- We use left factorization to transform it into the form

$$\begin{array}{rcl} A & \to & \alpha A' \\ A' & \to & \beta \mid \gamma. \end{array}$$

- Now we can apply the productions immediately and unambiguously.

Example

Example (Left Factoring)

- In the earlier example, we had the productions

$$C \longrightarrow \mathbf{id} == \mathbf{num} \mid \mathbf{id}\ != \mathbf{num} \mid \mathbf{id} < \mathbf{num}$$

- To perform left factoring, introduce a nonterminal C':

$$\begin{aligned} C &\longrightarrow \mathbf{id}\ C' \\ C' &\longrightarrow == \mathbf{num} \mid\ != \mathbf{num} \mid < \mathbf{num} \end{aligned}$$

Example

Example (Left Factoring)

- Consider the grammar of **`if`** statements.

$$\begin{array}{rcl} S & \to & \texttt{if}\ C\ \texttt{then}\ S\ \texttt{else}\ S \\ & & \mid \texttt{if}\ C\ \texttt{then}\ S \end{array}$$

- We rewrite it as

$$\begin{array}{rcl} S & \to & \texttt{if}\ C\ \texttt{then}\ SS' \\ S' & \to & \texttt{else}\ S \mid \varepsilon \end{array}$$

Table-Driven LL Parsing

- To build the parsing table, we need three concepts:
  - Nullability
  - The FIRST function
  - The FOLLOW function

Nullability

Definition (Nullable)

A nonterminal A is nullable if

$$A \Rightarrow^* \varepsilon.$$

Nullability

- Clearly, A is nullable if it has a production

$$A \rightarrow \varepsilon.$$

- But A is also nullable if there are, for example, productions

$$\begin{aligned} A &\rightarrow BC \\ B &\rightarrow A \mid aC \mid \varepsilon \\ C &\rightarrow aB \mid Cb \mid \varepsilon \end{aligned}$$

Nullability

- In other words, A is nullable if there is a production

$$A \rightarrow \varepsilon,$$

or there is a production

$$A \rightarrow B_1 B_2 \ldots B_n,$$

where B_1, B_2, ..., B_n are nullable.

Example

Example (Nullability)

- In the grammar

$$
\begin{aligned}
E &\rightarrow T\ E' \\
E' &\rightarrow +\ T\ E' \mid \varepsilon \\
T &\rightarrow F\ T' \\
T' &\rightarrow *\ F\ T' \mid \varepsilon \\
F &\rightarrow (\ E\ ) \mid \mathbf{id} \mid \mathbf{num}
\end{aligned}
$$

- E' and T' are nullable.
- E, T, and F are not nullable.

Example

Example (Nullability)

Nonterminal	Nullable
E	No
E'	Yes
T	No
T'	Yes
F	No

FIRST and FOLLOW

Definition (FIRST)

FIRST(α) is the set of all terminals that may appear as the first symbol in a replacement string of α.

Definition (FOLLOW)

FOLLOW(α) is the set of all terminals that may follow α in a derivation.

- Given a grammar G, we may define the functions FIRST and FOLLOW on the strings of symbols of G.

FIRST

- For a grammar symbol X, FIRST(X) is computed as follows.
  - For every terminal X, FIRST$(X) = \{X\}$.
  - For every nonterminal X, if

$$X \rightarrow Y_1 Y_2 \ldots Y_n$$

    is a production, then
    - FIRST$(Y_1) \subseteq$ FIRST(X).
    - Furthermore, if $Y_1, Y_2, \ldots, Y_k$ are nullable, then

$$\text{FIRST}(Y_{k+1}) \subseteq \text{FIRST}(X).$$

FIRST

- We are concerned with FIRST(X) only for the nonterminals of the grammar.
- FIRST(X) for terminals is trivial.
- According to the definition, to determine FIRST(A), we must inspect all productions that have A on the left.

Example

Example (FIRST)

- Let the grammar be

$$\begin{aligned} E &\rightarrow T\ E' \\ E' &\rightarrow +\ T\ E' \mid \varepsilon \\ T &\rightarrow F\ T' \\ T' &\rightarrow *\ F\ T' \mid \varepsilon \\ F &\rightarrow (E) \mid \mathbf{id} \mid \mathbf{num} \end{aligned}$$

Example

Example (FIRST)

- Find FIRST(E).
  - E occurs on the left in only one production

$$E \quad \to \quad T\ E'$$

  - Therefore, FIRST$(T) \subseteq$ FIRST(E).
  - Furthermore, T is not nullable.
  - Therefore, FIRST$(E) =$ FIRST(T).
  - We have yet to determine FIRST(T).

Example

Example (FIRST)

- Find FIRST(T).
  - T occurs on the left in only one production

$$T \rightarrow F\ T'$$

  - Therefore,

$$\text{FIRST}(F) \subseteq \text{FIRST}(T).$$

  - Furthermore, F is not nullable.
  - Therefore,

$$\text{FIRST}(T) = \text{FIRST}(F).$$

  - We have yet to determine FIRST(F).

Example

Example (FIRST)

- Find FIRST(F).
  - FIRST$(F) = \{\,(, \textbf{id}, \textbf{num}\}$.
- Therefore,
  - FIRST$(E) = \{\,(, \textbf{id}, \textbf{num}\}$.
  - FIRST$(T) = \{\,(, \textbf{id}, \textbf{num}\}$.

Example

Example (FIRST)

- Find FIRST(E').
  - FIRST$(E') = \{+\}$.
- Find FIRST(T').
  - FIRST$(T') = \{*\}$.

Example

Example (FIRST)

Nonterminal	Nullable	FIRST
E	No	$\{\,(, \mathbf{id}, \mathbf{num}\}$
E'	Yes	$\{+\}$
T	No	$\{\,(, \mathbf{id}, \mathbf{num}\}$
T'	Yes	$\{*\}$
F	No	$\{\,(, \mathbf{id}, \mathbf{num}\}$

FOLLOW

- For a grammar symbol X, FOLLOW(X) is defined as follows.
  - If S is the start symbol, then $\$ \in$ FOLLOW(S).
  - If $A \to \alpha B \beta$ is a production, then
    $$\text{FIRST}(\beta) \subseteq \text{FOLLOW}(B).$$
  - If $A \to \alpha B$ is a production, or $A \to \alpha B \beta$ is a production and β is nullable, then
    $$\text{FOLLOW}(A) \subseteq \text{FOLLOW}(B).$$

FOLLOW

- We are concerned about FOLLOW(X) only for the nonterminals of the grammar.
- According to the definition, to determine FOLLOW(A), we must inspect all productions that have A on the right.

Example

Example (FOLLOW)

- Let the grammar be

$$
\begin{array}{rcl}
E & \rightarrow & T\ E' \\
E' & \rightarrow & +\ T\ E' \mid \varepsilon \\
T & \rightarrow & F\ T' \\
T' & \rightarrow & *\ F\ T' \mid \varepsilon \\
F & \rightarrow & (E) \mid \mathbf{id} \mid \mathbf{num}
\end{array}
$$

Example

Example (FOLLOW)

- Find FOLLOW(E).
  - E is the start symbol, therefore
$$\$ \in \text{FOLLOW}(E).$$
  - E occurs on the right in only one production.
$$F \quad \to \quad (E)$$
  - Therefore
$$\text{FOLLOW}(E) = \{\$,)\}.$$

Example

Example (FOLLOW)

- Find FOLLOW(E').
  - E' occurs on the right in two productions.

$$\begin{aligned} E &\rightarrow T\ E' \\ E' &\rightarrow +\ T\ E' \end{aligned}$$

  - Therefore,

$$\text{FOLLOW}(E') = \text{FOLLOW}(E) = \{\$,)\,\}.$$

Example

Example (FOLLOW)

- Find FOLLOW(T).
  - T occurs on the right in two productions.

$$\begin{array}{rcl} E & \rightarrow & T\ E' \\ E' & \rightarrow & +\ T\ E' \end{array}$$

  - Therefore, FOLLOW$(T) \supseteq$ FIRST$(E') = \{+\}$.
  - However, E' is nullable, therefore it also contains FOLLOW$(E) = \{\$,)\}$ and FOLLOW$(E') = \{\$,)\}$.
  - Therefore,

$$\text{FOLLOW}(T) = \{+, \$,)\}.$$

Example

Example (FOLLOW)

- Find FOLLOW(T').
  - T' occurs on the right in two productions.

$$\begin{array}{rcl} T & \rightarrow & F\ T' \\ T' & \rightarrow & *\ F\ T' \end{array}$$

  - Therefore,

$$\text{FOLLOW}(T') = \text{FOLLOW}(T) = \{\$, \texttt{)}\,, +\}.$$

Example

Example (FOLLOW)

- Find FOLLOW(F).
  - F occurs on the right in two productions.

$$\begin{aligned} T &\rightarrow F\ T' \\ T' &\rightarrow *\ F\ T'. \end{aligned}$$

  - Therefore,

$$\text{FOLLOW}(F) \supseteq \text{FIRST}(T') = \{*\}.$$

  - However, T' is nullable, therefore it also contains FOLLOW$(T) = \{+, \$,)\}$ and FOLLOW$(T') = \{\$,), +\}$.
  - Therefore,

$$\text{FOLLOW}(F) = \{*, \$,), +\}.$$

Example

Example (FOLLOW)

Nonterminal	Nullable	FIRST	FOLLOW
E	No	$\{\,(\,,\mathbf{id},\mathbf{num}\}$	$\{\$,\,)\,\}$
E'	Yes	$\{+\}$	$\{\$,\,)\,\}$
T	No	$\{\,(\,,\mathbf{id},\mathbf{num}\}$	$\{\$,\,)\,,+\}$
T'	Yes	$\{*\}$	$\{\$,\,)\,,+\}$
F	No	$\{\,(\,,\mathbf{id},\mathbf{num}\}$	$\{*,\$,\,)\,,+\}$

Assignment

Homework

- The grammar

$$R \rightarrow R \cup R \mid RR \mid R^* \mid (R) \mid \mathbf{a} \mid \mathbf{b}$$

  generates all regular expressions on the alphabet $\Sigma = \{\mathbf{a}, \mathbf{b}\}$.
- Using the result of the exercise from the previous lecture, find FIRST(X) and FOLLOW(X) for each nonterminal X in the grammar.