FIRST and FOLLOW sets

2013-10-05 23:10:47 Partha Biswas

FIRST(X) for all grammar symbols X

Apply following rules:

1. If X is terminal, FIRST(X) = {X}.
2. If $X \rightarrow \varepsilon$ is a production, then add ε to FIRST(X).
3. If X is a non-terminal, and $X \rightarrow Y_1\ Y_2 \ldots Y_k$ is a production, and ε is in all of FIRST(Y_1), …, FIRST(Y_k), then add ε to FIRST(X).
4. If X is a non-terminal, and $X \rightarrow Y_1\ Y_2 \ldots Y_k$ is a production, then add a to FIRST(X) if for some i, a is in FIRST(Y_i), and ε is in all of FIRST(Y_1), …, FIRST(Y_{i-1}).

Applying rules 1 and 2 is obvious. Applying rules 3 and 4 for FIRST($Y_1\ Y_2 \ldots Y_k$) can be done as follows:

Add all the non-ε symbols of FIRST(Y_1) to FIRST($Y_1\ Y_2 \ldots Y_k$). If $\varepsilon \in$ FIRST(Y_1), add all the non-ε symbols of FIRST(Y_2). If $\varepsilon \in$ FIRST(Y_1) and $\varepsilon \in$ FIRST(Y_2), add all the non-ε symbols of FIRST(Y_3), and so on. Finally, add ε to FIRST($Y_1\ Y_2 \ldots Y_k$) if $\varepsilon \in$ FIRST(Y_i), for all $1 \le i \le k$.

Example:

Consider the following grammar.

```
E → E + T | T

T → T * F | F

F → (E) | id
```

Grammar after removing left recursion:

```
E → TX

X → +TX | ε

T → FY

Y → *FY | ε

F → (E) | id
```

For the above grammar, following the above rules, the FIRST sets could be computed as follows:

FIRST(E) = FIRST(T) = FIRST(F) = {(, id}

FIRST(X) = {+, ε}

FIRST(Y) = {*, ε}

FOLLOW(A) for all non-terminals A

Apply the following rules:

1. If \$ is the input end-marker, and S is the start symbol, $\$ \in$ FOLLOW(S).
2. If there is a production, $A \rightarrow \alpha B \beta$, then (FIRST(β) – ε) ⊆ FOLLOW(B).
3. If there is a production, $A \rightarrow \alpha B$, or a production $A \rightarrow \alpha B \beta$, where $\varepsilon \in$ FIRST(β), then FOLLOW(A) ⊆ FOLLOW(B).

Note that unlike the computation of FIRST sets for non-terminals, where the focus is on *what a non-terminal generates*, the computation of FOLLOW sets depends upon *where the non-terminal appears on the RHS of a production*.

Example:

For the above grammar, the FOLLOW sets can be computed by applying the above rules as follows.

FOLLOW(E) = {\$,)}

FOLLOW(E) ⊆ FOLLOW(X) [in other words, FOLLOW(X) contains FOLLOW(E)]

Since there is no other rule applicable to FOLLOW(X),

FOLLOW(X) = {\$,)}

FOLLOW(T) ⊆ FOLLOW(Y) …. (1)

(FIRST(X) – ε) ⊆ FOLLOW(T) i.e., {+} ⊆ FOLLOW(T) …. (2)

Also, since ε ∈ FIRST(X), FOLLOW(E) ⊆ FOLLOW(T)

i.e., {\$,)} ⊆ FOLLOW(T) …. (3)

Putting (2) and (3) together, we get:

FOLLOW(T) = {$,), +}

Since, there is no other rule applying to FOLLOW(Y), from (1), we get:

FOLLOW(Y) = {$,), +}

Since $\varepsilon \in$ FIRST(Y), FOLLOW(T) $\subseteq$ FOLLOW(F) and FOLLOW(Y) $\subseteq$ FOLLOW(F). Also, (FIRST(Y) – ε) $\subseteq$ FOLLOW(F). Putting all these together:

FOLLOW(F) = FOLLOW(T) $\cup$ FOLLOW(Y) $\cup$ (FIRST(Y) – ε) = {$,), +, *}