FIRST and FOLLOW sets

To compute $FIRST(X)$ for all grammar symbols *X*, apply the following rules until no more terminals or $ϵ$ can be added to any FIRST set.

1. If *X* is a terminal, then $FIRST\left(X\right)=\left\{X\right\}$
2. If *X* is a nonterminal, and $FIRST\left(X\right)=Y\_{1}Y\_{2}…Y\_{n}$ is a production rule, then
	* everything in $FIRST\left(Y\_{1}\right)$ is in $FIRST\left(X\right)$
	* if $Y\_{1}\rightarrow ϵ$ contains, then everything in $FIRST\left(Y\_{2}\right)$ is also in $FIRST\left(X\right)$
	* repeat for $Y\_{2}\rightarrow ϵ$, and so on…
3. If $X\rightarrow ϵ$, then add $ϵ$ to $FIRST\left(X\right)$

To compute $FOLLOW(A)$ for all nonterminals *A*, apply the following rules until nothing can be added to any FOLLOW set.

1. Place $ in $FOLLOW\left(S\right)$, where *S* is the start symbol, and $ is the input right end-marker (i.e. end of the program code).
2. If there is a production rule $X\rightarrow αAβ$, then everything in $FIRST\left(β\right)$ except $ϵ$ is in $FOLLOW\left(A\right)$
3. If there is a production rule $X\rightarrow αA$, or a production $X\rightarrow αAβ$ where $FIRST\left(β\right)$ contains $ϵ$, then everything in $FOLLOW\left(X\right)$ is in $FOLLOW\left(A\right)$

Given our LL grammar:

$$S \rightarrow E$$

$$E \rightarrow TE' $$

$E^{'}\rightarrow +TE^{'} \left| -TE^{'} \right| ϵ$

$$T \rightarrow FT' $$

$T'\rightarrow \*FT^{'} \left| / FT^{'} \right| ϵ$

$$F \rightarrow \left(E\right) | id$$

Compute the FIRST sets:

* $FIRST\left(F\right)= \left\{\begin{matrix} (,&id\end{matrix} \right\}$
* $FIRST\left(T'\right)= \left\{\begin{matrix} \* ,&/,&ϵ\end{matrix} \right\}$
* $FIRST\left(T\right)= FIRST\left(F\right)= \left\{\begin{matrix} (,&id\end{matrix} \right\}$
* $FIRST\left(E'\right)= \left\{\begin{matrix} + ,&-,&ϵ\end{matrix} \right\}$
* $FIRST\left(E\right)= FIRST\left(T\right) = \left\{\begin{matrix} (,&id\end{matrix} \right\}$
* $FIRST\left(S\right)= FIRST\left(E\right) = \left\{\begin{matrix} (,&id\end{matrix} \right\}$

Compute the FOLLOW sets:

* $FOLLOW\left(S\right)= \left\{ \$ \right\}$
* $FOLLOW\left(E\right)= \left\{ ) \right\} ∪ FOLLOW\left(S\right) = \left\{\begin{matrix} ),&\$\end{matrix} \right\}$
* $FOLLOW\left(E'\right)= FOLLOW\left(E\right)= \left\{\begin{matrix} ),&\$\end{matrix} \right\}$
* $FOLLOW\left(T\right)= FIRST\left(E'\right) ∪ FOLLOW\left(E/E'\right)= \left\{\begin{matrix} \begin{matrix}+,&-,&)\end{matrix},&\$\end{matrix} \right\}$
* $FOLLOW\left(T'\right)= FOLLOW\left(T\right)= \left\{\begin{matrix} \begin{matrix}+,&-,&)\end{matrix},&\$\end{matrix} \right\}$
* $FOLLOW\left(F\right)= FIRST\left(T'\right) ∪ FOLLOW\left(T/T'\right) = \left\{\begin{matrix} \begin{matrix}\begin{matrix}\*,&/,&+\end{matrix},&-,&)\end{matrix},&\$\end{matrix} \right\}$

Generate the predictive parsing table, $M\left[A, α\right]$, (for LL grammars) from the FIRST/FOLLOW sets.

For each production rule $A\rightarrow α$ of the grammar, do the following:

1. For each terminal $a$ in $FIRST\left(A\right)$, add $A\rightarrow α$ to $M\left[A, a\right]$
2. If $ϵ$ is in $FIRST\left(α\right)$, then for each terminal *b* in $FOLLOW\left(A\right)$, add $A\rightarrow α$ to $M\left[A, b\right]$.
3. Every other entry in the parsing table, $M\left[A, a\right]$, implicitly generates an **error**

|  |  |
| --- | --- |
| Non-Terminal | Input Symbol |
| id | + | - | \* | / | ( | ) | $ |
| *S* | $$S\rightarrow E$$ |  |  |  |  | $$S\rightarrow E$$ |  |  |
| *E* | $$E\rightarrow TE'$$ |  |  |  |  | $$E\rightarrow TE'$$ |  |  |
| *E’* |  | $$E^{'}\rightarrow +TE'$$ | $$E^{'}\rightarrow -TE'$$ |  |  |  | $$E^{'}\rightarrow ϵ$$ | $$E^{'}\rightarrow ϵ$$ |
| *T* | $$T\rightarrow FT'$$ |  |  |  |  |  |  |  |
| *T’* |  |  |  | $$T^{'}\rightarrow \*FT'$$ | $$T^{'}\rightarrow /FT'$$ |  | $$T^{'}\rightarrow ϵ$$ | $$T^{'}\rightarrow ϵ$$ |
| *F* | $$F\rightarrow id$$ |  |  |  |  |  |  |  |