FIRST Sets

Dr. Mattox Beckman

University of Illinois at Urbana-Champaign
Department of Computer Science

Objectives

- Compute the FIRST sets for the nonterminal symbols of a grammar.

The Problem

- Given a grammar for a language L, how can we recognize a sentence in L?
- Solution: Divide and conquer: Given a symbol E …
  - What symbols indicate that the symbol E is just starting? (FIRST Set)
  - What symbols should we expect to see after we have finished parsing an E?

Misleadingly simple example:

$$\begin{array}{ll} S \to xEy & \text{FIRST}(E) = \{z, q\} \\ E \to zE & \text{FOLLOW}(E) = \{y\} \\ E \to q & \end{array}$$

- Important because a parser can see only a few tokens at once.

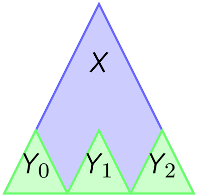
Algorithm

We can compute the FIRST set by a simple iterative algorithm.
For each symbol X:

1. If X is a terminal, then $FIRST(X) = \{X\}$.
2. If there is a production $X \to \epsilon$, then add ϵ to $FIRST(X)$.
3. If there is a production $X \to Y_1 Y_2 \cdots Y_n$, then add $FIRST(Y_1 Y_2 \cdots Y_n)$ to $FIRST(X)$:
   - If $FIRST(Y_1)$ does not contain ϵ, then $FIRST(Y_1 Y_2 \cdots Y_n) = FIRST(Y_1)$.
   - Otherwise, $FIRST(Y_1 Y_2 \cdots Y_n) = FIRST(Y_1)/\epsilon \cup FIRST(Y_2 \cdots Y_n)$.
   - If all of $Y_1, Y_2, \ldots Y_n$ have ϵ then add ϵ to $FIRST(X)$.

Diagram

$$X \to Y_0\ Y_1\ Y_2$$



- If there is a production $X \to Y_1 Y_2 \cdots Y_n$, then add $FIRST(Y_1 Y_2 \cdots Y_n)$ to $FIRST(X)$:
  - If $FIRST(Y_1)$ does not contain ϵ, then $FIRST(Y_1 Y_2 \cdots Y_n) = FIRST(Y_1)$.
  - Otherwise, $FIRST(Y_1 Y_2 \cdots Y_n) = FIRST(Y_1)/\epsilon \cup FIRST(Y_2 \cdots Y_n)$.
  - If all of $Y_1, Y_2, \ldots Y_n$ have ϵ then add ϵ to $FIRST(X)$.

Small Examples

Example 1

$S \to x\ A\ B$

FIRST set of S is $\{x\}$.

Example 2

$A \to \epsilon$

$A \to y$

$A \to z\ q$

FIRST set of A is $\{y, z, \epsilon\}$.

Example 3

$B \to A\ q$

$B \to r$

FIRST set of B is $\{y, z, q, r\}$.

Example 4

$C \to A\ A$

$C \to B$

FIRST set of C is $\{y, z, q, r, \epsilon\}$.

FIRST Set Example

Grammar

$S \rightarrow \texttt{if}\ E\ \texttt{then}\ S\ ;$

$S \rightarrow \texttt{print}\ E;$

$E \rightarrow E + E$

$E \rightarrow P\ \texttt{id}$

$P \rightarrow * P$

$P \rightarrow \epsilon$

Result

S={}

E={}

P={}

Action

Step 1: Create a list of symbols.

FIRST Set Example

Grammar

$S \rightarrow \texttt{if}\ E\ \texttt{then}\ S\ ;\ \Leftarrow$

$S \rightarrow \texttt{print}\ E;\ \Leftarrow$

$E \rightarrow E + E$

$E \rightarrow P\ \texttt{id}$

$P \rightarrow *\ P\ \Leftarrow$

$P \rightarrow \epsilon\ \Leftarrow$

Result

S={if, print }

E={}

P={ϵ, *}

Action

Step 2: Add terminals starting productions, and all ϵ.

FIRST Set Example

Grammar

$S \rightarrow \texttt{if}\ E\ \texttt{then}\ S\ ;$

$S \rightarrow \texttt{print}\ E;$

$E \rightarrow E + E$

$E \rightarrow P\ \texttt{id} \Leftarrow$

$P \rightarrow * P$

$P \rightarrow \epsilon$

Result

S={$\texttt{if}, \texttt{print}$ }

E={$*$, $\texttt{id}$}

P={ϵ, $*$}

Action

Step 3: Check productions. Add $\mathit{FIRST}(P\texttt{id})$ to $\mathit{FIRST}(E)$.

FIRST Set Example

Grammar

$S \rightarrow \texttt{if}\ E\ \texttt{then}\ S\ ;$

$S \rightarrow \texttt{print}\ E;$

$E \rightarrow E + E \Leftarrow$

$E \rightarrow P\ \texttt{id}$

$P \rightarrow * P$

$P \rightarrow \epsilon$

Result

S={$\texttt{if}, \texttt{print}$}

E={$*$, $\texttt{id}$}

P={ϵ, $*$}

Action

Step 4: Check productions: $E \rightarrow E + E$ adds nothing. We're done.

Another FIRST Set Example

Grammar

$S \rightarrow A\mathrm{x}$
$S \rightarrow B\mathrm{y}$
$S \rightarrow \mathrm{z}$
$A \rightarrow 1CB$
$A \rightarrow 2B$
$B \rightarrow 3B$
$B \rightarrow C$
$C \rightarrow 4$
$C \rightarrow \epsilon$

Result

S ={}
A={}
B ={}
C={}

Action

Create a chart.

Another FIRST Set Example

Grammar

$S \to Ax$
$S \to By$
$S \to z \Leftarrow$
$A \to 1CB \Leftarrow$
$A \to 2B \Leftarrow$
$B \to 3B \Leftarrow$
$B \to C$
$C \to 4 \Leftarrow$
$C \to \epsilon \Leftarrow$

Result

S ={ z}
A={ 1, 2}
B ={ 3}
C={ ϵ, 4}

Action

Add initial terminals and ϵs.

Another FIRST Set Example

Grammar

$S \rightarrow A\mathrm{x} \Leftarrow$
$S \rightarrow B\mathrm{y}$
$S \rightarrow \mathrm{z}$
$A \rightarrow 1CB$
$A \rightarrow 2B$
$B \rightarrow 3B$
$B \rightarrow C$
$C \rightarrow 4$
$C \rightarrow \epsilon$

Result

S ={z, 1, 2}
A={1, 2}
B ={3}
C={ϵ, 4}

Action

Add $FIRST(A\mathrm{x})$ to $FIRST(S)$.

Another FIRST Set Example

Grammar

$S \rightarrow A\mathrm{x}$
$S \rightarrow B\mathrm{y}$ $\Leftarrow$
$S \rightarrow \mathrm{z}$
$A \rightarrow 1CB$
$A \rightarrow 2B$
$B \rightarrow 3B$
$B \rightarrow C$
$C \rightarrow 4$
$C \rightarrow \epsilon$

Result

S ={z, 1, 2, 3}
A={1, 2}
B ={3}
C={ϵ, 4}

Action

Add $FIRST(B\mathrm{y})$ to $FIRST(S)$. Note that there is still more to be added to $FIRST(B)$! We will

Another FIRST Set Example

Grammar

$S \to A\mathrm{x}$
$S \to B\mathrm{y}$
$S \to \mathrm{z}$
$A \to 1CB$
$A \to 2B$
$B \to 3B$
$B \to C \Leftarrow$
$C \to 4$
$C \to \epsilon$

Result

S ={z, 1, 2, 3}
A={1, 2}
B ={3, 4, ϵ}
C={ϵ, 4}

Action

Add $FIRST(C)$ to $FIRST(B)$. At this point we should iterate again to see if anything changes.

Another FIRST Set Example

Grammar

$S \rightarrow A\mathtt{x}$ $\Leftarrow$

$S \rightarrow B\mathtt{y}$

$S \rightarrow \mathtt{z}$

$A \rightarrow \mathtt{1}CB$

$A \rightarrow \mathtt{2}B$

$B \rightarrow \mathtt{3}B$

$B \rightarrow C$

$C \rightarrow \mathtt{4}$

$C \rightarrow \epsilon$

Result

S ={z, 1, 2, 3}

A={1, 2}

B ={3, 4, ϵ }

C={ϵ, 4}

Action

Add $FIRST(A\mathtt{x})$ to $FIRST(S)$ again. Nothing happens ...

Another FIRST Set Example

Grammar

$S \rightarrow A\mathrm{x}$
$S \rightarrow B\mathrm{y} \Leftarrow$
$S \rightarrow \mathrm{z}$
$A \rightarrow 1CB$
$A \rightarrow 2B$
$B \rightarrow 3B$
$B \rightarrow C$
$C \rightarrow 4$
$C \rightarrow \epsilon$

Result

S ={z, 1, 2, 3, 4, y}
A={1, 2}
B ={3, 4, ϵ }
C={ϵ, 4}

Action

Add $FIRST(B\mathrm{y})$ to $FIRST(S)$ again. The 4 gets propagated. Since B could be ϵ we need to add

Another FIRST Set Example

Grammar

$S \rightarrow A\mathrm{x}$
$S \rightarrow B\mathrm{y}$
$S \rightarrow \mathrm{z}$
$A \rightarrow 1CB$
$A \rightarrow 2B$
$B \rightarrow 3B$
$B \rightarrow C \Leftarrow$
$C \rightarrow 4$
$C \rightarrow \epsilon$

Result

S ={z, 1, 2, 3, 4, y}
A={1, 2}
B={3, 4, ϵ }
C={ϵ, 4}

Action

Add $FIRST(C)$ to $FIRST(B)$ again. We are done.