First Sets

Dr. Mattox Beckman

Illinois Institute of Technology
Department of Computer Science

Objectives

- Be able to explain the purpose of a first set.
- Be able to compute the first set.

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:

$$\begin{array}{ll} S \rightarrow xEy & \text{First}(E) = \{z, q\} \\ E \rightarrow zE & \text{Follow}(E) = \{y\} \\ E \rightarrow q & \end{array}$$

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

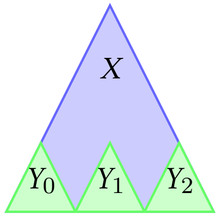
Algorithm

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

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

$S \to$ `if` E `then` S ;
$S \to$ `print` E;
$E \to E + E$
$E \to P$ `id`
$P \to * P$
$P \to \epsilon$

Result

S={}
E={}
P={}

Action

Step 1: Create a list of symbols.

First Set Example

Grammar

$S \to$ `if` E `then` S ; $\Leftarrow$
$S \to$ `print` E; $\Leftarrow$
$E \to E + E$
$E \to P$ `id`
$P \to * P \Leftarrow$
$P \to \epsilon \Leftarrow$

Result

S={`if`, `print`}
E={}
P={ϵ, $*$}

Action

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

First Set Example

Grammar

$S \to$ `if` E `then` S ;
$S \to$ `print` E;
$E \to E + E$
$E \to P\,$`id` $\Leftarrow$
$P \to * \, P$
$P \to \epsilon$

Result

S={`if`, `print` }
E={$*$, `id`}
P={ϵ, $*$}

Action

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

First Set Example

Grammar

$S \to$ `if` E `then` S ;
$S \to$ `print` E;
$E \to E + E$ $\Leftarrow$
$E \to P\,$`id`
$P \to * \, P$
$P \to \epsilon$

Result

S={`if`, `print` }
E={$*$, `id`}
P={ϵ, $*$}

Action

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

Another First Set Example

Grammar

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

Result

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

Action

Create a chart.

Another First Set Example

Grammar

$S \to A$x
$S \to B$y
$S \to$ 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

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

Result

S ={z, 1, 2}
A={1, 2}
B ={3}
C={ϵ, 4}

Action

Add *First*(Ax) to *First*(S).

Another First Set Example

Grammar

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

Result

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

Action

Add *First*(By) to *First*(S). Note that there is still more to be added to *First*(B)! We will have to revisit this step later.

Another First Set Example

Grammar

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

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

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

Result

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

Action

Add *First*(Ax) to *First*(S) again. Nothing happens...

Another First Set Example

Grammar

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

Result

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

Action

Add $First(By)$ to $First(S)$ again. The 4 gets propagated. Since B could be ϵ we need to add y.

Another First Set Example

Grammar

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

Result

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

Action

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