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:

$S \rightarrow xEy$	FIRST(E)	$=\{z, q\}$
$E \rightarrow zE$	FOLLOW(E)	$=\{y\}$
$E \rightarrow q$		

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

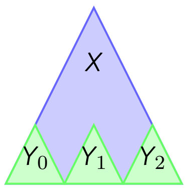
Algorithm

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

$$X \rightarrow Y_0\, Y_1\, Y_2$$



- 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 $FIRST(X)$.

Small Examples

Example 1

$S \rightarrow x\, A\, B$

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

Example 2

$A \rightarrow \epsilon$
$A \rightarrow y$
$A \rightarrow z\, q$

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

Example 3

$B \rightarrow A\, q$
$B \rightarrow r$

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

Example 4

$C \rightarrow A\, A$
$C \rightarrow B$

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

FIRST Set Example

Grammar

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

Result

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

Action

Step 1: Create a list of symbols.

FIRST Set Example

Grammar

$S \rightarrow$ `if` E `then` S ; $\Leftarrow$
$S \rightarrow$ `print` E; $\Leftarrow$
$E \rightarrow E + E$
$E \rightarrow P$ `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

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

Result

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

Action

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

Objectives FIRST Sets Examples

FIRST Set Example

Grammar

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

Result

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

Action

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

Objectives FIRST Sets Examples

Another FIRST Set Example

Grammar

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

Result

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

Action

Create a chart.

Objectives FIRST Sets Examples

Another FIRST Set Example

Grammar

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

Result

S ={ z}
A={ 1, 2}
B ={ 3}
C={ ϵ, 4}

Action

Add initial terminals and ϵs.

Another FIRST Set Example

Grammar

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

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

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

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

Another FIRST Set Example

Grammar

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

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

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

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

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

Result

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

Action

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

Another FIRST Set Example

Grammar

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

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.