Compilers

Arthur Hoskey, Ph.D.
Farmingdale State College
Computer Systems Department

- First Sets

Today's Lecture

- First set. Indicates which terminals begin the strings derivable from a nonterminal.

- First(A). This is the set of terminals derivable from the nonterminal A.

- Calculate first sets for the following productions (lowercase are terminals):

A → stu **s is a terminal and begins the string to it is in First(A)**
B → xyz **x is a terminal and begins the string to it is in First(B)**

First(A) = { s }
First(B) = { x }

First Set

- To find the first set of a nonterminal we must take into account ALL productions with the nonterminal on the LHS.

- Calculate first sets for the following productions:

A → stu
A → pq
B → xyz

Check both A productions when calculating First(A). s begins one and p begins the other. Both s and p should be in the First(A)

First Set – Multiple Productions for Non-terminal

- To find the first set of a nonterminal we must take into account ALL productions with the nonterminal on the LHS.

- Calculate first sets for the following productions:

A → stu
A → pq

Check both A productions when calculating First(A). s begins one and p begins the other. Both s and p should be in the First(A)

B → xyz

<u>ANSWER</u>
First(A) = { s, p }
First(B) = { x }

First Set – Multiple Productions for Non-terminal

- If a non-terminal begins the RHS then you must add that non-terminal's first set.

- Calculate first sets for the following productions:

A → stu
A → Bq
B → xyz

B appears as the leftmost symbol in an A production we must include First(B) in First(A).

First Set – Non-terminals in RHS

- If a non-terminal begins the RHS then you must add that non-terminal's first set.

- Calculate first sets for the following productions:

A → stu

A → Bq   **B is a non-terminal so First(B) must be added to First(A)**

B → xyz

                   x is from First(B)

ANSWER
First(A) = { s, x }
First(B) = { x }

First Set – Non-terminals in RHS

- If a non-terminal can derive an empty string (λ) then the empty string must be added to the first set.

- Calculate first sets for the following productions:

A → stu
A → λ   **A can derive the empty string (λ)**
                so λ must be included in First(A)
B → xyz

Note: Both lambda (λ) and epsilon (ε) can be used to signify an empty string.

First Set – Empty Strings

- If a non-terminal can derive an empty string (λ) then the empty string must be added to the first set.

- Calculate first sets for the following productions:

A → stu   **A can derive the empty string (λ)**
A → λ   ←   **so λ must be included in First(A)**
B → xyz

ANSWER
First(A) = { s, λ }
First(B) = { x }

First Set – Empty Strings

- If the first RHS symbol is a non-terminal, and that non-terminal can derive λ then you must also include the first set of the next symbol minus λ.

- Calculate first sets for the following productions:

A → Bstu
B → xyz
B → λ

First Set – Empty Strings

- If the first RHS symbol is a non-terminal, and that non-terminal can derive λ then you must also include the first set of the next symbol.

- Calculate first sets for the following productions:

A → Bstu

B → xyz

B → λ

A must include First(B)-λ. We must also take into account that B can derive λ. Since B can derive λ we must also include the first set of the symbol that follows B (this is the terminal s).

ANSWER

First(A) = { s, x }

First(B) = { x, λ }

First(A) does NOT contain λ. This is because there is no way to derive λ from A.

In general, λ should only be included if ALL symbols in the rhs can derive λ.

First Set – Empty Strings

- What are the first sets for the nonterminals A, B, and C?

A → BCdef
B → mn
B → λ
C → qr
C → λ

First Set – Example

- What are the first sets for the nonterminals A, B, and C?

A → BCdef   ←

B → mn

B → λ

C → qr

C → λ

<u>**Calculating First(A)**</u>
B is a nonterminal so First(B) is included.
B can derive λ so we must also include First(C)-λ.
C can derive λ so we must include First(g) (g is a terminal so it only contains g).

Answer

First(A) = { m, q, d }   ←

First(B) = { m, λ }

First(C) = { q, λ }

First(A) does NOT contain λ. This is because there is no way to derive λ from A.

In general, λ should only be included if ALL symbols in the rhs can derive λ.

First Set – Example

- What are the first sets for the nonterminals A, B, C, and D?

A → BCD
B → mn
B → λ
C → qr
C → λ
D → st
D → λ

First Set – Example

- What are the first sets for the nonterminals A, B, C, and D?

A → BCD
B → mn
B → λ
C → qr
C → λ
D → st
D → λ

Calculating First(A)
B is a nonterminal so First(B)-λ is included.
B can derive λ so we must also include First(C)-λ.
C can derive λ so we must also include First(D)-λ.
D is at the end of the production A→BCD and can derive λ so now include λ in the First(A).

First(A) includes λ because all symbols on the rhs of A productions can derive λ

Answer
First(A) = { m, q, s, **λ** }
First(B) = { m, λ }
First(C) = { q, λ }
First(D) = { s, λ }

First Set – Example

- **End of Slides**

End of Slides