

Algorithms: FIRST and FOLLOW Sets

CS236 - Discrete Structures
Brett Decker and Michael Goodrich
FALL 2021

Terminology

Before discussing the algorithms, let's review terminology. The set T defines all *terminals* of the grammar. The *nonterminals* are defined either by the set N or by looking at the vocabulary (V) and removing all terminals. Everything in the vocabulary is called a symbol. Thus, all nonterminals and terminals are symbols. The set P defines all *productions*. A production consists of the *left-hand side* and *right-hand side* of the production arrow (\rightarrow). In the grammars we care about—*context-free grammars*—the left-hand side will only contain a single, **nonterminal**. Here is an example production: $A \rightarrow aB$

A is a nonterminal because it is on the left-hand side of the production. aB is on the right-hand side of the production. For this document, we will use the convention that upper-case letters are nonterminals and lower-case letters are terminals. When in doubt, use T , V or N , and the left-hand side of productions to determine which symbols are terminals and which are nonterminals.

Algorithm 1: Computing FIRST sets

We compute FIRST sets on productions. When computing the FIRST set of a the production use these steps (let's say A is the nonterminal on the *right-hand* side of the production):

- 1) Look at the leftmost symbol on the *right-hand* side of the production
- 2.1) If the leftmost symbol is a terminal then
that terminal is the only symbol in the FIRST set of the production $A \rightarrow \dots$
- 2.2) Else, if the leftmost symbol is a nonterminal (let's call it B) then
add all terminals in the FIRST sets of all productions with B on the *left-hand* side into the FIRST set of the original production $A \rightarrow B \dots$

Algorithm 2: Computing FOLLOW sets

When computing the FOLLOW set of a nonterminal A use these steps:

- 1) Find all the productions that *produce* the nonterminal A
- 2) For each production in step 1, find all places where the nonterminal A is produced and what single symbol is produced immediately afterward
- 3.1) For each symbol found in step 2, If the symbol is a terminal, then
add that terminal into the FOLLOW set of A
- 3.2) Else, if the symbol is a nonterminal (let's call it B), then
add each element in the FIRST set of B into the FOLLOW set of A
- 3.3) Else, if there is no symbol, go up another level, by repeating steps 1-3 for B
(if we go up to the starting nonterminal then add the end of input symbol, $\#$, to the FOLLOW set of A)