

CSEN1002 Compilers Lab, Spring Term 2021  
Task 6: First and Follow

Due: Week starting 05.06.2021

## 1 Objective

For this task you will implement the algorithms computing the functions *First* and *Follow*, introduced in Lecture 4 of CSEN1003, for the *variables* of a given context-free grammar. Recall that a CFG is a quadruple  $(V, \Sigma, R, S)$  where  $V$  and  $\Sigma$  are disjoint alphabets (respectively, containing *variables* and *terminals*),  $R \subseteq V \times (V \cup \Sigma)^*$  is a set of *rules*, and  $S \in V$  is the *start variable*.

## 2 Requirements

- We make the following assumptions about input CFGs for simplicity.
  - a) The set  $V$  of variables consists of upper-case English symbols.
  - b) The start variable is the symbol  $S$ .
  - c) The set  $\Sigma$  of terminals consists of lower-case English symbols other than “e”.
  - d) The letter “e” represents  $\varepsilon$ .
- You should implement two functions, **First** and **Follow**, which take an input string encoding a CFG and return a string encoding of the *First*, respectively the *Follow*, set of each variable of the grammar.
- A string encoding a CFG is a semi-colon-separated sequence of items. Each item represents a largest set of rules with the same left-hand side and is a comma-separated sequence of strings. The first string of each item is a member of  $V$ , representing the common left-hand side. The first string of the first item is  $S$ .
- For example, consider the CFG  $(\{S, T, L\}, \{i, a, b, c, d\}, R, S)$ , where  $R$  is given by the following productions.

$$\begin{aligned} S &\longrightarrow S c T \mid T \\ T &\longrightarrow a S b \mid i a L b \mid \varepsilon \\ L &\longrightarrow S d L \mid S \end{aligned}$$

This CFG will have the following string encoding.

S, ScT, T; T, aSb, iaLb, e; L, SdL, S

- The output of each of **First** and **Follow** is, similar to the input, a semi-colon-separated sequence of items, where each item is a comma-separated pair. The first element of each pair is a variable of the grammar and the second element is a string representing the *First* or, respectively, the *Follow* set of that variable. The symbols in these strings should appear in alphabetical order. (\$) always appears last.) The items themselves should appear in the order in which their respective variables appear in the input CFG.
- For example, the result of calling **First** on the above CFG may have the following form

S, acei; T, aei; L, acdei

Similarly, the result of calling **Follow** may be as follows

S, bcd\$; T, bcd\$; L, b

### 3 Evaluation

- Your implementation will be tested by running **First** and **Follow** on five CFGs.
- You get one point for each correct output; hence, a maximum of ten points.

### 4 Online Submission

- You should submit your code at the following link.

<https://forms.gle/9PwVnsqxMWMG5uMD9>

- Submit one Java file (.java) containing executable code.
- The filename should be of form <LabNo\_ID\_Name>.  
For example, P13\_40\_1234\_John\_Smith.java
- Online submission is due on Thursday, June 3rd, by 23:59.