German University in Cairo
Department of Computer Science
Assoc. Prof. Haythem O. Ismail

CSEN1002 Compilers Lab, Spring Term 2022
Task 6: First and Follow

Due: Week starting 07.05.2022

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, Σ, R, S) where V and Σ 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 Σ of terminals consists of lower-case English symbols other than "`e`".
  d) The letter "`e`" represents ε.
- You should implement a class constructor `FFCFG`, which takes an input string encoding a CFG, and two functions, `First` and `Follow`, which 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 $(\{S, T, L\}, \{\mathtt{i}, \mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}\}, R, S)$, where R is given by the following productions.

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

  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`

- Important Details:
  - Your implementation should be done within the template file "`FFCFG.java`" (uploaded to the CMS).
  - You are not allowed to change package, file, constructor, or method names/signatures.
  - You are allowed to implement as many helper classes/methods within the same file (if needed).
  - Public test cases have been provided on the CMS for you to test your implementation.
  - Please ensure that the public test cases run correctly without modification before coming to the lab to maintain a smooth evaluation process.
  - Private test cases will be uploaded before your session and will have the same structure as the public test cases.

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/3q8pRA6kK3KaQtLq7

- Submit one Java file (`FFCFG.java`) containing executable code.
- Online submission is due on Thursday, May , by 23:59.