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

CSEN1002 Compilers Lab, Spring Term 2023
Task 6: Context-Free Grammars First and Follow

Due: Week starting 25.04.2023

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 letters.
  b) The start variable is the symbol S.
  c) The set Σ of terminals consists of lower-case English letters (except the letter `e`).
  d) The letter "`e`" represents ε.
- You should implement a class constructor `CfgFirstFollow`, and two methods; `first`, and `follow`.
- `CfgLeftRecElim`, a class constructor, takes one parameter which is a string description of a CFG and constructs a CFG instance. A string encoding a CFG is of the form $V\#T\#R$.
  - V is a string representation of the set of variables; a semicolon-separated sequence of upper-case English letters, starting with S.
  - T is a string representation of the set of terminals; a semicolon-separated sequence of alphabetically sorted lower-case English letters.
  - R is a string representation of the set of rules. R is a semicolon-separated sequence of pairs. Each pair represents a largest set of rules with the same left-hand side. Pairs are of the form i/j where i is a variable of V and j is a string representation of set of right-hand sides—a comma-separated sequence of strings. These pairs are sorted by the common left-hand side i based on the ordering of V.
- For example, consider the CFG $G_1 = (\{S, T, L\}, \{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}, \mathtt{i}\}, R, S)$, 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;T;L#a;b;c;d;i#S/ScT,T;T/aSb,iaLb,e;L/SdL,S`

- The output of each of `first` and `follow` is a semi-colon-separated sequence of items, where each item is a /-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 first.) 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 G_1 may have the following form

  `S/acei;T/aei;L/acdei`

  Similarly, the result of calling `follow` on G_1 may be as follows

  `S/$bcd;T/$bcd;L/b`

- Important Details:
  - Your implementation should be done within the template file "`CfgFirstFollow.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/kwErNZbz9RdAviRs6`

- Submit one Java file (`CfgFirstFollow.java`) containing executable code.
- **Online submission is due by the end of your lab session.**