Project 2: Instructions

Help Center

Programming Details:

Materials for Download

- **Java Version**
- **Python Version** **(Please use Python 2.7.5)**

In this assignment, you are expected to fill in a program that implements the CYK algorithm for context free grammars. The file-parsing part and data structures have been finished. What is left to fill in is the following method in CYK.java: (Note: For Python version, the names of files, methods and variables are exactly the same as Java version)

```
// CYK algorithm
// Calculate the array X
// w is the string to be processed
void calcCYK(int [] w)
```

Details of the CYK algorithm are in slide 7~12 in lecture 15_cfl5. Basically, the whole program is to read in strings over alphabet {a,b}, and then use the context-free grammar defined on slide 8 of lecture 15_cfl5:
S -> AB
A -> BC | a
B -> AC | b
C -> a | b

For each string w of length L, the program will call the CYK algorithm to calculate the table X defined on slide 7 of lecture 15_cfl5. In other words, you are expected to calculate $X_{ij}, 0 \leq i \leq j \leq L-1$ **(in Java, array is 0-based). To facilitate computation, we use numerical values to represent variables and terminals. Here are the definitions of arrays you will need:**

```
int [][] production = new int[maxProductionNum+1][3];
//Productions in Chomsky Normal Form (CNF)
//production[i][0] is the number for the variable (0~3, 0: S 1: A, 2: B, 3: C)
//If this production is of the form A->BC (two variables), then production[i][1] and production[i][2]
will contain the numbers for these two variables
//If this production is of the form A->a (a single terminal), then production[i][1] will contain the
number for the terminal (0 or 1, 0: a, 1: b), production[i][2]=-1
```

```
boolean [][][] X;
//X[i][j][s]=true if and only if variable s (0~3, 0: S 1: A, 2: B, 3: C) is in X_ij defined in CYK
//Suppose the length of string to be processed is L, then 0 <= i <= j <= L-1
```

Here is a method we have implemented which you may find helpful:
//check whether (a,b,c) exists in production defined above

//e.g. If you want to check whether S->AB is a production, just call existProd(0,1,2); if you want to check whether A->a is a production, just call existProd(1,0,-1)
boolean existProd(int a,int b,int c)

Here's the sample input for your program, which is exactly the same as in slide 12 of lecture 15_cfl5.
sample input (sampleCYK.in):
ababa

Submission Details:

After you fill in the program in CYK.java, you can run Submit.java. This program will first run your CYK.java on sample input, and only after you get it right on samples, can you submit your programs.
In submission, you will select the programs you want to submit (Select "All of the above"), and it will need your website credentials (note that the password should be the assignment password, which you can find when you click "Programming Assignment" on the website). Then, the website will compare your output on testCYK.in with the solution. If it's correct, you will get full score and see "Correct!" on the screen; otherwise, you should go back and check your filled part to make it correct.
Example of what you will see in the terminal screen when you run Submit.java:

```
==
== [automata-class] Submitting Solutions | Programming Exercise 2
==
**********************
Test on sample input (sampleCYK.in).....

You output is correct for sample input. Now start submitting to servers.....
**********************


== Select which part(s) to submit:
== 1) 2-1 [ CYK.java ]
== 2) All of the above
==
Enter your choice [1-2]:
2
Login (Email address): xxxxx@yyyy.com
Password: abcdefg

== Connecting to automata-class ...

== [automata-class] Submitted Homework 2 - Part 1 - 2-1
== Correct!
```