CS251: Introduction to Language Processing

Tierce-1 Exam (2021-22-W Semester)

Max. Points: 100 **Duration:** 1 hour 30 minutes

February 4, 2022

Instructions

- Question paper has of 3 pages containing 3 questions.
- All the questions are compulsory.

Question-1 (Lexical Analysis)

Assume that the IIT Bhilai ID card has a unique ID number consisting of 8 digits as the following.

D1	D2	D3	D4	D5	D6	D7	D8

Details of the Digits:

- The digit D1 represents the type of cardholder as per the following.

Value of D1	**Type of the Card Holder**
1	Regular Student
2	Visiting Student
3	Regular Employee
4	Contract Employee
5-9	Other categories
0	Invalid

- D2 and D3 denote the year of admission for students and year joining for employees.
- D4 through D7 denote the serial number within the group as per the following.
  1. For students (i.e., D1 is either 1 or 2)

Serial Number	**Programme**
0000-1999	Ph.D students
2001-3999	Master students
4000-9999	Under graduate students

2. For employees, it denotes the serial number from 0000 to 9999.

- D8 denotes the dependent family information. For employees, the following table is to be used for D8. For students, D8 should be 0; any other value is considered invalid.

Value	**Semantics**
0	Self
1-9	Dependent

Given the above information, write a lexical analyzer that, given a character stream, it prints the tokens corresponding to it as per the following.

Type of Person	**TOKEN**
Ph.D students	PHD_STUDENT
Master students	MASTER_STUDENT
Under graduate students	UG_STUDENT
Regular employee	EMPL_REG
Dependent of regular employee	EMPL_REG_DEP
Contract employee	EMPL_CON
Dependent of contract employee	EMPL_CON_DEP
Other category type	OTHER
Any invalid ID. For example, invalid characters, incorrect length	INVALID

Note that it is sufficient if you just write down the regular expressions and the corresponding tokens. You need not write the complete lex program.

[20 Points]

Question-2 (Top Down Parsing)

Consider the following grammar:

$$\begin{aligned} S &: SA \\ &\mid b \\ &\mid \epsilon \\ A &: (A) \\ &\mid a \end{aligned}$$

Part-(a):

1. Compute the FIRST and FOLLOW sets for all the non terminals
2. Construct the LL(1) parser table
3. Show that the grammar is not LL(1) with a justification

Part-(b):

1. Rewrite the above grammar such that it becomes LL(1)
2. Write down the FIRST and FOLLOW sets for all the non terminals for the above grammar.
3. Construct the LL(1) parser table
4. Show that the revised grammar is LL(1) with a justification

[**45 Points**]

Question-3 (Bottom Up Parsing)

Consider the following grammar:

$$\begin{aligned} S &: AA \\ A &: aA \\ &\mid b \end{aligned}$$

1. Construct the DFA for LR(0)
2. Construct parser table for LR(0)
3. Is the above grammar LR(0)? Justify your answer.

[**35 Points**]