School of Mathematical and Computational Sciences Indian Association for the Cultivation of Science Compiler Construction: COM 5202

Tutorial III (29 January, 2025)

M. Sc Semester IV: 2024-2025

Instructor: Goutam Biswas

Exercise 1. Consider the regular expression $(a + b)^*abb(a + b)^*$. Concatenate it with the right end marker # and answer the following questions.

- (a) Draw the syntax tree of the concatenated regular expression.
- (b) Label the positions of leaf nodes with a unique number and decorate the nodes of the tree with firstpos() and lastpos() data ((fp,lp)).
- (c) Compute followpos() for different positions.
- (d) Construct the state transition diagram of the NFA.
- (e) Construct the DFA by subset construction.
- (f) Finally identify the equivalent states and draw the minimal DFA.

Exercise 2. Consider the following flex specification. Write actions for three patterns to count the number of characters, words (string over $\{a, \dots, z, A, \dots, Z\}$ and lines in the input file.

```
%{
#include <stdio.h>
#include <string.h>
int ctr=0, word=0, line=0;
%}
%option noyywrap
%%
            {/* Action - 1 */}
[a-zA-Z]+
            {/* Action - 2 */}
\n
            {/* Action - 3 */ }
%%
int main(){
    yylex();
    printf("lines: %d, words: %d, characters: %d\n",
            line, word, ctr);
    return 0;
}
                                                                 [10]
Exercise 3.
```

Design a scanner without using flex) for the following set of tokens. {a, b, ab, aab, abb, aaa, baab, aabba, bbaba}. An input to the scanner is an element of {a, b}⁺ terminated by 'c'. There may

```
be blanks (' ') as separators in between strings of \{a, b\}^+ e.g.
bababba ab babaa bbabbabc.
```

- (a) Draw a deterministic state transition diagram for the set of tokens.
- (b) Write a C program that reads an input and emits the sequence of maximal length tokens.The input file name is passed as a command line argument to the C program.
- (c) The C file name should be <roll no>.3.c. Send the file to goutamamartya@gmail.com.

Following are a few input-output.

Input	Output
ac	$\langle a \rangle$
bc	< b >
aac	< a > < a >
abc	$\langle ab \rangle$
bac	$\langle b \rangle \langle a \rangle$
bbc	< b > < b >
aaac	$\langle aaa \rangle$
aabc	< aab >
abac	$\langle ab \rangle \langle a \rangle$
abbc	< abb >
bbaaac	< b > < b > < aaa >
baabc	< baab >
bbababaababa bbbbaab aababba baaababbbabc	< bbaba > < baab > < ab > < a >
	< b > < b > < b > < b a a b > < a a b >
	$\langle abb \rangle \langle a \rangle \langle b \rangle \langle aaa \rangle \langle b \rangle$
	$\langle abb \rangle \langle b \rangle \langle ab \rangle$