

School of Mathematical and Computational Sciences
Indian Association for the Cultivation of Science

Compiler Construction: COM 5202

Tutorial IV (28 January, 2026)

M. Sc Semester IV: 2025-2026

Instructor: Goutam Biswas

Exercise 1. [10]

Design a scanner without using *flex* for the following set of tokens.

$\{a, b, ba, bba, baa, bbb, abba, bbaab, aabab\}$.

An input to the scanner is an element of $\{a, b\}^+$. There may be blanks (‘ ’) as separators in between strings of $\{a, b\}^+$ e.g. bababba ab babaa bbabbab.

- (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.
- (c) The input file name is passed as a command line argument to the C program.
- (d) Open the file using the system call `int open(const char *pathname, int flags)`. call `int fstat(int fd, struct stat *statbuf)` and get the file size from `statBuf->st_size`.
- (e) Read the entire file in a buffer and perform tokenization.
- (f) Your C file name should be `<roll no>.4.c`. Send the file to `goutamamartya@gmail.com`.

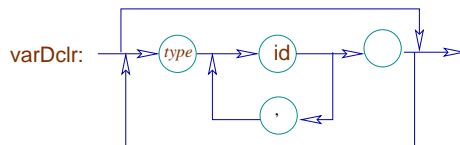
Following are a few input-output pairs:

Input	Output
a	< a >
b	< b >
ba	< ba >
bba	< bba >
baa	< baa >
bbb	< bbb >
abba	< abba >
bbaab	< bbaab >
aabab	< aabab >
babaaababababaaab	< ba >< baa >< a >< ba >
aabbabbabbbababbb	< ba >< ba >< baa >< a >< b >
bbababaababa bbbbaab aababba baaababbbab	< a >< abba >< bba >< bbb >
	< a >< ba >< bbb >
	< bba >< ba >< baa >< ba >
	< ba >< bbb >< baa >< b >
	< aabab >< ba >< baa >< a >
	< ba >< bbb >< a >< b >

Exercise 2. Consider the following language of variable decleration:

```
int a, b, c;
float x, y;
```

(a) A syntax diagram is as follows:



Modify it to include multi-dimension array.

(b) A CFG is as follows:

$$\begin{aligned}
 \langle \text{VDP} \rangle &\rightarrow \varepsilon \mid \langle \text{VD} \rangle \langle \text{VD_OPT} \rangle \\
 \langle \text{VD} \rangle &\rightarrow \langle \text{TYPE} \rangle \text{id} \langle \text{ID_OPT} \rangle \\
 \langle \text{ID_OPT} \rangle &\rightarrow \varepsilon \mid , \text{id} \langle \text{ID_OPT} \rangle \\
 \langle \text{VD_OPT} \rangle &\rightarrow ; \mid ; \langle \text{VD} \rangle \langle \text{VD_OPT} \rangle \\
 \langle \text{TYPE} \rangle &\rightarrow \text{int} \mid \text{float} \mid \dots
 \end{aligned}$$

Modify it to include multi-dimensional array.

(c) A Backus-Naur Form is as follows:

$$\begin{aligned}
 \langle \text{VDP} \rangle &::= \varepsilon \mid \langle \text{VD} \rangle ; \{ \langle \text{VD} \rangle ; \} \\
 \langle \text{VD} \rangle &::= \langle \text{TYPE} \rangle \text{id} \{ , \text{id} \}
 \end{aligned}$$

Modify it to include multi-dimensional array.

Exercise 3.

- (a) What is the minimum number of states including the dead state of a DFA over $\{0, 1, 2, 3\}$ that accepts strings with no repeated digits?
- (b) Find a regular expression of all strings of lower case letters where the letters are in ascending lexicographic order.

Exercise 4.

- (a) Consider the following grammar G_1 with the start symbol E . Draw the parse tree corresponding to the expression $id - ic / id$.

$$\begin{aligned}
 E &\rightarrow E - T \mid T \\
 T &\rightarrow T / F \mid F \\
 F &\rightarrow (E) \mid id \mid ic.
 \end{aligned}$$

- (b) Consider the following grammar G_2 with the start symbol E and draw the parse tree corresponding to the expression $id - ic / id$.

$$\begin{aligned}
 E &\rightarrow T E' \\
 E' &\rightarrow - T E' \mid \varepsilon \\
 T &\rightarrow F T' \\
 T' &\rightarrow / F T' \mid \varepsilon \\
 F &\rightarrow (E) \mid id \mid ic.
 \end{aligned}$$

(c) Is $L(G_1) = L(G_2)$?

(d) Consider the following grammar G_3 with the start symbol E . Draw the parse tree corresponding to the expression $id - ic - id$.

$$\begin{aligned} E &\rightarrow T - E \mid T \\ T &\rightarrow T / F \mid F \\ F &\rightarrow (E) \mid id \mid ic. \end{aligned}$$

(e) Is $L(G_1) = L(G_3)$?

(f) Consider the following grammar G_4 with the start symbol E . Draw the parse tree corresponding to the expression $id - ic / id$.

$$\begin{aligned} E &\rightarrow E / T \mid T \\ T &\rightarrow T - F \mid F \\ F &\rightarrow (E) \mid id \mid ic. \end{aligned}$$

(g) Is $L(G_1) = L(G_4)$?