

## Practical 8

**Title: WALex Program to extract HTML tags from .html file.**

**Hint:** In this practical, you will create a Lex program that reads an HTML file and extracts all the HTML tags (elements enclosed within < and >). The extracted tags will be written to an output file.

**Program:**

```
% {
    #include <stdio.h>
% }

%%

\<[^\>]*\> { // Matches anything between '<' and '>' (HTML tags)
    printf("%s\n", yytext);      // Print the matched HTML tag to the console
    fprintf(yyout, "%s\n", yytext); // Write the matched HTML tag to output.txt
}

.|\\n; // Matches any other character or newline (ignored)

%%

int yywrap() {
    return 1;
}

int main() {
    printf("Jay Dalsaniya \n");
    printf("92100103336 \n");
    yyin = fopen("index.html", "r"); // Input file (HTML file)
    yyout = fopen("output.txt", "w"); // Output file (to store HTML tags)

    yylex(); // Start lexical analysis

    fclose(yyin); // Close input file
    fclose(yyout); // Close output file
    return 0;
}
```

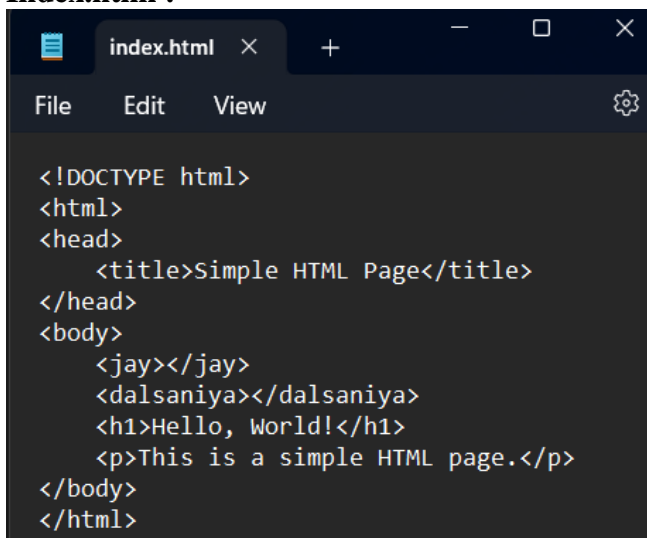
### Output:

```
F:\sem 7\CD\practical8>lex p8.l

F:\sem 7\CD\practical8>gcc lex.yy.c

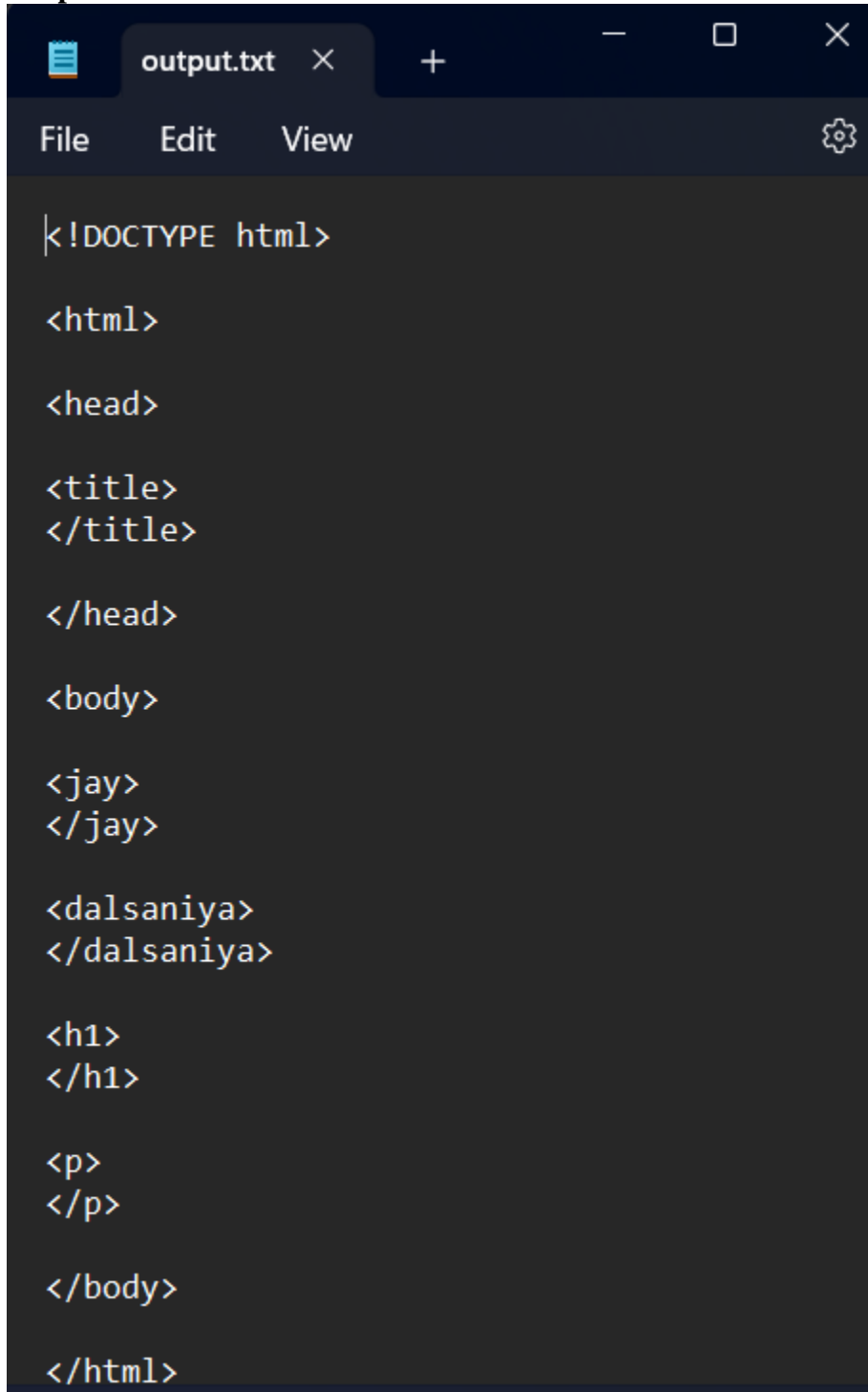
F:\sem 7\CD\practical8>a.exe
Jay Dalsaniya
92100103336
<!DOCTYPE html>
<html>
<head>
<title>
</title>
</head>
<body>
<jay>
</jay>
<dalsaniya>
</dalsaniya>
<h1>
</h1>
<p>
</p>
</body>
</html>
```

### Index.html :



```
index.html x + - □ ×
File Edit View ⚙
<!DOCTYPE html>
<html>
<head>
  <title>Simple HTML Page</title>
</head>
<body>
  <jay></jay>
  <dalsaniya></dalsaniya>
  <h1>Hello, World!</h1>
  <p>This is a simple HTML page.</p>
</body>
</html>
```

Output.txt :



```
<!DOCTYPE html>

<html>

<head>

<title>
</title>

</head>

<body>

<jay>
</jay>

<dalsaniya>
</dalsaniya>

<h1>
</h1>

<p>
</p>

</body>

</html>
```