Form 6 ICT SBA – Case Study 4

Text File Analysis

Form 6 _____ Name _____ No. _____ Date_____

In this project, you are required to write a Pascal program to achieve the following tasks:

1. Read a text file.
2. Display the content of a text file.
3. Count the number of characters excluding punctuation marks.
4. Count the number of words.
5. Count the number of paragraphs.
6. Display the frequencies of letters.
7. Find the frequency of a given word/expression.
8. Justify the paragraphs in the file.

In the text file, words are separated by space and paragraphs are separated by an empty line. Only comma and full-stop will be used as punctuation marks. No number will be used in the text file. The text file will be prepared by using an editor program like NOTEPAD. The content of a sample text file is given below.

```
Try to identify the program involved.  One way to do this is
to see if the error occurs when the program is not running.
If you suspect a memory resident program or device driver,
try by passing it when your computer starts.

Try changing the order in which you load device drivers and
Memory resident programs. This might help because some
errors occur only under specific memory conditions.

If you are using compressed floppy disks with automounting enabled,
you might encounter error messages or other problems while using File
Manager.
```

The project should be done with modular design so that every function will be displayed in a **menu** screen. Users can pick any function from the menu. All outputs of the project are basically assumed to be sent out to the screen.

[Sample Output]

```
Enter the filename of the text file you want to read: File1.txt
```

```
Text File Analyst

**********************************
(1) Display text file
(2) Find the number of characters
(3) Find the number of words
(4) Find the number of paragraphs
(5) Display frequencies of letters
(6) Find the frequency of a given word/expression
(7) Justify the paragraphs
(8) Read another text file
(9) Quit
**********************************

Enter your choice: 1
```

```
Try to identify the program involved.  One way to do this is
to see if the error occurs when the program is not running.
If you suspect a memory resident program or device driver,
try by passing it when your computer starts.

Try changing the order in which you load device drivers and
Memory resident programs. This might help because some
errors occur only under specific memory conditions.

If you are using compressed floppy disks with automounting enabled,
you might encounter error messages or other problems while using File
Manager.

<<< Press <Enter> to return. >>>
```

```
Enter your choice: 2

There are 440 characters.

<<< Press <Enter> to continue. >>>
```

```
Enter your choice: 3

There are 90 words.

<<< Press <Enter> to continue. >>>
```

```
Enter your choice: 4

There are 3 paragraphs.

<<< Press <Enter> to continue. >>>
```

```
Enter your choice: 5
```

```
Letter    Frequency
------------------
  A           18
  B            4
  C           16
  D           17
  E           50
  F            7
  G           15
  H           17
  I           35
  J            0
  K            1
  L            9
  M           19
  N           28
  O           40
  P           13
  Q            0
  R           46
  S           31
  T           29
  U           17
  V            6
  W            6
  X            0
  Y           16
  Z            0
------------------
Total         440

<<< Press <Enter> to continue. >>>
```

```
Enter your choice: 6
```

```
Enter the word/expression you want to count:
if

Frequency of 'if' is 5

<<< Press <Enter> to continue. >>>
```

```
Enter your choice: 7
```

```
Enter the paragraph width (30-70): 60

Try  to identify the program involved.  One way to do this
is  to   see   if the error occurs when the program is not
running.  If  you   suspect  a  memory  resident program or
device driver,  try by passing it when your computer starts.

Try  changing  the order in which you load   device drivers
and Memory  resident  programs. This might help because some
errors occur only under specific memory conditions.

If   you   are   using  compressed  floppy   disks  with
automounting enabled,  you  might  encounter error messages
or  other  problems  while  using   File Manager.

<<< Press <Enter> to return. >>>
```

Data Structures

The following data structure will be used in the program:

- A 1-dimensional string array:

  ```
  line : array[1..99] of string;
  ```

  will be used to store the lines of string read from the text file.

  Here, we assume that the maximum number of lines in the text file is 99.

- A 1-dimensional integer array:

  ```
  frequency : array['A'..'Z'] of integer;
  ```

  will be used in the procedure **`Frequency_of_Letters`** to store the number of occurrences of each letter in the passage.

Procedures

The program consists of the following main procedures

- **`procedure Read_File`**
  - Read the passage from the text file (e.g. File1.txt) and store the lines of texts into the string array `line`.
- **`procedure Display_file`**
  - Display the passage read from the text file onto the screen.
- **`procedure Number_of_Characters`**
  - Count and display the number or characters (letters) in the passage.
- **`procedure Number_of_Words`**
  - Count and display the number of words in the passage.
- **`procedure Number_of_Paragraphs`**
  - Count and display the number of paragraphs in the passage.
- **`procedure Frequency_of_Letters`**
  - Count and display the number of occurrences of each letter in the passage.
- **`procedure Frequency_of_expresion`**
  - Count and display the number of occurrences of an expression entered by the user.
- **`procedure Justify_Paragraphs`**
  - Convert the passage into a right-justified one according to the paragraph width entered by the user.
  - This procedure is a little bit more difficult. You may consider it as a bonus part**.

Hints:

- For the **`procedure Number_of_Words`**, the idea is to take the characters in each line one by one and, in each case, determine whether a new word is found or not. A boolean variable `new_word` is used as a flag to indicate whether the current character is from a new word or not. You may refer to the following algorithm.

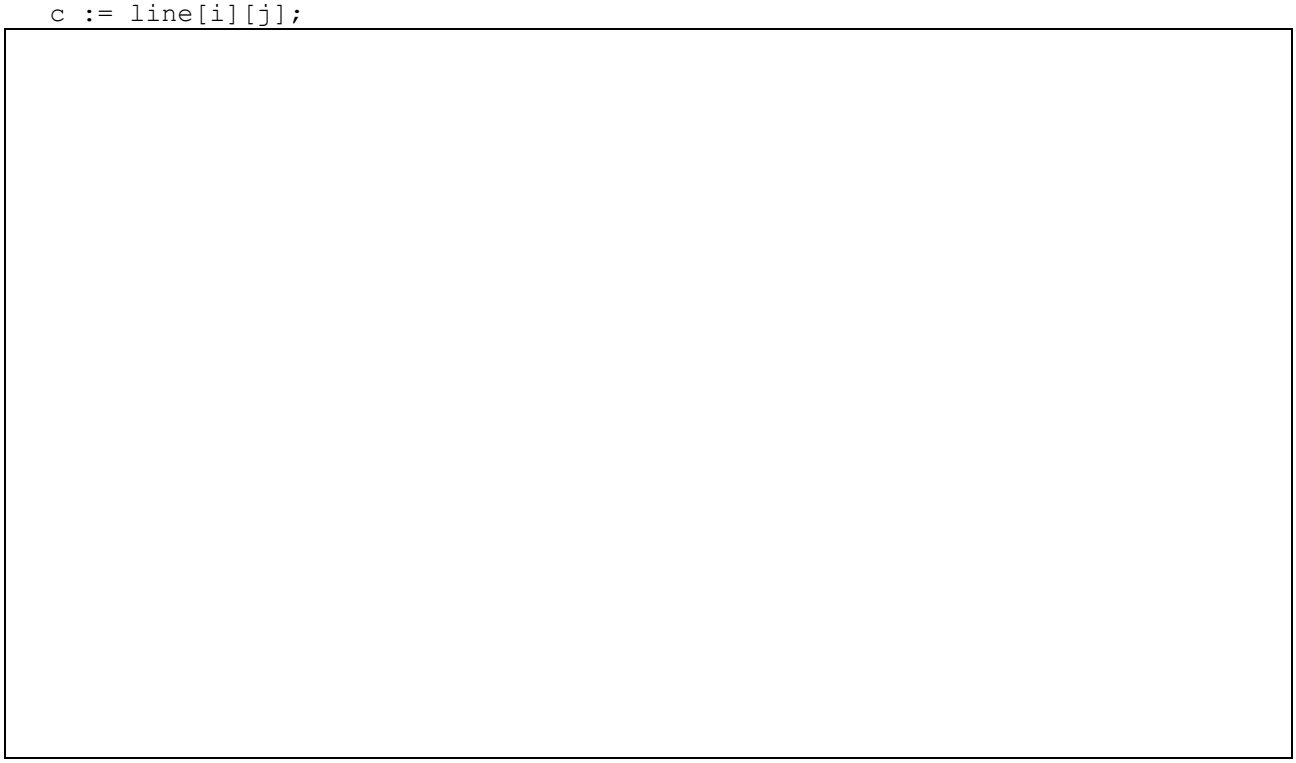
```
number ← 0
For i from 1 to number_of_lines,
   new_word ← true
   For j from 1 to length of line i,
      c ← jth character from line i
      If the jth character lies between 'A' and 'Z' or 'a' and 'z' then
         If new_word
             number ← number + 1
             new_word ← false
      Else
         new_word ← true
```

- For the **`procedure Number_of_Paragraphs`**, you may use similar idea as the **`procedure Number_of_Words`**.

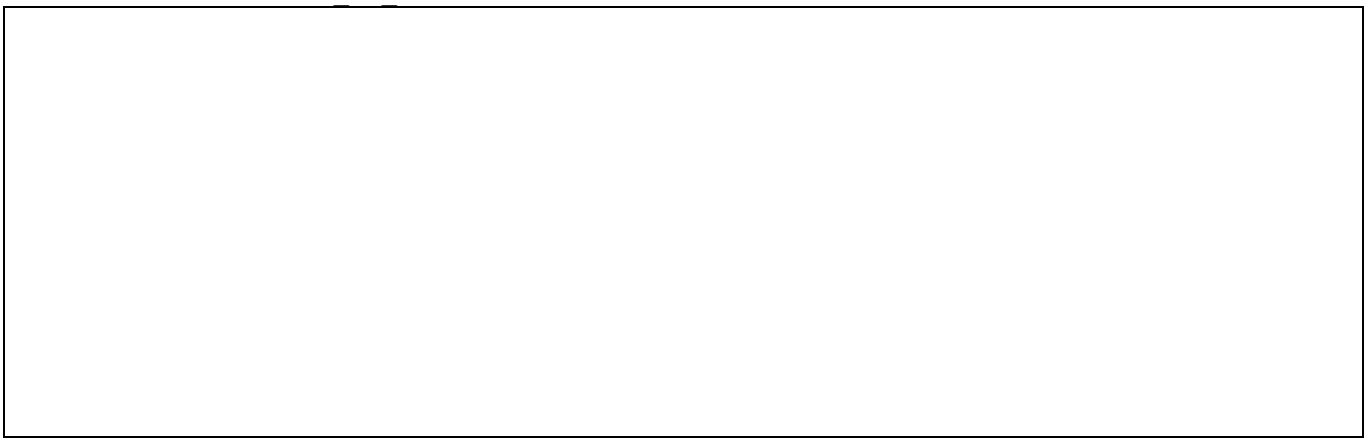
Complete the program below by referring to the results of running the executable program (Text_File_Analyst.exe).

```
program Text_File_Analyst;
uses Crt;
var
  line : array[1..99] of string;
  i, j, number_of_lines, choice, number : integer;


procedure Read_File;
var
  filename : string;
  text_file : text;
begin
  clrscr;
  write('Enter the filename of the text file you want to read: ');
  readln(filename);
  assign(text_file, filename);
  ______________________________;
  number_of_lines := 0;
  while not eof(text_file) do
    begin
      number_of_lines := ________________________________________;
      readln(_________________, line[number_of_lines])
    end;
  close(text_file)
end;


procedure Display_File;
begin
  clrscr;
  for i := 1 to number_of_lines do
    writeln(_________________________________);
  writeln;
  write('<<< Press <Enter> to return. >>>');
  readln
end;


procedure Number_of_Characters;
var
  c : char;
begin
  number := 0;
  for i := 1 to number_of_lines do
    for j := 1 to _________________________________ do
      begin
        c := line[i][j];
        if ((c >= 'a') and (c <= 'z')) __________ ((c >= 'A') and (c <= 'Z')) then
          number := number + 1;
      end;
  writeln('There are ', number, ' characters.');
  writeln;
  write('<<< Press <Enter> to continue. >>>');
  readln
end;
```

```
procedure Number_of_Words;
var
  new_word : boolean;
  c : char;
begin
  number := 0;
  for i := 1 to number_of_lines do
    begin
      new_word := true;
      for j := 1 to length(line[i]) do
        begin
          c := line[i][j];
```



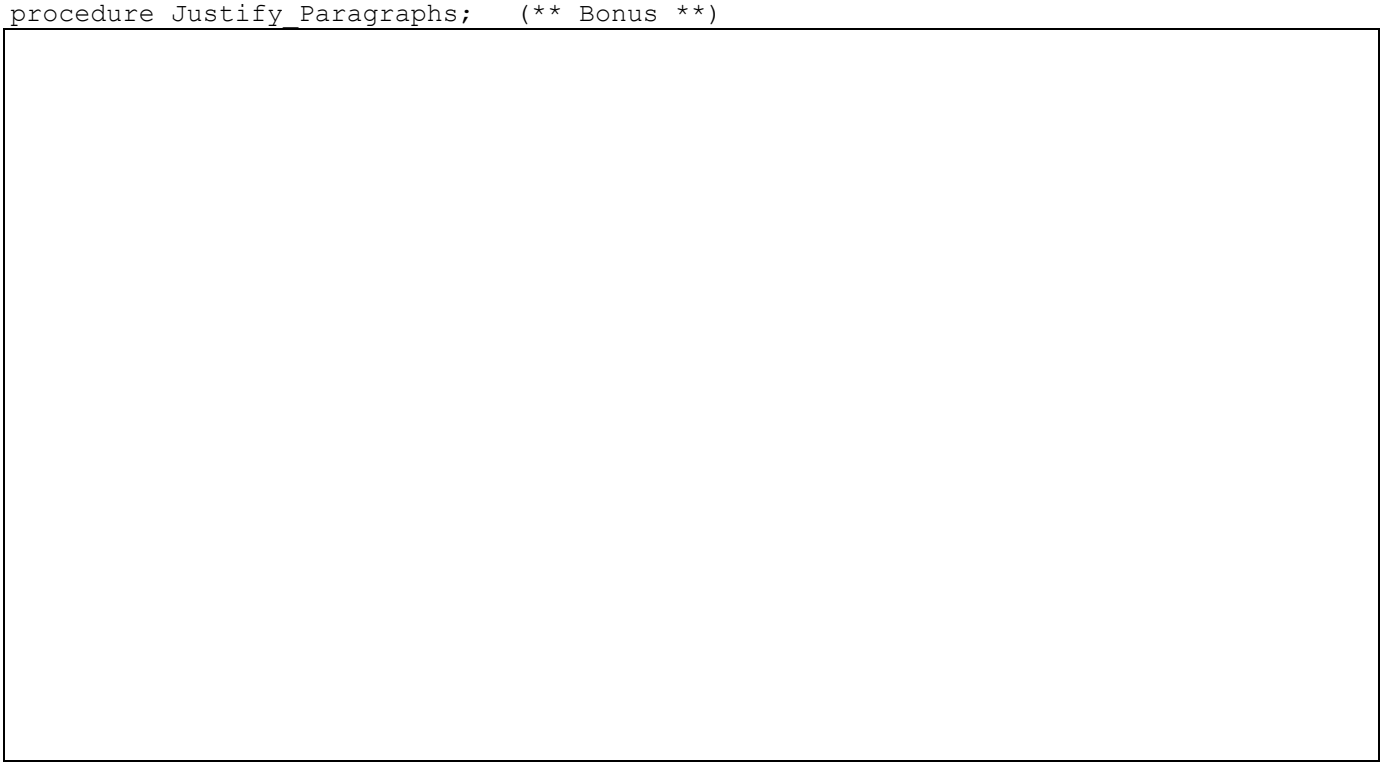
```
        end
    end;
  writeln('There are ', number, ' words.');
  writeln;
  write('<<< Press <Enter> to continue. >>>');
  readln
end;

procedure Number_of_Paragraphs;
var
  new_paragraph : boolean;
begin
  number := 0;
  new_paragraph := true;
  for i := 1 to number_of_lines do
```



```
  writeln('There are ', number, ' paragraphs.');
  writeln;
  write('<<< Press <Enter> to continue. >>>');
  readln
end;
```

```
procedure Frequency_of_Letters;
var
  frequency : array['A'..'Z'] of integer;
  c : char;
  total : integer;
begin
  clrscr;
  total := 0;
  for c := 'A' to 'Z' do
    frequency[c] := 0;
  for i := 1 to _____________________________ do
    for j := 1 to length(line[i]) do
      begin
        c := upcase(line[i][j]);
        frequency[c] := ____________________________________ ;
        if c in ['A'..'Z'] then
          total := total + 1;
      end;
  writeln('Letter   Frequency');
  writeln('--------------------');
  for c := 'A' to 'Z' do
    writeln(c:3, frequency[c]:15);
  writeln('--------------------');
  writeln('_____________________________________');
  writeln;
  write('<<< Press <Enter> to continue. >>>');
  readln
end;


procedure Frequency_of_expresion;
var
  expression, up_expression : string;
  total : integer;
begin
  clrscr;
  writeln('Enter the word/expression you want to count: ');
  readln(expression);
  up_expression := upcase(expression);
  total := 0;
  for i := 1 to ___________________________________ do
    for j := 1 to length(line[i])-length(up_expression) + 1 do
      begin
         if up_expression = upcase(_____________________________________________________) then
           total := total + 1;
      end;
  writeln;
  writeln('Frequency of ''', expression, ''' is ', total);
  writeln;
  write('<<< Press <Enter> to continue. >>>');
  readln
end;
```

```
procedure Justify_Paragraphs;   (** Bonus **)
```



```
begin  { Main program }
  Read_File;
  repeat
    clrscr;
    writeln('Text File Analyst');
    writeln;
    writeln('**********************************');
    writeln('(1) Display text file');
    writeln('(2) Find the number of characters');
    writeln('(3) Find the number of words');
    writeln('(4) Find the number of paragraphs');
    writeln('(5) Display frequencies of letters');
    writeln('(6) Find the frequency of a given word/expression');
    writeln('(7) Justify the paragraphs');
    writeln('(8) Read another text file');
    writeln('(9) Quit');
    writeln('**********************************');
    writeln;
    write('Enter your choice: ');
    readln(choice);
    writeln;
    case choice of
      1 : _____________________________;
      2 : _____________________________;
      3 : _____________________________;
      4 : _____________________________;
      5 : _____________________________;
      6 : _____________________________;
      7 : _____________________________;
      8 : _____________________________;
    end;
  until _____________________________
end.
```