tcg crest
Inventing Harmonious Future

Institute for Advancing Intelligence, TCG CREST
(TCG Centres for Research and Education in Science and Technology)

Introduction to Computer Programming and Data Structures
Assignments 04

Maximum Marks: 150
Clarification Deadline: 2023-Mar-07 Submission Deadline: **2023-Mar-11**

Assignment problem # AP0401

- Problem: Frequency analysis: Suppose you are given a dictionary of English words. Let $\mathbb{A} = \{a, b, \ldots, z\}$ be the set of lowercase letters.

  1. Compute the non-case-sensitive frequency frequencies of each letter.
     Let $fr_a, fr_b, \ldots, fr_z$ be such frequencies. Then, for each $a \in \mathbb{A}$, non-case-sensitive frequency fr_a counts occurrences of both a and A.
  2. Compute the non-case-sensitive frequencies g_{ab} for each pair of letters $ab \in \mathbb{A} \times \mathbb{A}$.

- Input: path of the dictionary (source file download link)

- Output: List of letter/pair of letters and its frequency, separated by a space. Example

  - $a\ fr_a$
  - $b\ fr_b$
  - $\vdots$
  - $z\ fr_z$
  - $aa\ fr_{aa}$
  - $ab\ fr_{ab}$
  - $\vdots$
  - $zz\ fr_{zz}$

[20]

Assignment problem # AP0402

- Find and replace: given a text file and display the content of the file in the terminal (upto 1000 lines). Then ask for user input for two words `word_find` and `word_replace`.

  1. Find the word textttword_find throughout the text file and replace with `word_replace`. Save the file with a new name.

2. Display the changed words in bold in the terminal. Finally show the number of changes in a new line. Your code should be case sensitive.

- Input: path of the dictionary
- Output: New file content upto 1000 lines.

[40]

Assignment problem # AP0403

- Write a program in C to print frequencies, in ascending order, of all unique elements in a given 2D float array.
- Input matrix must be taken from the file titled "input_matrix.txt".
- The 2D input matrix is kept in the input file as follows
  $n\ m$
  $a_{11}\ a_{12}\ \ldots\ a_{1m}$
  $a_{21}\ a_{22}\ \ldots\ a_{1m}$
  $\vdots$
  $a_{n1}\ a_{n2}\ \ldots,\ a_{nm}$

[20]

Assignment problem # AP0404

- Problem: *Linked-list and string*: Suppose you are given a list of strings.
- Create a singly link list. Each node in the list should contain a character pointer to a word and a link to the next node.
- Take the strings from the file "list_of_strings.txt" one by one, create a node, add the string in that node, and append the node in front of the link list.
- Write a function to print the strings stored in the nodes of the linked list.
- Rearrange the nodes so that strings are sorted in ascending order in the link list. Use `strcmp`() from `string.h` to compare strings.
- Write a function in C to check whether a given string is a sub-sequence of at least one of the string from the dictionary.
- Write a function that, given a string of at least length 5, deletes all the nodes containing the given string as substring.
- Print the values in the linked list after every operation.
- Write user menu/choice accordingly.

[70]