**HOTEL REVIEWS SENTIMENT ANALYSIS WITH ARRAYS AND LINKED LISTS**

*(Working Duration: 2 September 2024 – 4 October 2024 – 30 Marks)*

**Background:**

TripAdvisor hotel reviews are essential tools for modern travellers, shaping their choices and improving their travel experiences. Serving as a premier platform for user-generated content, TripAdvisor aggregates a vast array of reviews contributed by guests worldwide. These reviews provide invaluable insight into various aspects of accommodations, which include service quality, amenities, cleanliness, and overall guest satisfaction. To make well-informed decisions that are aligned with their expectations and preferences, travellers rely on this wealth of firsthand information.

The significance of TripAdvisor reviews lies in their authenticity and relevance. Unlike traditional marketing materials, these reviews are candid and based on personal experiences, providing a realistic portrayal of what guests can expect. This transparency helps travellers navigate through a multitude of options and choose accommodations that best suit their needs. By reading reviews, travellers gain confidence in their choices, knowing they are supported by the collective wisdom of fellow travellers. This empowerment not only aids in selecting suitable lodging but also enhances the overall travel planning process.

Ultimately, TripAdvisor reviews serve as more than just a resource; they become trusted companions in the journey towards memorable travel experiences. They enable travellers to anticipate and prepare for their stays, ensuring they can fully enjoy their trips possible.

**Question:**

As programmers at TripAdvisor, your team's task is to develop a streamlined sentiment analysis program focused specifically on TripAdvisor hotel reviews using C++. The primary goal of this program is to ***accurately tally the number of feedback comments associated with a specified rating column within a given CSV file***. To achieve this objective, the program will utilize predefined lists of positive and negative words stored in separate text files: negative-words.txt and positive-words.txt.

In addition to analysing the sentiment of the feedback comments, the project entails the creation of two distinct programs: one employing arrays and the other utilizing linked lists. These programs are designed to efficiently manage and organize data extracted from three essential sources: ***tripadvisor\_hotel\_reviews.csv***, ***negative-words.txt***, and ***positive-words.txt***.

**Key functionalities anticipated from both programs include but not limited to:**

1. Implementing data structures (arrays or linked lists) to store and organize hotel feedback, ratings, positive word lists, and negative word lists sourced from text files or CSV files.
2. Ensuring efficient adding, retrieval, modification, and access of data elements required for sentiment analysis using your chosen data structures.
3. Implement functions to count the frequency of positive and negative words in each review using both arrays and linked lists. Display the number of positive and negative words found in each review and summarize the overall sentiment scores (positive / 4, 5, negative / 1 - 2, neutral / 3).

**Sample of output:**

This hotel is amazing, great service! The room was clean, but the staff was unfriendly. The location is perfect, highly recommended. The food was terrible, and the noise was unbearable.

***Notes:***

***Here is an example equation for calculating a sentiment analysis score in this system:***

***Raw Sentiment Score =*** *Positive Word Count – Negative Word Count*

***Normalized Score =*** *(Raw Sentiment Score − Min Raw Score​) / (Max Raw Score −Min Raw Score)*

*Where:*

*N is based on total of words count (positive & negative)*

*Min Raw Score = −N*

*Max Raw Score = +N*

***Sentiment Score (1 – 5) =*** *1 + ( 4 × Normalized Score )*

***Raw Sentiment Score =*** *4 – 3 = 1*

***Normalized Score =*** *((1) – (-7)) / ((7) – (-7)​) = 8 / 14 = 0.57*

*Where:*

*Min Raw Score = −7*

*Max Raw Score = +7*

***Sentiment Score (1 – 5) =*** *1 + ( 4 × 0.57 ) = 3.28*

1. Compare the provided rating in CSV file with the sentiment analysis result to evaluate the accuracy of the rating assigned to that specific feedback.

**Sample Idea 1:**

***Sentiment Score (1 – 5) = 3***

***Rating given by user = 4***

***Analysis output:***

User's subjective evaluation does not match the sentiment score provided by the analysis. There is an inconsistency between the sentiment score generated by the analysis and the user's evaluation of the sentiment.

**Sample Idea 2:**

***Sentiment Score (1 – 5) = 3***

***Rating given by user = 3***

***Analysis output:***

User's subjective evaluation matches the sentiment score provided by the analysis.

There is a consistency between the sentiment score generated by the analysis and the user's evaluation of the sentiment.

1. Calculate and display the overall sentiment of the reviews based on the total counts of positive and negative words.

**Sample Idea 1:**

Total Reviews = 300

Total Counts of positive words = 400

Total Counts of negative words = 300

Frequency of each word used in overall reviews, listed in ascending order based on frequency:

recommended = 5 times

amazing = 10 times

great = 15 times

unfriendly = 16 times

unbearable = 17 times

perfect = 22 times

terrible = 22 times

***…***

Maximum used word in the reviews : terrible, perfect …

Minimum used word in the reviews : recommended

1. Generate detailed reports or summaries that highlight trends identified through this analysis.
2. Your team might also incorporate new and essential features that leverage specific data structures in this system.
3. Ensure robust error handling and data validation to manage unexpected formats or missing data.

**Lab Work #1 – Program & Video Guidelines (15 Marks)**

1. A team can only contain a maximum of **FOUR (4)** members.
2. Your team is required to utilize C++ programming to develop this prototype, comprising two programs in this section.
3. **Built-in containers such as <list>, <vector>, etc. are not allowed** in this assignment. All containers are self-created.

*Refer to the link:* [*https://www.geeksforgeeks.org/containers-cpp-stl/*](https://www.geeksforgeeks.org/containers-cpp-stl/) *for further information on built-in containers in STL C++.*

1. The program will be enhanced if your team effectively implements and compares various useful searching or sorting algorithms in your developed prototype.
2. The evaluation criteria for this lab work #1 include assessing the clarity and structural design of the code, as well as the quality of comments and adherence to good programming practices. (e.g., indentation, meaningful identifier names, comments, etc.).
3. **This task requires an individual submission.**

Each team member is required to upload a personal video recording to the Moodle system by **4 October 2024**, Friday of Week 8, no later than 5:00 pm.

Using up to maximum **5-minutes** of video, demonstrating and elucidating specific aspects of your code development. Relate your explanation to the workload matrix distribution table provided in the Word documentation.

**Notes:**

* *Please compress all videos to under 100mb before submitting.*
* *If the length of the video exceeds the specified time, it will only be assessed up to that specified time.*
* *Videos must be produced at 1x speed and cannot be adjusted (i.e. accelerated) to meet demo video duration requirements.*

The video recording file must adhere to the following name format:

*“<GroupNo>\_<student ID >.mp4”*

For example, **“G1\_TP012345.mp4”**

1. Refer to the **Page 6** for marking criteria of this Lab Evaluation Work #1 submission.

**Solution Work – Documentation Guidelines (15 Marks)**

1. Create a **Word document** (refer Appendix 1) that explains the overall efficiency and limitations of your developed system
2. Clearly demonstrate the differences in implementation and performance between arrays and linked lists based on the ***operations performed*** in this specific sentiment analysis scenario.
3. Present results that clearly show the advantages and disadvantages of using arrays and linked lists in this specific sentiment analysis scenario.
4. Include a summary of all your hypotheses, observations, experimental results, justifications, and reflections, etc. in your word documentation.
5. Your team must include a workload distribution matrix table after the cover page, showing each member's contributions to the project.

*Note that this table will impact each member's personal final mark in Lab Evaluation Work #1 based on their stated contribution percentage.*

1. If you use some code which has been taken or adapted from another source (book, magazine, internet, forum, etc.) then this must be **cited and referenced** using **APA Referencing Style within your source code**, and this must be mentioned explicitly in the **report**. Failure to reference code properly will be treated as plagiarism.
2. **This task requires a group submission.**

Team leader is required to upload a word documentation and a system solution ZIP file to the Moodle system by **4 October 2024**, Friday of Week 8, no later than 5:00 pm.

The documentation should be named using format:

*“<GroupNo>\_<student ID-leader>\_<student ID-member1>\_<student ID-member2>.docx”.*

For example, **“G1\_TP012345\_TP012344\_TP012123.docx”**

*Refer to the sample submission below.*

A screenshot of a computer

Description automatically generated

The zip file must adhere to the following name format:

*“<GroupNo>\_<student ID-leader>\_<student ID-member1>\_<student ID-member2>.zip”*

For example, **“G1\_TP012345\_TP012344\_TP012123.zip”**

1. Refer to **Page 6** for marking criteria of this Solution Work submission.

**Summary: What Do You Need to Hand in During this Assignment Submission?**

1. This assignment requires **TWO (2)** submissions by your team, which include the following:
2. **Lab Work #1 – Individual Submission (15 Marks)**

* A maximum of 5-minutesvideo recording

1. **Solution Work – Group Submission (15 Marks)**

* C++ solution in zip folder, inclusive the *.cpp, .hpp and csv/text files.*
* Word documentation

1. Your team will need to submit all your C++ solution, video recordings and word documentation to the Moodle system by / on **4 October 2024,** Friday of Week 8, before / on 5.00pm.

MARKING CRITERIA

(Lab Evaluation Work #1 - 15 MARKS)

This Lab Evaluation Work #1 will be evaluated according to the following performance criteria:

|  |  |  |
| --- | --- | --- |
| **Assessment Components** | **Inclusive** | **15 Marks** |
| *CLO3: Lab Evaluation Work #1 – 5 Minutes Video Recording* ***(Individual)*** | | |
| Practical Skills: Use of Data Structures & Algorithms + Personal Understanding | | |
| * Utilization of data structures | Technical Proficiency |  |
| * Implementation of relevant algorithms | Technical Proficiency |  |
| * Demonstrates understanding of data structures/algorithms used | Comprehensive Understanding |  |
| * Justifies choices of structures/algorithms | Insightful Justification |  |

MARKING CRITERIA

(Solution Work - 15 MARKS)

This solution work will be evaluated according to the following performance criteria:

|  |  |  |
| --- | --- | --- |
| **Assessment Components** | **Inclusive** | **15 Marks** |
| *CLO2: Solution Work - Documentation* ***(Group)*** | | |
| Theoretical Explanation  (e.g., Data Structures, Algorithms) | Thorough and Comprehensive |  |
| Code / Input Output Screenshots | Detailed and Comprehensive |  |
| Development Observations / System Effectiveness  (Time and Space Complexity) | Analytical and Insightful |  |
| Summary Discussions  (Strengths and Weaknesses of Solution, Data Structures, and Algorithms) | Insightful and Evaluative |  |
| Other Relevant / Importance Discussions | Clearly Highlighted and Emphasized |  |
| Content Organization | Well-Structured and Logical |  |

Approximation of Total Pages for the solution work documentation: **40 (max).**

Approximation of Words for the documentation: **3000 words (min)**

**APPENDIX 1:**

**SOLUTION WORK - DOCUMENTATION**

**(15 MARKS)**

The report outline as below:

* **Cover Page**
* Module Code and Name
* Intake Code
* Group Number
* Member List.
* **Workload Matrix Table with signature**
* **Design and Implementation**
* Describe the system's functions.
* Elaborate on the data manipulations in chosen data structures.
* Source code including data structure, class, and algorithm that was developed.
* System input output screenshots
* **Result & Discussions**
* System efficiency (*including actual execution time, time efficiency, space efficiency etc.*).
* Summarize and discuss all the observations / thoughts made during the development.
* Critically summarize the strengths and weaknesses of the code you developed.
* **Conclusion, Future Works and Reflection**
* Summary of your works
* Describe your future work in terms of your system weaknesses, if any.
* Additional personal thoughts on this assignment, if any.
* **References (if applicable)**
* **Appendix (if applicable)**

Approximation of Total Pages : **40 (max).**

Approximation of Words : **3000 words (min)**