Due Date: 28.03.2024, 23:59

CENG 112 - Data Structures

Homework 1: Word Frequency Counter

This homework will cover the topics given below:

- Bag ADT
- File I/O
- Generics
- Arrays

Please read the whole document carefully.

In this homework, you are expected to implement a "Word Frequency Counter" in Java. In this application, there is a text and a dictionary. The text contains words, and the dictionary contains the words that the text contains. Also, there are three types of words: rare, frequent, and most frequent. For each word type, there is a relevant dictionary: rare words dictionary, frequent words dictionary, and most frequent words dictionary.

This application separates words from the text and transfers them to the relevant dictionaries such as the rare words dictionary or the frequent words dictionary. The text, dictionary, rare words dictionary, frequent words dictionary, and most frequent words dictionary are all bag data structures. In this homework, you are expected to implement a **secure** and **resizable array-based** bag data structure.

The text can be parsed from "text.txt". Each word is separated by a space. After parsing, each word is added to both text and dictionary bags. Before addition, the words that are concatenated with punctuation such as commas, semicolons, etc. are modified so that the punctuations are removed from these words. For example, the word "computer," is modified as "computer" before adding to the text and dictionary.

A word is a rare word if its frequency in the text is smaller than 5. It is a frequent word if its frequency in the text is between 5 and 8. Finally, a word is a most-frequent word if its frequency in the text is greater than 8.

Your program **shouldn't** take user inputs. The **output** of your program will be:

Text: number of words and content.
Rare Words Dictionary: number of words and content.
….
Most Frequent Words Dictionary: number of words and content.

In this homework, your code **must have** the interface, class and their fields, and method implementations given below. The bullets (I), (C), (F), and (M) stand for interface, class, fields, and method respectively. **Make sure that you not only implement the given interfaces, classes, and methods but also use them (especially methods). For each necessary operation of the homework, there is at least one method you can use. Also, remember to control data structures for their availability such as if a bag is empty or not.**

I. IBag

M. `public boolean add(T newEntry);`

```
    M. public boolean isEmpty();
    M. public boolean isFull();
    M. public T removeByIndex(int index);
    M. public boolean remove(T anEntry);
    M. public T remove();
    M. public int getFrequencyOf(T anEntry);
    M. public int getIndexOf(T anEntry);
    M. public boolean contains(T anEntry);
    M. public void clear();
    M. public void displayItems();
    M. public int getCurrentSize();
    M. public T[] toArray();
C. Bag<T> implements IBag<T>
C. Word
    F. private String content;
    F. private int length;
    M. public String toString();
    M. public boolean equals(Object obj);
C. Dictionary<Word> extends Bag<Word>
C. RareWordsDictionary<Word> extends Bag<Word>
C. FrequentWordsDictionary<Word> extends Bag<Word>
C. MostFrequentWordsDictionary<Word> extends Bag<Word>
C. Text<Word> extends Bag<Word>
    M. public void separate(Dictionary<Word> dictionary,
RareWordsDictionary<Word> rareWordsDictionary,
FrequentWordsDictionary<Word> frequentWordsDictionary,
MostFrequentWordsDictionary<Word> mostFrequentWordsDictionary)
C. FileIO
    M. public static void readFile(Text<Word> text,
Dictionary<Word> dictionary);
C. WordFrequencyCounter
    M. public static void main(String[] args);
```

- This is a 2-person group assignment. However, inter-group collaboration is not allowed!
- All assignments are subject to plagiarism detection and the suspected solutions (derived from or inspired by the solution of other groups) will be graded as zero.
- It is not allowed to use Java Collections Framework.
- Your code should be easy to read and test: **Keep your code clean**. **Avoid duplication and redundancy. Follow Java Naming Conventions**. Use ***relative paths*** instead of absolute ones. 🔗

Submission Rules

All submissions must:

- be performed via **Microsoft Teams** by only one of the group members,
- be exported as an Eclipse Project and saved in ZIP format,
- include all necessary data files (if any TXT, CSV, JSON, etc.) in the right directory,
- follow a specific naming convention such that CENG112_HW2_*groupID*.

Eclipse Project: CENG112_HW1_*G05*
Exported Archive File: CENG112_HW1_*G05*.zip
Submissions that do not comply with the rules above are penalized.