

## PROGRAMMAZIONE II – A. A. 2017/18: Primo Progetto Intermedio

[consegna entro il 26 novembre 2017]

### Parte 1.

Il progetto ha l'obiettivo di applicare i concetti e le tecniche di programmazione orientate agli oggetti prese in considerazione durante il corso. Lo scopo del progetto è lo sviluppo di un componente software di supporto alla gestione di insiemi di dati.

Si richiede di progettare, realizzare e documentare la collezione `DataCounter<E>`. `DataCounter<E>` è una collezione che permette di associare un valore numerico a ogni oggetto di tipo `E`. Intuitivamente la collezione si comporta come un dizionario con la differenza che associa un valore numerico ad ogni dato di tipo `E` presente nella struttura. Le operazioni sono definite descritte di seguito.

```
public interface DataCounter<E> {  
    // incrementa il valore associato all'elemento data di tipo E  
    public void incCount(E data);  
    // restituisce il numero degli elementi presenti nella collezione  
    public int getSize();  
    // restituisce il valore corrente associato al parametro data,  
    // e 0 se data non appartiene alla collezione  
    public int getCount(E data);  
    // restituisce un iteratore (senza remove) per la collezione  
    public Iterator<DataCounter<E>> getIterator();  
}
```

1. Si definisca la specifica completa come interfaccia Java del tipo di dato astratto `DataCounter<E>`, fornendo le motivazioni delle scelte effettuate.
2. Si definisca l'implementazione del tipo di dato astratto `DataCounter<E>`, fornendo due implementazioni: la prima basata sulla classe `Hashtable<K,V>`, la seconda sulla classe `TreeMap<K,V>`. In entrambi i casi si devono comunque descrivere sia la funzione di astrazione sia l'invariante di rappresentazione. Si discutano le caratteristiche delle due implementazioni.

### Parte 2

Per valutare il comportamento dell'implementazione proposta si realizzi, utilizzando `DataCounter<E>`, una applicazione che determina la frequenza delle parole in un documento di testo, partendo da quelle più frequenti e utilizzando l'ordine alfabetico in caso di parole con la medesima frequenza. Ad esempio, il calcolo della frequenza applicato al paragrafo precedente produrrebbe come risultato

Number of characters (including spaces): 295

Number of characters (without spaces): 256

Number of words: 43

## word frequency count

Order	word frequency count	Occurrences
1.	di	2
2.	e	2
3.	in	2
4.	la	2
5.	frequenza	2
6.	parole	2
7.	utilizzando	2
8.	alfabetico	1
9.	applicazione	1
10.	caso	1
11.	che	1
12.	comportamento	1
13.	con	1
14.	datacounter	1
15.	dell	1
16.	delle	1
17.	determina	1
18.	documento	1
19.	frequenti	1
20.	il	1
21.	implementazione	1
22.	l	1
23.	medesima	1
24.	ordine	1
26.	partendo	1
27.	per	1
28.	più	1
29.	proposta	1
30.	quelle	1
31.	realizzi	1
32.	si	1
33.	testo	1
34.	un	1
35.	una	1
36.	valutare	1

Quando si lavora con la frequenza di parole nei documenti, è spesso auspicabile considerare solo le radici delle parole. In questo modo, "dorme", "dormire" e "dormiente" sono tutti considerati la stessa parola. Questo processo si chiama *stemming* e viene utilizzato nei motori di ricerca e in molti altri contesti. Nel caso del progetto, non è richiesto di effettuare effettivamente lo *stemming*, ma suggeriamo di normalizzare i documenti in ingresso nel modo seguente

1. convertire tutte le lettere in lettere minuscole, quindi "La" e "la" sono la stessa parola
2. rimuovere tutta la punteggiatura dal testo

### **Modalità di consegna**

- Il progetto deve essere svolto e discusso col docente individualmente. Il confronto con colleghi mirante a valutare soluzioni alternative durante la fase di progetto è incoraggiato.
- Il progetto deve essere costituito da
  - i file sorgente contenenti il codice sviluppato e le corrispondenti batterie di test, ove tutto il codice deve essere adeguatamente commentato;
  - una relazione di massimo due pagine, che descrive le principali scelte progettuali ed eventuali istruzioni per eseguire il codice.
- La consegna va fatta inviando per email al Prof. Ferrari o al Prof. Gadducci con oggetto “[PR2A] Consegna Progetto Intermedio 1” e “[PR2B] Consegna Progetto Intermedio 1”, rispettivamente

### **Altre informazioni**

- Per quanto riguarda il progetto, i docenti risponderanno solo a eventuali domande riguardanti l’interpretazione del testo, e non commenteranno soluzioni parziali prima della consegna.