

Huffman Εργασία 2020

ΟΜΑΔΑ 1: Αντωνόπουλος Διογένης

Στεφανίδου Άρτεμις

Χύσκαϊ Βασίλης

Μέρος 1ο

Η παράγραφος αυτή αναφέρεται στον τρόπο σκέψης και υλοποίησης του πρώτου μέρους της εργασίας.

- 1) Αρχικά, δημιουργήσαμε έναν πίνακα 3ων θέσεων για να μπορέσουμε να αποθηκεύσουμε τα τρία διαφορετικά url που δόθηκαν από την εκφώνηση.
- 2) Στη συνέχεια δημιουργήσαμε τα αντικείμενα για τα url καθώς και τρεις BufferedReader για να μπορέσουμε να τα διαβάσουμε και να καταγράψουμε το πόσες φορές εμφανίζεται το κάθε γράμμα(από τα πρώτα 128 του ASCII) και στα τρία url μαζί.

```
url[0] = new URL ("https://www.gutenberg.org/files/1342/1342-0.txt");  
url[1] = new URL ("https://www.gutenberg.org/files/11/11-0.txt");  
url[2] = new URL ("https://www.gutenberg.org/files/2701/2701-0.txt");  
  
BufferedReader[] reader = new BufferedReader[3];
```

- 3) Για την διαδικασία αυτή ορίσαμε:

➤ Έναν πίνακα 128 θέσεων ,που η κάθε θέση αντιστοιχεί στον counter εμφάνισης του κάθε γράμματος.

```
int[] chars = new int[128];
```

➤ Μία μεταβλητή που κρατάει την int τιμή του γράμματος που επέστρεψε το read().

```
int charValue = 0;
```

➤ Μία for που θα εκτελεστεί 3 φορές με το σκεπτικό ότι η ίδια διαδικασία θα γίνει και για τα τρία url.

➤ Ένα while και μία read() για να διαβάσουμε χαρακτήρα-χαρακτήρα και τα τρία url.

```
for (int i = 0; i < 3; i++) {  
  
    reader[i] = new BufferedReader(new InputStreamReader(url[i].openStream()));  
  
    while ((charValue = reader[i].read()) != -1) {  
        if (charValue < 128) {  
            chars[charValue]++;  
        }  
    }  
}
```

- Ένα if για να πάρουμε μόνο όσα είναι μεταξύ 0 και 127 και να αυξήσουμε τον αντίστοιχο μετρητή του γράμματος.

4) Στη συνέχεια κλείσαμε τα αρχεία-url.

```
reader[i].close();
```

5) Δημιουργήσαμε έναν BufferedWriter για να τυπώσουμε τις θέσεις του πίνακα με τις αντίστοιχες συχνότητες για κάθε γράμμα-ειδικό χαρακτήρα ,μέσω μιας for ,στο αρχείο “frequencies.dat” όπως μας ζητήθηκε από την εκφώνηση .(μικρή αλλαγή στον κώδικα του 1ου μέρους για να μας διευκολύνει στο 2ο μέρος)

```
//Create a file and print the results there
try (BufferedWriter outputStream = new BufferedWriter(new FileWriter("frequencies.dat"))) {
    for (int i = 0; i < 128; i++) {
        outputStream.write(i + " -> " + chars[i] + "\n");
        outputStream.flush();
    }
}
```

```
//Create a file and print the results there
try (BufferedWriter outputStream = new BufferedWriter(new FileWriter("frequencies.dat"))) {
    for (int i = 0; i < 128; i++) {
        outputStream.write( chars[i] + "\n");
        outputStream.flush();
    }
}
```

Παραδείγματα Εκτέλεσης του Κώδικα μας:

Ακολουθεί παράδειγμα από το file frequencies.dat ,που δημιουργείτε στο φάκελο του project (εκεί που είναι και το rom.xml):

2621

2098

1450

1876

1106

1082

Μέρος 2ο

Η παράγραφος αυτή αναφέρεται στον τρόπο σκέψης και υλοποίησης του δεύτερου μέρους της εργασίας.

- 1) Ορίσαμε μία κλάση για τους κόμβους του δέντρου ,όπως μας ζητήθηκε στην εκφώνηση,με το όνομα Node η οποία περιέχει:

```
public class Node implements Comparable<Node>, Serializable {
```

- Ένα χαρακτήρα (σε περίπτωση που ο κόμβος είναι φύλλο ο, χαρακτήρας είναι ένα γράμμα από τα πρώτα 128 του ASCII, ενώ αν δεν είναι τότε ο χαρακτήρας ισούτε με το '\0').
- Μια συχνότητα (που θα είναι ή η συχνότητα του γράμματος ,αν είναι φύλλο,ή το άθροισμα των συχνοτήτων των δύο παιδιών του node)
- Δύο παιδιά,ένα δεξί και ένα αριστερό (που θα είναι null αν είναι leaf)

Αυτά τα χαρακτηριστικά του Node τα βάλουμε **private** γιατί δε θέλουμε να είναι προσβάσιμα εκτός κλάσης και **final** γιατί δεν αλλάζουν ,παρά μόνο κατά την αρχικοποίηση τους που γίνεται μέσω του **Constructor**.

```
//Node Characteristics
private final char character;
private final int frequency;
private final Node left;
private final Node right;

//Constructor
public Node(char character, int frequency, Node left, Node right) {
    this.character = character;
    this.frequency = frequency;
    this.left = left;
    this.right = right;
}
```

- Την μέθοδο compareTo επειδή κάναμε **implement** το **Interface Comparable<T>** γιατί θέλαμε να έχουμε τη δυνατότητα σύγκρισης μεταξύ των αντικειμένων που θα δημιουργήσουμε.Στην ουσία έπρεπε να υλοποιήσουμε και να κάνουμε override την μέθοδο αυτή η οποία μας γυρνάει τη διαφορά δύο συχνοτήτων δύο διαφορετικών κόμβων.Δηλαδή συγκρίνει τις δύο συχνότητες και γυρνάει <0 αν η συχνότητα του n είναι μεγαλύτερη ,>0 αν είναι μικρότερη και 0 αν είναι ίσες.

```
//Compare the frequencies from two nodes
@Override
public int compareTo(Node n) {
    return this.frequency - n.frequency;
}
```

- Μία μέθοδο που ελέγχει αν ο κόμβος αυτός που της δώσαμε ως παράμετρο είναι φύλλο ή όχι.Για να είναι φύλλο ένας κόμβος πρέπει και τα δύο του παιδιά να είναι null.Τα φύλλα είναι αυτά που περιέχουν ένα χαρακτήρα από τους 128 του ASCII.Προς το παρόν όμως τη μέθοδο αυτή δε τη χρειαστήκαμε κάπου αλλά θεωρήσαμε πως θα μας είναι χρήσιμη στη συνέχεια της εργασίας και ότι είναι καλό να την έχουμε για κάθε καινούριο αντικείμενο Node που θα δημιουργούμε.

```
//Check if node is leaf:Has character
public boolean isLeaf() {
    return right == null && left == null;
}
```

- Τους getters για τα χαρακτηριστικά του Node που χρειαστήκαμε να χρησιμοποιήσουμε τις τιμές τους και σε άλλη κλάση.

```
public char getCharacter() {
    return character;
}

public int getFrequency() {
    return frequency;
}

public Node getLeftChild() {
    return left;
}

public Node getRightChild() {
    return right;
}
```

2) Ορίσαμε επίσης μία κλάση για το δέντρο Huffman όπως μας ζητήθηκε στην οποία υλοποιούμε όλη τη διαδικασία δημιουργίας ενός δέντρου Huffman. Η κλάση περιέχει:

```
public class Huffman implements Serializable {
```

- Μία μέθοδο makeTree για να μπορέσουμε να υλοποιήσουμε τον αλγόριθμο του Huffman που να παράγει ένα δέντρο χρησιμοποιώντας ως είσοδο τον πίνακα συχνοτήτων.

```
public Node makeTree(int[] array) {
```

- **Υλοποίηση μεθόδου :**

- ✓ Δημιουργήσαμε έναν πίνακα από Node με 128 θέσεις (μία για τον κάθε χαρακτήρα) έτσι ώστε να τα αρχικοποιήσουμε εύκολα με μία for loop και να μπορέσουμε να φτιάξουμε ένα δέντρο (ενός κόμβου) για τον κάθε ASCII χαρακτήρα βάζοντας κάθε φορά, πέρα από τον χαρακτήρα, την αντίστοιχη συχνότητα από τον πίνακα που δεχτήκαμε ως παράμετρο, και τα παιδιά του να είναι null.

```
//Create an array of nodes. Each node represents a character
Node[] treeNodes = new Node[128];
for(char i = 0; i < array.length; i++) {
    treeNodes[i] = new Node(i, array[i], null, null);
}
```

- ✓ Χρησιμοποιήσαμε την MinHeap του εργαστηρίου για να ταξινομήσουμε τα 128 Nodes και να έχουμε στην κορυφή του δέντρου το Node με τη μικρότερη συχνότητα.

```
//Create the MinHeap for our Nodes using the right constructor to use heapify
MinHeap<Node> h =new ArrayMinHeap<>(treeNodes);
```

- ✓ Όσο το size της MinHeap είναι μεγαλύτερο από 1(αντίστοιχο στις διαφάνειες του μαθήματος:Μέχρι να μείνει ένα δέντρο)καλούμε την deleteMin η οποία διαγράφει και γυρνάει το πρώτο Node από τη MinHeap άρα και αυτό με τη μικρότερη συχνότητα και το αποθηκεύουμε σε ένα προσωρινό Node .Αφού η deleteMin περιέχει στον κώδικα της την fixDown πάει να πει ότι έγιναν οι απαραίτητοι έλεγχοι και ενέργειες έτσι ώστε πρώτο στοιχείο στην h να ξαναείναι το μικρότερο από όσα έμειναν.Άρα ξανακαλούμε την deleteMin και γίνεται η ίδια διαδικασία που αναφέρθηκε και παραπάνω. Στη συνέχεια τα συγχωνεύσαμε σε ένα νέο δυαδικό δέντρο που έχει αριστερό παιδί ρίζας το 1ο και δεξί παιδί ρίζας το 2ο και δώσαμε συχνότητα ίση με το άθροισμα των συχνοτήτων των παιδιών του και χαρακτήρα τον κενό '\0'. Τέλος,προσθέσαμε το νέο node στην MinHeap για να μπορέσουμε και πάλι να βρούμε τα δύο μικρότερα και να συνεχίσουμε την διαδικασία μέχρι το size να γίνει ένα.

```
//Create the Huffman Tree with the help of MinHeap
while(h.size() > 1) {
    Node leftChild = h.deleteMin();
    Node rightChild = h.deleteMin();
    Node parent = new Node('\0',leftChild.getFrequency() + rightChild.getFrequency(),leftChild,rightChild);
    h.insert(parent);
}
```

- ✓ Στο τέλος,καλούμε την deleteMin για να πάρουμε το root του δέντρου Huffman που δημιουργήσαμε.

Παρατήρηση:Στην ουσία το size μειώνεται κατά ένα κάθε φορά αφού διαγράφουμε 2 και προσθέτουμε ένα node στην MinHeap.Ο πατέρας ,το node δηλαδή που έχει δύο παιδιά και δώσαμε συχνότητα ίση με το άθροισμα των συχνοτήτων των παιδιών του ,να μεν μπαίνει μόνο αυτός με τη βοήθεια του insert στην MinHeap αλλά δεν παύει να δείχνει και στα δύο παιδιά που είχε πριν μπει.

3) Τέλος, χρησιμοποιήσαμε την κλάση App από το πρώτο μέρος της εργασίας:

- Δημιουργήσαμε σε αυτή ένα αντικείμενο Scanner στο οποίο δώσαμε σαν όρισμα το αρχείο frequencies.dat για να μπορέσουμε στη συνέχεια να το διάβασουμε και να πάρουμε την απαραίτητη πληροφορία που χρειαζόμαστε για το δέντρο Huffman.

```
Scanner scanner = new Scanner(new File("frequencies.dat"));
```

- Για να το πετύχουμε αυτό ορίσαμε έναν πίνακα με 128 θέσεις για να αποθηκεύσουμε τις συχνότητες των 128 χαρακτήρων ASCII που θα πάρουμε από το αρχείο.
- Χρησιμοποιήσαμε μία for και το nextInt() για να διαβάσουμε κάθε φορά έναν ακέραιο από το αρχείο ,δηλαδή μία συχνότητα και να την αποθηκεύουμε στην αντίστοιχη θέση του πίνακα array.

```
int[] array = new int[128];
for(int i= 0; i < 128; i++) {
    array[i] = scanner.nextInt();
}
```

- Δημιουργήσαμε ένα αντικείμενο Huffman για να μπορέσουμε στη συνέχεια να χρησιμοποιήσουμε την μέθοδο που έχει αυτή η κλάση.

```
//Call class HuffmanTree to create an object for that by the default constructor
Huffman tree = new Huffman();
```

- Αποθηκεύσαμε το αντικείμενο του δέντρου στο αρχείο tree.dat χρησιμοποιώντας Java Object Serialization(γι αυτό το λόγο έχουμε κάνει και 'implements Serializable' στην κλάση Huffman).Η makeTree φτιάχνει ένα δέντρο Huffman και μας γυρνάει το root του δέντρου αυτού και στην ουσία όλο το δέντρο αφού έχει pointers που δείχνουν σε αυτό.Έτσι, καλώντας την με όρισμα τον πίνακα συχνοτήτων των χαρακτήρων,μας δημιουργεί το Huffman με αυτές τις συχνότητες και μας γυρνάει το root του.Το node δηλαδή που χρειαζόμαστε για να αποθηκεύσουμε ολόκληρο το δέντρο στο tree.dat.

```
//Create the huffman tree by calling makeTree and write it in tree.dat
try (ObjectOutputStream objectOut = new ObjectOutputStream(new FileOutputStream("tree.dat"))) {
    objectOut.writeObject(tree.makeTree(array));
}
```

Δύο λόγια για την MinHeap αφού την έχουμε υλοποιήσει και στο εργαστήριο:

- ✓ Η insert που την χρησιμοποιήσαμε και στην εργασία αυξάνει κατά ένα το size του πίνακα της MinHeap και αποθηκεύει εκεί το Node(στην προκειμένη περίπτωση)που προσθέσαμε. Στη συνέχεια,καλεί την fixup η οποία λέει όσο $k > 1$ δηλαδή όσο το αντικείμενο που έδωσα δεν είναι η ρίζα του δέντρου και όσο το αντικείμενο είναι πιο μεγάλο από τον πατέρα του τότε κάνει swap τον πατέρα με το παιδί του και πήγαινε στην θέση που ήταν πριν ο πατέρας για να ξαναγίνει το while.Με αυτόν τον τρόπο εξασφαλίζουμε ότι διαρκώς το root θα είναι και το μικρότερο αλλά και ότι η MinHeap είναι διατεταγμένη σωστά κάθε φορά που προσθέτουμε ένα αντικείμενο σε αυτήν.
- ✓ Η deleteMin αποθηκεύει σε μια προσωρινή μεταβλητή το πρώτο στοιχείο του πίνακα της MinHeap(άρα και το ελάχιστο)και στη θέση του τοποθετεί το τελευταίο στοιχείο.Στη συνέχεια,διαγράφει την τελευταία θέση του πίνακα και καλεί τη fixdown.Η fixdown με τη σειρά της κάνει τους απαραίτητους ελέγχους για να βρει το μικρότερο από τα δύο παιδιά του Node που το βάλαμε στο root και μετά συγκρίνει το min των παιδιών αυτών με τον πατέρα.Αμα ο πατέρας είναι μεγαλύτερος τότε κάνει swap με το παιδί αλλιώς παραμένει στη θέση που ήταν.Τέλος,η deleteMin γυρνάει την προσωρινή μεταβλητή.