

Data Structures

Πρώτο Παραδοτέο

Γεώργιος-Νεκτάριος Εγγλέζος it21824
Αντώνης Δρόσος it218115
Ανδρέας Σαραντόπουλος it218140

Εισαγωγή

Για την υλοποίηση της εργασίας χρειαστήκαμε 2 κλάσεις και τα τρία txt αρχεία (Βρίσκονται στο root του φακέλου Project) από τα οποία παίρνουμε το κείμενο.

1. Main.java
2. Text.java
3. First_File.txt
4. Second_File.txt
5. Third_File.txt

Main.java

Η **Main** φτιάχνει έναν πίνακα (array) τύπου Integer μεγέθους 128 (0-127), ο οποίος θα χρησιμοποιηθεί για να μετρήσουμε πόσες φορές εμφανίστηκε ο κάθε χαρακτήρας στα τρία αρχεία κειμένου. Τέλος καλεί την μέθοδο **Text.Extract_Text()** που εκτελεί όλες τις απαραίτητες μεθόδους.

Text.java

Περιέχει όλες τις απαραίτητες μεθόδους για διάβασμα από τα τρία αρχεία κειμένου (txt), υπολογισμό συχνότητας χαρακτήρων στον πίνακα που πήραμε από την **Main** και τέλος την δημιουργία του αρχείου **“frequencies.dat”** που περιέχει τα αποτελέσματα του προγράμματος.

Κώδικας

Main.java

```
private static int[] freq= new int[128];
```

Περιέχει τον private πίνακα integer μεγέθους 128 (0-127) όσα είναι δηλαδή και τα νούμερα για το ascii και φτιάχτηκε ώστε να χρησιμοποιηθεί ως μετρητής για κάθε χαρακτήρα.

```
for (int i=0;i<freq.length;i++)  
{  
    freq[i]=0;  
}
```

Αρχικοποιούμε τις τιμές του πίνακα σε 0 για να είναι πιο εύκολο να τις αυξήσουμε στην συνέχεια με ένα απλό `freq[i]++`;

```
Text.extractText(freq);
```

Καλώ την μέθοδο `extractText(int[] freq)` που βρίσκεται στην κλάση `Text.class` και περνάω ως παράμετρο τον πίνακα που έφτιαξα.

Text.java

static String[] *fileNames*= new String[3];

Ένας πίνακας τύπου String με τρεις θέσεις που περιέχει τα ονόματα των τριών αρχείων κειμένου (txt).

- **extractText**(int[] freq)

Η μέθοδος αυτή είναι εκείνη που καλεί και συνδυάζει όλες τις υπόλοιπες ώστε να παραχθεί στο τέλος το αρχείο frequencies.dat.

Περιέχει μια επανάληψη **for** (int i=0;i<3;i++) που τρέχει τον ίδιο κώδικα και για τα τρία αρχεία κειμένου(txt).

Μέσα σε αυτήν την επανάληψη:

Καλώ την μέθοδο **getText**(String fileName)

περνώντας ως παράμετρο το όνομα του αρχείου που τρέχει *fileNames[i]* και μου επιστρέφει σε ένα **BufferedReader** **inputStream** ολόκληρο το κείμενο που περιέχει το txt.

Σε περίπτωση που δεν βρήκε το αρχείο μου επιστρέφει null και ελέγχω τα εξής πράγματα:

```
if (inputStream==null && i<2)
{
    continue;
}
else if (inputStream==null && i==2)
{
    break;
}
else
{
    while ((currentLine = inputStream.readLine())
           != null) {
        char[] arr;
        if (currentLine != "") {
            arr = extractTextFromLine(currentLine);
            makeTheArray(arr, freq);
        }
    }
}
```

Στην περίπτωση που λείπει κάποιο από τα πρώτα δύο αρχεία από το Project βγαίνει από την συγκεκριμένη επανάληψη με την εντολή **continue**; χωρίς να τρέξει κάποιον άλλο κώδικα μέσα σε αυτήν και συνεχίζει με τα επόμενα αρχεία.

Στην περίπτωση που λείπει το τρίτο αρχείο βγαίνει τελείως από την επανάληψη με την εντολή **break**; και εκτελεί τις υπόλοιπες μεθόδους.

Αν το αρχείο που τρέχει στην επανάληψη υπάρχει τότε διαβάζουμε την κάθε γραμμή ξεχωριστά με την επανάληψη **while** `((currentLine = inputStream.readLine()) != null)` Ελέγχουμε αν είναι κενή γραμμή και αν δεν είναι καλούμε την μέθοδο

private static char[] extractTextFromLine(String line) η οποία παίρνει ως παράμετρο την γραμμή (currentLine) και μας επιστρέφει όλους τους χαρακτήρες (char) που βρίσκονται μέσα σε αυτή την γραμμή σε έναν πίνακα τύπου char.

Στην συνέχεια καλούμε την συνάρτηση **private static void makeTheArray(char[] arr, int[] freq)** που παίρνει τους δύο πίνακες και υπολογίζει τις συχνότητες.

Κλείνουμε τον BufferedReader **inputStream.close();** Και τέλος καλούμε την μέθοδο **makeDatFile(int[] freq)** που φτιάχνει το αρχείο frequencies.dat .

- **getText(String fileName)**

Είναι τύπου BufferedReader και παίρνει ως είσοδο το όνομα του αρχείου: **private static BufferedReader getText(String fileName) throws FileNotFoundException**

Έχω μέσα σε μια try-catch ένα **FileNotFoundException** ώστε να ελέγχω αν το αρχείο υπάρχει στον φάκελο.

Μέσα στην try φτιάχνω ένα `bufferReader` ώστε να διαβάζω από το αρχείο και γυρίζω ότι πάρω. Στην περίπτωση που δεν υπάρχει το αρχείο γυρίζω `null`.

```
try {  
    BufferedReader inputStream = new BufferedReader(new  
    FileReader(fileName+".txt"));  
    return inputStream;  
}  
catch (FileNotFoundException ex)  
{  
    System.out.println("file " + fileName + " was not  
    found.");  
    return null;  
}
```

- **extractTextFromLine(String line)**

Παίρνει μια γραμμή και γυρίζει ένα `array` από χαρακτήρες με όλους τους χαρακτήρες από την πρόταση.

Έχουμε έναν πίνακα χαρακτήρων με μέγεθος `line.length()`:

```
char[] arr = new char[line.length()];
```

Τέλος μια επανάληψη που αποθηκεύει όλους τους χαρακτήρες στις θέσεις του πίνακα:

```
for (int i=0;i<line.length();i++)  
{  
    char temp=line.charAt(i);  
    if ( temp!=' ')  
    {  
        arr[i] = line.charAt(i);  
    }  
}
```

- **makeTheArray(char[] arr, int[] freq)**

```
private static void makeTheArray(char[] arr, int[] freq)
{
    for (int i=0;i<arr.length;i++)
    {
        int asciiOfA = (int) arr[i];
        if (asciiOfA<128)
        {
            freq[asciiOfA]++;
        }
    }
}
```

Η μέθοδος αυτή παίρνει σαν παραμέτρους τους δύο πίνακες, τρέχει με μια for loop όλο τον πίνακα arr και ανάλογα με τι τιμή ascii έχει ο χαρακτήρας πηγαίνει στο ανάλογο index του freq και αυξάνει κατά 1. Η μετατροπή από χαρακτήρας σε ascii γίνεται με ένα απλό cast σε int.

- `makeDatFile(int[] freq)`

```
private static void makeDatFile(int[] freq) throws IOException
{
    PrintWriter out = new PrintWriter(new
    BufferedWriter(new FileWriter("frequencies.dat")));
    for (int i=0;i<freq.length;i++)
    {
        out.write(i+": "+freq[i]+",");
    }
    out.close();
    System.out.println("Frequencies.dat is ready in
    the project folder!");
}
```

Χρησιμοποιώ ένα `BufferedWriter` για να δημιουργήσω και να γράψω στο αρχείο **frequencies.dat** . Περνάω όλο τον πίνακα `freq` με μια `for loop` και με την εντολή `out.write` γράφω στο αρχείο όλες τις τιμές την μία δίπλα στην άλλη. Τέλος κλείνω τον `bufferedWriter` με την εντολή `out.close()`.

Ενδεικτικά Αποτελέσματα

Περίπτωση 1:

Όλα τα αρχεία είναι μέσα στον φάκελο ,τρέχω την εντολή “**mvn package**” στο terminal για να κάνω compile το πρόγραμμα , την εντολή “**java -cp target/Project-1.0-SNAPSHOT.jar org.hua.project.Main**” για να τρέξω το jar που βρίσκεται στον φάκελο target και αφού βγάλει το μήνυμα πως φτιάχτηκε το frequencies.dat τρέχω την εντολή “**cat frequencies.dat**” για να δω ότι όντως φτιάχτηκε και περιέχει τα frequencies .

```
geonek@DESKTOP-PK8471L:~/Project$ java -cp target/Project-1.0-SNAPSHOT.jar org.hua.project.Main
Frequencies.dat is ready in the project folder!
geonek@DESKTOP-PK8471L:~/Project$ cat frequencies.dat
0:409147,1:0,2:0,3:0,4:0,5:0,6:0,7:0,8:0,9:0,10:0,11:0,12:0,13:0,14:0,15:0,16:0,17:0,18:0,19:0,20:0,
21:0,22:0,23:0,24:0,25:0,26:0,27:0,28:0,29:0,30:0,31:0,32:0,33:2719,34:22,35:4,36:8,37:3,38:2,39:11,
40:347,41:347,42:217,43:0,44:31243,45:3279,46:15811,47:63,48:239,49:437,50:173,51:151,52:127,53:149,
54:101,55:115,56:127,57:98,58:626,59:5914,60:0,61:0,62:0,63:1674,64:5,65:3609,66:2621,67:2098,68:145
0,69:1876,70:1106,71:1082,72:2231,73:6697,74:600,75:308,76:1597,77:2581,78:1273,79:1000,80:1697,81:4
09,82:1029,83:2744,84:3749,85:294,86:189,87:2149,88:28,89:735,90:23,91:10,92:0,93:10,94:0,95:2144,96
:0,97:128252,98:25722,99:38393,100:65761,101:204231,102:34929,103:33606,104:104115,105:107638,106:17
81,107:12557,108:69078,109:38706,110:112331,111:120723,112:26916,113:2033,114:92778,115:103575,116:1
46680,117:46427,118:15360,119:35949,120:2085,121:32138,122:1633,123:0,124:0,125:0,126:0,127:0,geonek
```

Περίπτωση 2:

Λείπει κάποιο/α από τα αρχεία κειμένου (txt) από τον φάκελο του Project και μας εμφανίζει ποια λείπουν ενώ φτιάχνει το αρχείο με τα υπόλοιπα.

```
geonek@DESKTOP-PK8471L:~/Project$ java -cp target/Project-1.0-SNAPSHOT.jar org.hua.project.Main
file First_File was not found.
file Third_File was not found.
Frequencies.dat is ready in the project folder!
geonek@DESKTOP-PK8471L:~/Project$ cat frequencies.dat
0:27497,1:0,2:0,3:0,4:0,5:0,6:0,7:0,8:0,9:0,10:0,11:0,12:0,13:0,14:0,15:0,16:0,17:0,18:0,19:0,20:0,21:0,22:0,
23:0,24:0,25:0,26:0,27:0,28:0,29:0,30:0,31:0,32:0,33:452,34:22,35:1,36:2,37:1,38:0,39:9,40:73,41:73,42:86,43:
0,44:2576,45:220,46:1227,47:11,48:24,49:66,50:12,51:12,52:8,53:13,54:7,55:8,56:11,57:9,58:249,59:193,60:0,61:
0,62:0,63:204,64:1,65:675,66:120,67:185,68:211,69:195,70:127,71:183,72:245,73:787,74:13,75:82,76:115,77:196,7
8:136,79:148,80:183,81:84,82:165,83:231,84:488,85:71,86:26,87:231,88:10,89:100,90:1,91:5,92:0,93:5,94:0,95:44
0,96:0,97:9193,98:1648,99:2861,100:5293,101:15317,102:2263,103:2775,104:7687,105:7889,106:223,107:1217,108:51
19,109:2269,110:7959,111:9400,112:1801,113:139,114:6520,115:7070,116:11817,117:3938,118:948,119:2749,120:172,
121:2510,122:79,123:0,124:0,125:0,126:0,127:0,geonek@DESKTOP-PK8471L:~/Project$
```

Περίπτωση 3:

Λείπουν όλα τα αρχεία και απλά θα φτιάξει ένα αρχείο με όλες τις θέσεις να έχουν μηδενική τιμή.

```
geonek@DESKTOP-PK8471L:~/Project$ java -cp target/Project-1.0-SNAPSHOT.jar org.hua.project.Main
file First_File was not found.
file Second_File was not found.
file Third_File was not found.
Frequencies.dat is ready in the project folder!
geonek@DESKTOP-PK8471L:~/Project$ cat frequencies.dat
0:0,1:0,2:0,3:0,4:0,5:0,6:0,7:0,8:0,9:0,10:0,11:0,12:0,13:0,14:0,15:0,16:0,17:0,18:0,19:0,20:0,21:0,22:0,23:0,24:0,25:0,26:0,27:0,28:0,29:0,30:0,31:0,32:0,33:0,34:0,35:0,36:0,37:0,38:0,39:0,40:0,41:0,42:0,43:0,44:0,45:0,46:0,47:0,48:0,49:0,50:0,51:0,52:0,53:0,54:0,55:0,56:0,57:0,58:0,59:0,60:0,61:0,62:0,63:0,64:0,65:0,66:0,67:0,68:0,69:0,70:0,71:0,72:0,73:0,74:0,75:0,76:0,77:0,78:0,79:0,80:0,81:0,82:0,83:0,84:0,85:0,86:0,87:0,88:0,89:0,90:0,91:0,92:0,93:0,94:0,95:0,96:0,97:0,98:0,99:0,100:0,101:0,102:0,103:0,104:0,105:0,106:0,107:0,108:0,109:0,110:0,111:0,112:0,113:0,114:0,115:0,116:0,117:0,118:0,119:0,120:0,121:0,122:0,123:0,124:0,125:0,126:0,127:0,geonek@DESKTOP-PK8471L:~/Project$
```