TASK:

Develop a Java prototype program to help you explore differences in compression performances.

I. Develop a program that converts a matrix of decimal integers to a matrix of binary code (codewords) and to a matrix of Gray code.

The input and output of your program should be read from and sent to a text file, namely input.txt and output.txt, respectively. The program should be able to read at least the following data as the input:

- (a) A one dimensional array of integers (decimal values ranging from 0 to 255, i.e. [0,255], inclusive 0 and 255)
- (b) Two dimensional array of integers [0,255]
- (c) Three dimensional array of integers [0,255].

The output of the program should contain the following:

- (a) A one dimensional array of binary codewords and their Gray codes
- (b) Two dimensional array of binary codewords and their Gray codes
- (c) Three dimensional array of binary codewords and their Gray codes.
- II. You should apply ONE of the lossless compression algorithms below to the sources of the binary and Gray code:
- (1) Run-length
- (2) Huffman coding
- (3) Shannon-Fano coding
- (4) Arithmetic coding
- (5) Dictionary compression

III. Your program should report statistics of changes in the process. The compression efficiency should also be reported using the analytical techniques learnt from the course. While being encouraged to design good user interfaces, you are welcome to use essential user interfaces, such as tables (or diagrams).

ANSWER:

A. Algorithms

The flow-chart of Run Length Algorithm applied. (Rachesti, et al., 2017)

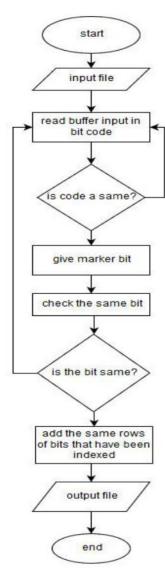


Figure 1 Run Length Algorithm flow-chart

B. Design

The class diagram for assignment code. For this class diagram creation, online tool draw.io https://www.draw.io/ was used.

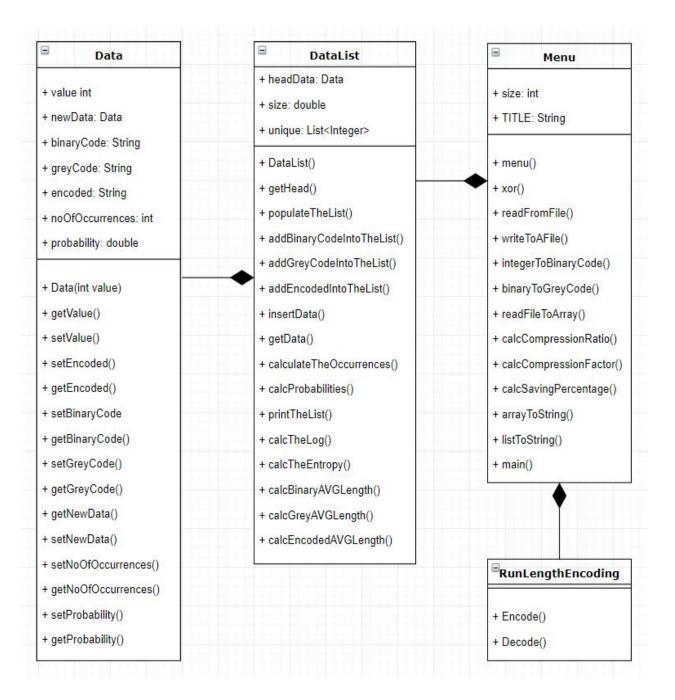


Figure 2 Class diagram

C. Demonstration

1. The input.txt file data for testing a program.



```
File Edit Format View Help

249,244,240,230,209,233,227,251,255,

248,245,93,81,120,97,193,254,

250,170,133,94,137,120,104,145,253,

241,116,118,107,134,138,96,92,163,

277,142,121,113,124,115,107,71,179,

234,106,84,125,97,108,125,106,204,

241,202,102,132,75,73,141,246,252,

253,252,24,239,178,199,242,250,245,

255,249,244,250,226,231,240,251,253
```

Figure 3 Input.txt file

2. The **output.txt** files . The output of the program code, consist of:

File 1:

- Data Input (original input.txt file);
- Binary data from outputBinary.txt file;
- Grey data from outputGrey.txt file;

```
output.txt - Notepad
File Edit Format View Help
CO3325 Data Compression coursework 2
 by RIZZO-anastasia 140359547
Data Input:
249,244,240,230,209,233,227,251,255,
248,245,93,81,120,97,193,254,
250,170,133,94,137,120,104,145,253,
241,116,118,107,134,138,96,92,163,
277,142,121,113,124,115,107,71,179,
234,106,84,125,97,108,125,106,204,
241,202,102,132,75,73,141,246,252,
253, 252, 24, 239, 178, 199, 242, 250, 245,
255,249,244,250,226,231,240,251,253
BinaryCode Data from outputBinary.txt file:
GreyCode Data from outputGrey.txt file:
```

Figure 4 File 1

File 2:

- Binary data after Run Length Encoding;
- Grey data after Run Length Encoding;

```
BinaryCode Data RLE:
11101120113011101120111021204130312011206120411011102120
111011101110411061202120812011101120212011102110111031
11101120511031301110712041301120311011202110311021
111081201110211071101120311021101212011
11103120111021101130112021101110212051101130211011102110211011201120
111011202120111011101110411041108120211011202110111011102110
111011102110211011201120112011201120111011101110111011102110111041101130111011103110111031\\
GreyCode Data RLE:
41102110511021104120312021102110411011105120141
411021204110511021102120411021102120311021107110
5110213021101110211031102110112031101110211031109110
41104110512011101120211021102110212021102110311051102110
312011203110911051106110112011203120211011
31102110111011103120211081102120111021207110111031102110
411031103110411031204120111031105110111081
```

Figure 5 File 2

File 3:

- Encoded input data from input.txt file;
- Decoded input from encoded data;

```
Encoded data:
1214191224121410121310121019122322171215111225
1214181214151913182112101917111913121514
12151011171011231914111317111210111014111415121513
12143116211811101711131411131819161912111613
1227111412111231131112142115111027211719
12131411101618141112151917111018111215111016121014
12141112101211101211131217151713111411121416121512
1215131215221412131911171811291214221510121415
122512141912241215102216121311121410121511121513
Decoded data:
249244240230209233227251255
248245938112097193254
25017013394137120104145253
2411161181071341389692163
27714212111312411510771179
2341068412597108125106204
2412021021327573141246252
25325224239178199242250245
255249244250226231240251253
```

Figure 6 File 3

File 4:

Calculations:

- Compression Ratio;
- Entropy;
- Average length;
- Compression factor;
- Saving percentage.

```
Compression Ratio [Integer to Binary code] = 2.2052401746724892
Compression Ratio [Integer to Grey code] = 2.2052401746724892
Compression Ratio [Integer to Encoded data] = 1.7991266375545851

Entropy [Integer] = 84.64981096063539

Average length [Binary code] = 62.64473684210525

Average length [Grey code] = 62.64473684210525

Compression Factor [Integer to Binary code] = 0.4534653465346535

Compression Factor [Integer to Grey code] = 0.4534653465346535

Compression Factor [Integer to Encoded data] = 0.5558252427184466

Saving percentage [Integer to Binary code] = -120.5240174672489

Saving percentage [Integer to Grey code] = -120.5240174672489

Saving percentage [Integer to Encoded data] = -79.91266375545851
```

Figure 7 File 4

D. Discussion

In **Task I** the program code read an input data from input.txt file, convert it to the Binary and Grey codes, and send the result to the text file called *outputBinary.txt* and *outputGrey.txt*. The decision to create 2 files instead of 1 (file *output.txt*) was made in order to have a better code visualisation.

The number of a greyscale images, which represented together with its an array of integers, were found in internet after a quite long search.

The first choice was given to an image of handwriting number 8 used for machine learning purposes (Figure 8). Unfortunately, this image has a lot of zeros and only a little bit of visual data. During testing it shows confusing results.

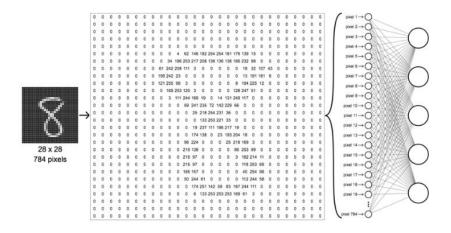
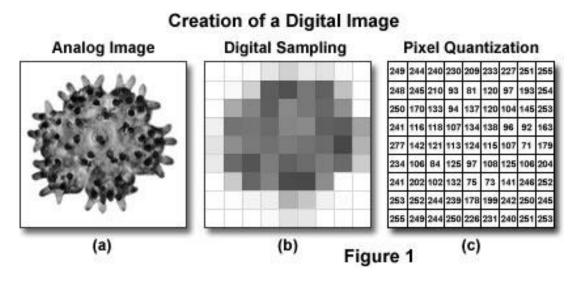


Figure 8 Grayscale image

Therefore, another image with better array of an integers was chosen from Hamamatsu web site, "Basic Properties of Digital Images" publication (Spring, et al., n.d.):



Figure~9~Screenshot~form~http://hamamatsu.magnet.fsu.edu/articles/digitalimage basics.html

For testing purposes 4 different input files were created from array of integers [0-255]: separated by space; separated by commas; separated by commas and space; all together in 1 line without spaces or commas between.

For this coursework was chosen an input file separated by commas without spaces.

In **Task II** the Run Length lossless compression algorithm (RLE) was implemented in a program code and applied to the data from *outputBinary.txt* and *outputGrey.txt*. The encoded Binary and Grey codes data were shown as a result.

Please, note: The Run Length algorithm was taken from the GitHub source (Yang, 2014), but changes in original code were applied as well.

In **Task III** the code, which reports statistics of changes in the process, was implemented. The following calculations was made:

- Compression Ratio;
- Entropy;
- Average length;
- Compression factor;
- Saving percentage.

An input file was decoded back through Run length decoding. All data were printed as an output.

After all calculations has been made, the statistic report of changes in a table below shows:

- the Compression ratio of Binary and Grey codes are the same (2.2052401746724892), but of RLE is smaller (1.7991266375545851);
- the AVG length of Binary and Grey codes are the same (62.64473684210525);
- the Compression factor of Binary and Grey codes are the same (0.4534653465346535), but of RLE is bigger (0.5558252427184466);
- the Saving percentage of Binary and Grey codes are the same (-120.5240174672489), but of RLE is bigger (-79.91266375545851);

Table for reporting statistics of changes in the process.

Parts	Prompt for Input	Display for Output
Encoding	249,244,240,230,209,233,227,251,255,248, 245,93,81,120,97,193,254,250,170,133,94, 137,120,104,145,253,241,116,118,107,134, 138,96,92,163,277,142,121,113,124,115,107, 71,179,234,106,84,125,97,108,125,106,204, 241,202,102,132,75,73,141,246,252,253,252, 24,239,178,199,242,250,245,255,249,244,250, 226,231,240,251,253	1214191224121410121310121019122322171215 1112251214181214151913182112101917111913 1215141215101117101123191411131711121011 1014111415121513121431162118111017111314 1113181916191211161312271114121112311311 1214211511102721171912131411101618141112 15191711101811121511101612101412141112101 21110121113121715171311141112141612152215 13121522141213191117181129121422151012141 51225121419122412151022161213111214101215 11121513
Decoding	1214191224121410121310121019122322171 2151112251214181214151913182112101917 1119131215141215101117101123191411131 7111210111014111415121513121431162118 1110171113141113181916191211161312271 1141211123113111214211511102721171912 1314111016181411121519171110181112151 1101612101412141112101211101211131217 1517131114111214161215221513121522141 2131911171811291214221510121415122512 1419122412151022161213111214101215111 21513	249244240230209233227251255248245938112097 193254250170133941371201041452532411161181 071341389692163277142121113124115107711792 341068412597108125106204241202102132757314 124625225325224239178199242250245255249244 250226231240251253
Binary Data from output Binary.txt file	249,244,240,230,209,233,227,251,255,248, 245,93,81,120,97,193,254,250,170,133,94, 137,120,104,145,253,241,116,118,107,134, 138,96,92,163,277,142,121,113,124,115,107, 71,179,234,106,84,125,97,108,125,106,204, 241,202,102,132,75,73,141,246,252,253,252, 24,239,178,199,242,250,245,255,249,244,250, 226,231,240,251,253	10 100 1001 10 100 100 10 10 10 11 0 10 1
Grey Data from output Grey.txt file	249,244,240,230,209,233,227,251,255,248, 245,93,81,120,97,193,254,250,170,133,94, 137,120,104,145,253,241,116,118,107,134, 138,96,92,163,277,142,121,113,124,115,107, 71,179,234,106,84,125,97,108,125,106,204, 241,202,102,132,75,73,141,246,252,253,252, 24,239,178,199,242,250,245,255,249,244,250, 226,231,240,251,253	11 110 1101 11 110 110 11 110 0 11 10 0 11 0 1101 11 10 10 11 11 110 11 111 1

Binary Data from output Binary.txt file after Run Length Encoding	10 100 1001 10 100 100 10 100 0 10 11 0 10 0 1001 10 11 11	11101120112021101120111011301110211011 201120211051101110411011103110111021102110 112011301110112011102120413031201120612041 101110212011101110111041106120212081201110 112021201110211011104110112051103130111071 204130112031101120211031103110812011102110 711011203110211012120211031201110211011301 120211011102120511011302110111021102110112 011201110112021201110111
Grey Data from output Grey.txt file after Run Length Encoding	11 110 1101 11 110 110 11 110 0 11 10 0 11 0 1101 11 1	411021105110211041203120211021104110111051 201811021204110511021102120411021102120311 021107110511021302110111021103110211011203 110111021103110911041104110512011101120211 021102110212021102110311051102110312011203 110911051106110112011203120211041102110111 011103120211081102120111021207110111031102 110411031103110411031204120111031105110111 014110111103110211031202120311031105110711 041101511021105110211051105110411051201211 0
Computing		
Compression Ratio (Binary)	10 100 1001 10 100 100 10 100 0 10 11 0 10 0 1001 10 11 11	2.2052401746724892
	11 110 1101 11 110 110 11 110 0 11 10 0 11 0 1101 1101 11 1	

Compression Ratio (Grey)	0 100 1 10 110 1 10 1100 1101 101 1101 11 1	2.2052401746724892
Compression Ratio (RLE)	1214191224121410121310121019122322171 2151112251214181214151913182112101917 1119131215141215101117101123191411131 7111210111014111415121513121431162118 1110171113141113181916191211161312271 1141211123113111214211511102721171912 1314111016181411121519171110181112151 1101612101412141112101211101211131217 1517131114111214161215221513121522141 2131911171811291214221510121415122512 1419122412151022161213111214101215111 21513	1.7991266375545851
Entropy (Integer)	249,244,240,230,209,233,227,251,255,248, 245,93,81,120,97,193,254,250,170,133,94, 137,120,104,145,253,241,116,118,107,134, 138,96,92,163,277,142,121,113,124,115,107, 71,179,234,106,84,125,97,108,125,106,204, 241,202,102,132,75,73,141,246,252,253,252, 24,239,178,199,242,250,245,255,249,244,250, 226,231,240,251,253	84.64981096063539
AVG length (Binary)	10 100 1001 10 100 100 10 100 0 10 11 0 10 0 1001 10 11 11	62.64473684210525
AVG length (Grey)	11 110 1101 11 110 110 11 110 0 11 10 0 11 0 1101 11 10 10 11 11 100 11 111 1	62.64473684210525

	1100 1 11 111 1 0 101 11 0 110 11 110 1 11 0 11 1 1 0 11 1 1 0 11 1 1 1 0 11 1 1 0 11 1 1 0 11 1 1 0 1	
Compression Factor (Binary)	10 100 1001 10 100 100 10 10 0 10 11 0 10 0 1001 10 11 11	0.453465346535
Compression Factor (Grey)	11 110 1101 11 110 110 11 110 0 11 10 0 11 0 1101 11 10 10 11 11 110 111 110 11 11 11	0.453465346535
Compression Factor (RLE)	1214191224121410121310121019122322171 2151112251214181214151913182112101917 1119131215141215101117101123191411131 7111210111014111415121513121431162118 1110171113141113181916191211161312271 1141211123113111214211511102721171912 1314111016181411121519171110181112151 1101612101412141112101211101211131217 1517131114111214161215221513121522141 2131911171811291214221510121415122512 1419122412151022161213111214101215111 21513	0.5558252427184466

Saving percentage (Binary)	10 100 1001 10 100 100 10 100 0 10 11 0 10 0 1001 10 11 11 10 10 111 10 101 1 10 101 101 10 100 10	-120.5240174672489
Saving percentage (Grey)	11 110 1101 11 110 110 11 110 0 11 10 0 11 0 1101 11 1	-120.5240174672489
Saving percentage (RLE)	1214191224121410121310121019122322171 2151112251214181214151913182112101917 1119131215141215101117101123191411131 7111210111014111415121513121431162118 1110171113141113181916191211161312271 1141211123113111214211511102721171912 1314111016181411121519171110181112151 1101612101412141112101211101211131217 1517131114111214161215221513121522141 2131911171811291214221510121415122512 1419122412151022161213111214101215111 21513	-79.91266375545851

Figure 10. Table for reporting statistics of changes in the process

References

Rachesti, D. A., Purboyo, T. W. & Prasasti, A. L., 2017. *Comparison of Text Data Compression Using Huffman, Shannon-Fano, Run Length Encoding, and Tunstall Methods.* s.l., s.n., pp. 13618-13622.

Spring, K. R., Russ, J. C. & Davidson, M. W., n.d. *Hamamatsu*, s.l.: s.n.

Yang, 2014. Algorithms, s.l.: s.n.