NCU 2022

Data Structure

Assignment #4

ARCD & Huffman Comparison

鄭宇軒

通訊二 · 110503534

Objective

This project aims to compare the performance of two compression algorithms —— **ARCD** and **Huffman**.

Two open source libraries have been used:

ARCD: https://github.com/wonder-mice/arcd

Huffman: https://github.com/drichardson/huffman

The time required to perform text encoding and decoding using each algorithm will be recorded.

Since there are different factors affecting the result of each run, the algorithms are run multiple times. The **average time is obtained for better accuracy**.

The time used for file IO is not counted, only encoding and decoding.

Test Program

A test program **`compare_test.c`** has been created to test the performance. It will encode and decode a given text file using the two algorithms.

The user can specify the test file and the number of iterations in the command line.

Each time the algorithms run, the starting and ending time will be recorded using the **`libev`** library. A time difference value, in milliseconds, will be returned.

When all the iterations have been completed, the program calculates the average time results.

Compiling the test program

```
gcc -o compare_test.out compare_test.c -lm -lev ./arcd/arcd.c
./arcd/arcd_test.c -lev ./arcd/adaptive_model.c
./huffman/huffman_test.c -lev -lm ./huffman/huffman.c
```

Running the test program

```
./compare_test.out <PATH_OF_TESTDATA> <ITERATIONS>
```

Sample

```
DSA Assignment 4/main$ ./compare_test.out data_1Mw.txt

Running test using 'data_1Mw.txt' for 100 iterations...

===== Results =====
ARCD Encode:  0.405693 ms
Huff Encode:  0.400036 ms
Difference:   0.005658 ms

ARCD Decode:  0.574303 ms
Huff Decode:  0.561552 ms
Difference:   0.012751 ms


Test complete.
```

Findings & Analysis

- *Tested on Ubuntu 22.10, kernel 5.19.0*
- *Using 100 iterations*

Encoding

Data	Size	ARCD (ms)	Huffman (ms)	Difference (ms)
1 character	2 B	0.011661	**0.010281**	0.001380
100 words	654 B	0.341578	**0.325866**	0.015712
1k words	631 B	0.336762	**0.326169**	0.010593
10k words	662 B	0.361276	**0.350018**	0.011258
1M words	681 B	0.405693	**0.400036**	0.005658

Decoding

Data	Bytes	ARCD (ms)	Huffman (ms)	Difference (ms)
1 character	2 B	0.012023	**0.012014**	0.000010
100 words	654 B	0.492827	**0.486054**	0.006773
1k words	631 B	0.475531	**0.469802**	0.005729
10k words	662 B	0.509675	**0.507715**	0.001960
1M words	681 B	0.574303	**0.561552**	0.012751

From the data, we can see that the **Huffman algorithm uses less time** than ARCD for **both encoding and decoding**.

The time used by both algorithms increases as the number of words and data size increase.