Huffman Coding

Victor Venegas & Cristian Robles
Team 10

What is Huffman Coding?

- Lossless data compression algorithm
- Goal is to reduce the number of bits needed
- Based on frequency of letters, more frequent characters get shorter bits and less frequent get longer bits

Introduction to Huffman Coding

Numerical example - looking at the word ‘aggressive’

a g g r e s s i v e

1 2 3 4 5 6 7 8 9 10

each character uses 8 bits and because this word is 10 characters long this word uses 80 bits, without any compression

a - 010

g - 110

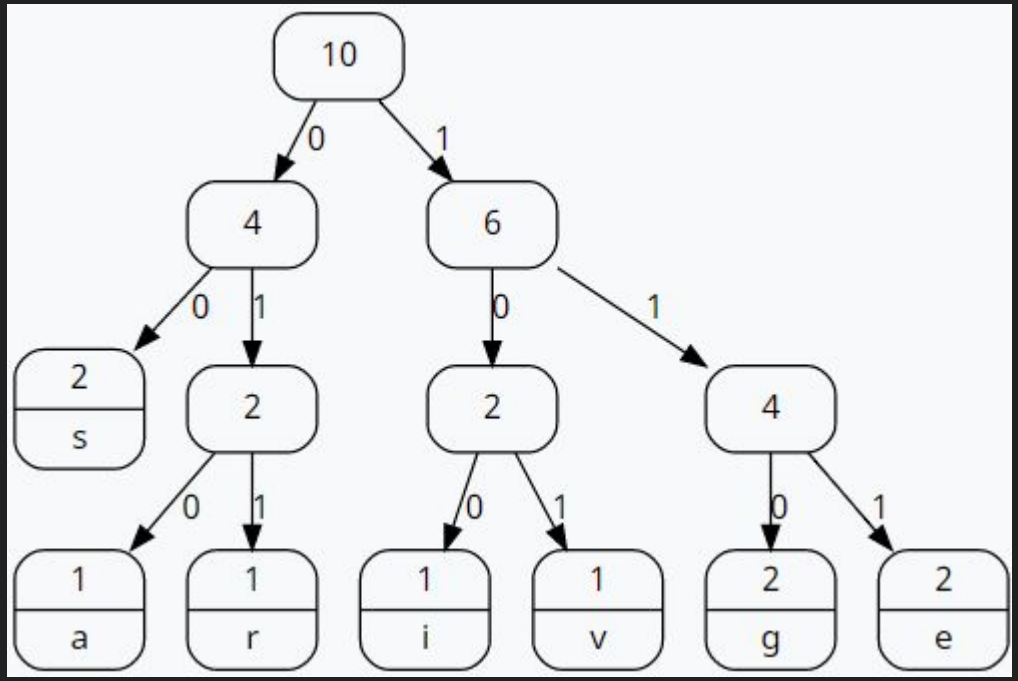
r - 011

e - 111

s - 00

i - 100

v - 101

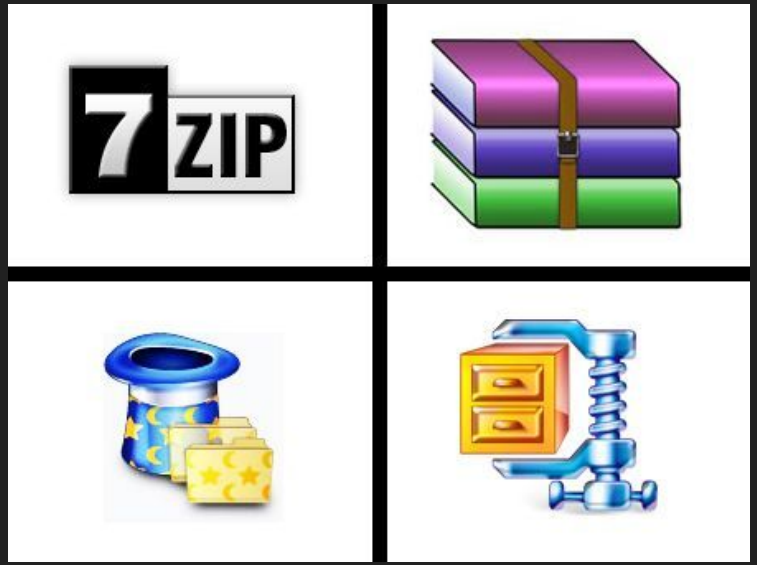


Huffman Tree

After using Huffman Coding we can now store ‘aggressive’ only using 35% of the bits originally used. It now comes out to
010 110 110 011 111 00 00 100 101 111 (28 bit)
instead of
01100001 01100111 01100111 01110010 01100101 01110011 01110011 01101001 01110110 01100101 (80 bit)

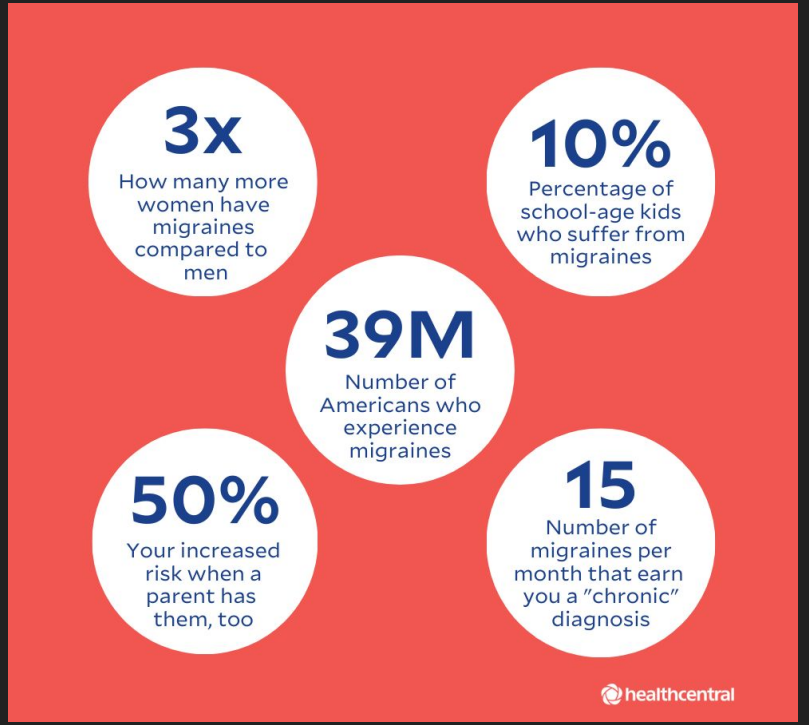
Introduction to Huffman Coding

Application fields - not solely used, more of a foundation used in programs that are used for compressing and decompressing files



Our Datasets Explained

In many other topics the accuracy of the output data is heavily reliant on the input. For example, if you are trying to predict if someone is likely to get a migraine it is necessary to have a large enough dataset to find a pattern that predicts migraines



Our Datasets Explained - Pt 2

Huffman codes however are not the same since we do not use them to predict anything and all compression is the same whether it be compressing an essay, a letter, or a birth certificate. So for our datasets we chose to compress old documents that can’t be deleted because they are important but aren’t currently needed freeing up space on computers

Our Datasets Examples - What we have compressed

Victor Venegas

Dr. Junliang Huang

FLIT 370

10 December 2020

Overworked Employees and Overlooked Automation

Due to recent backlash on convenience stores in Japan remaining open 24 hours 365 days a year, I have decided to research the question "Can we see a near future where automation is able to replace employees in convenience stores?" By looking at statistics, I would like to figure out if it is realistic for cashierless convenience stores to exist in a country where many people depend on these stores. More specifically, I will be looking at statistics and surveys regarding automation and what people think of automation in both Japan and the United States. I will also include crime rates to see how realistic it would be to have a cashier-less store.

During my recent trip to Japan I spent a lot of my time in their convenience stores. They were my lifesaver, providing fresh food and supplies at any time of the day. Unfortunately, going into different stores I always saw the same expression on the employee there, tired. On multiple occasions I would have to wake them up in order to pay, telling myself, "Wow, what is stopping me from just walking out with this." So with the news of these places fighting to close a couple of days a year is relieving because some of the workers were sure overworked. In the New York Times article regarding this outrage, it reads, "The government considers convenience stores part of the country's infrastructure, like highways and sewers." Although, I do believe these stores

should be always open and are essential for those who are not able to go into grocery stores or need supplies late at night and early morning, that is why I believe these jobs can be replaced by automation.

As mentioned before, Japan is seeing a lot of backlash for these stores that never close. The previously mentioned New York Times article explains this backlash, and I would like to specifically look at Mitoshi Matsumoto, owner of a 7-11 franchise store who wanted to close his store on New Year's Day and how the company responded to his request. In November 2019, Mr. Matsumoto requested to close his store on New Year's Day in order to allow himself and his two full-time employees a day off after years of working 14 hour shifts with no days of closing. He also said if his request was not approved, he would shut down the store anyway in order to protest the company's 24-hour demands. On December 20, 2019 the parent company told Mr. Matsumoto had received too many complaints of his store and he had 10 days to address these complaints, or else his store would be closed. I find this timing to be awfully convenient for the parent company considering Matsumoto wanted to close on New Year's Day. So did Matsumoto, he said the company wanted to shut him down in order to prevent others from realizing they could protest against the company too. To no one's surprise, 7-11 terminated Matsumoto's contract and closed his store. This wasn't the first time Mr. Matsumoto had issues with the company. He first opened the franchise store in 2012 with him and his wife, and unfortunately in 2018 she passed away. His wife was also a crucial employee there, sharing the long shifts with him, because he lost her he was understaffed and could not find anyone with such short notice. He requested to close his store for four hours a day until he could find a replacement but the parent company denied him, saying that would breach his contract and would cost a $155,000

Our Algorithm

Our Algorithm

Our Algorithm

Results

```
----jGRASP exec: java comp282p3
line 1 :sentence using all letters
line 2 :sphinx of black quartz judge my vow
a appears 2 times
b appears 1 times
c appears 1 times
d appears 1 times
e appears 1 times
f appears 1 times
g appears 1 times
h appears 1 times
i appears 1 times
j appears 1 times
k appears 1 times
l appears 1 times
m appears 1 times
n appears 1 times
o appears 2 times
p appears 1 times
q appears 1 times
r appears 1 times
s appears 1 times
t appears 1 times
u appears 2 times
v appears 1 times
w appears 1 times
x appears 1 times
y appears 1 times
```

Char counter

```
Huffman Code
w:0000
h:0001
l:0010
f:0011
t:01000
r:01001
x:01010
d:01011
y:01100
z:011010
m:011011
q:01110
i:01111
b:10000
c:10001
g:10010
n:10011
a:1010
o:1011
u:1100
v:11010
p:11011
j:11100
k:11101
s:11110
e:11111

----jGRASP: operation complete.
```

Huffman code representation of text file

Pros, Cons, & Limitations

Pros -

Save storage space

Most efficient in compressing longer texts

Favorable when dealing with characters that repeat themselves - like the English language

Lossless, you do not lose quality on an image or data on text

Cons -

Not very efficient when using with other languages with more characters

Limitations -

Our example only works on english characters - 256 character limit

We are limited to text files, jpg, mps etc excluded

Contributions

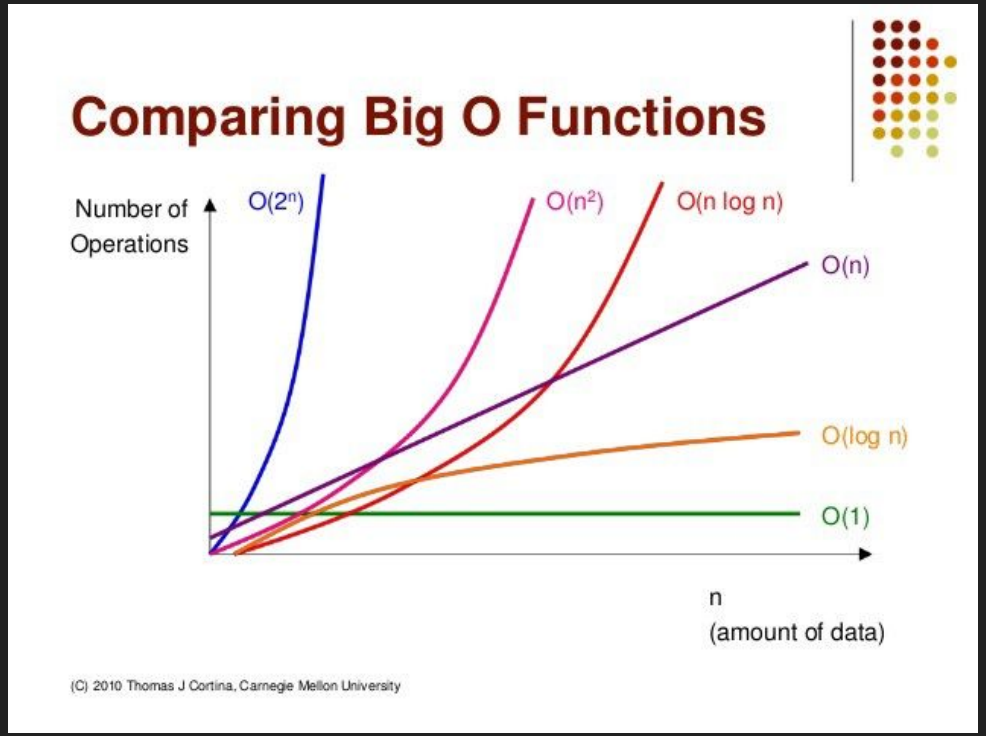
We would like to use it for the files from our previous classes as they are no longer needed but still useful to keep around in case of future necessity.

Time Complexity

Run length - O(n)

Huffman - O(nlogn)

n = number of unique characters



References - REMOVE ANNOTATIONS

A. E. El-Sharkawey, “Comparison between (RLE & Huffman and DWT) Algorithms for Data Compression,” *ResearchGate*. [Online]. Available: https://www.researchgate.net/publication/331563398_Comparison_between_RLE_Huffman_and_DWT_Algorithms_for_Data_Compression. [Accessed: 2019]. -COMPARISON

A. M. A. Ibrahim, “Comparison Between (RLE And Huffman) Algorithms for Lossless Data Compression,” *IJITR*. [Online]. Available: https://core.ac.uk/download/pdf/228547034.pdf. [Accessed: 2015]. -COMPARISON

“Huffman Coding: Greedy Algo-3,” *GeeksforGeeks*, 23-Mar-2021. [Online]. Available: https://www.geeksforgeeks.org/huffman-coding-greedy-algo-3/. - intro hufffman

R. Arshad, “Performance comparison of Huffman Coding and Double Huffman Coding,” *IEEE Xplore*. [Online]. Available: https://ieeexplore-ieee-org.libproxy.csun.edu/document/7845058. [Accessed: 24-Aug-2016]. -FUTURE WORK

R. Hashemian, “Memory efficient and high-speed search Huffman coding,” *IEEE Xplore*. [Online]. Available: https://ieeexplore-ieee-org.libproxy.csun.edu/document/469442. [Accessed: Oct-1995]. -INTRO

R. Patil, “Audio compression using dynamic Huffman and RLE coding,” *IEEE Xplore*. [Online]. Available: https://ieeexplore-ieee-org.libproxy.csun.edu/document/8321256. [Accessed: 19-Oct-2017]. -COMPARISON (AUDIO)

“Run Length Encoding,” *GeeksforGeeks*, 23-Mar-2021. [Online]. Available: https://www.geeksforgeeks.org/run-length-encoding/. -intro RLE