


Huffman


Your *LAST* Assignment



Announcements

- APT Set 6 – Due April 9
- Huffman – Due April 16
 - Burrows Wheeler Extra Credit – Due April 18

2




Today

- Bitwise representations
- File compression
- Huffman coding

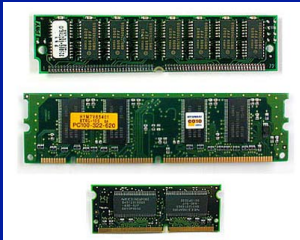
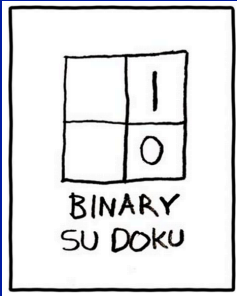
- Things you should know for the Huffman assignment

3




Primitives

- How are characters stored in memory?
 - 011101000110001101110000




4



ASCII

- American Standard Code for Information Interchange
 - Character encoding scheme
 - Characters mapped to numbers
 - A – 65
 - a – 97
 - ' ' (space) – 32

5



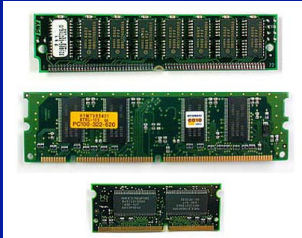
ASCII

t	c	p
116	99	112
01110100	01100011	01110000

<http://en.wikipedia.org/wiki/ASCII> 6

Primitives

- How are characters stored in memory?
 - 011101000110001101110000



7

ASCII

- 011101000110001101110000

Each 0 or 1 is a 'bit'
Each character is made of 8 bits
8 bits = 1 byte

8

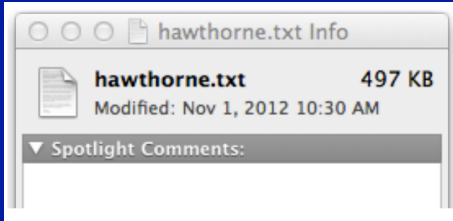
Memory

- 8 bits = 1 byte
- 1024 bytes = 1 kb
- 1 character saved in 1 byte

497 kB (kiloBytes)*

508,928 Bytes

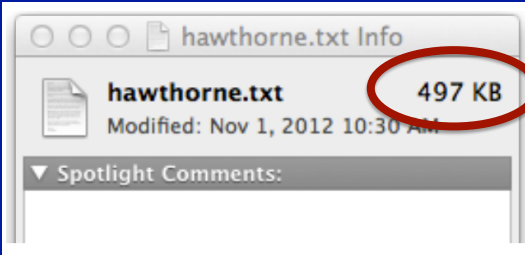
508,928 characters



A screenshot of a file information window titled 'hawthorne.txt Info'. It shows the file name 'hawthorne.txt' with a size of '497 KB' and a modification date of 'Nov 1, 2012 10:30 AM'. Below this is a section for 'Spotlight Comments'.

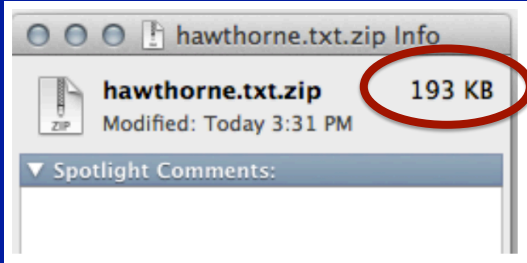
9

Compression



A screenshot of a file information window titled 'hawthorne.txt Info'. The file name is 'hawthorne.txt', the size is '497 KB', and the modification date is 'Nov 1, 2012 10:30 AM'. The '497 KB' value is circled in red.

How do they do that???



A screenshot of a file information window titled 'hawthorne.txt.zip Info'. The file name is 'hawthorne.txt.zip', the size is '193 KB', and the modification date is 'Today 3:31 PM'. The '193 KB' value is circled in red.

10



Compression

- ASCII – map each character to number represented as 8-bits
- Have to find a new mapping that uses fewer bits
 - 8 bits represent 256 characters
 - How do we represent 256 characters in fewer than 8 bits per character?

11



Game time



12




The image shows the iconic 'Wheel of Fortune' logo in large, glowing, golden letters. Below the logo is a circular game board with various colored segments containing letters and numbers. The scene is set against a night cityscape background with spotlights illuminating the board.

Compression

- What have we learned from Wheel of Fortune?

13



Huffman

- Variable length encoding
 - From our file
 - some characters are more common than others
 - encode
 - common characters $<$ bits
 - uncommon characters $>$ bits


14



Huffman

- AACCCAAABAADAE
- A : 8
- C : 3
- B : 1
- D : 1
- E : 1

15



Huffman

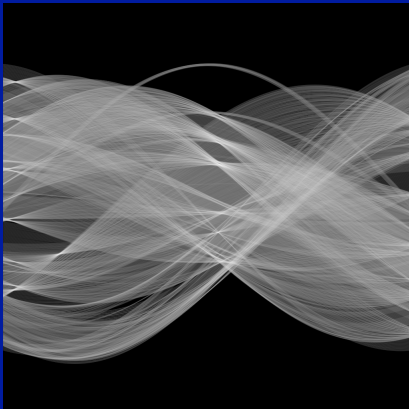

- Huffman coding
 - **I've made a huge mistake**
 - generate frequencies
 - make tree for each character
 - add trees to **PRIORITY QUEUE**
 - while (more than one tree)
 - remove two smallest trees (from **PRIORITY QUEUE**)
 - merge trees
 - add new tree to **PRIORITY QUEUE**

<http://goo.gl/Mq9xa>

16

Huffman

- Huffman (Huff) \neq Hough
- Hough transforms – Edge detection in images



17

Today

- Bitwise representations
- File compression
- Huffman coding

- Things you should know for the Huffman assignment

18