Unit 2 – Lesson 1

Bytes and File Sizes

Getting Started…

- We are going to give our 2-minute Flash-talks in groups

Objectives: You will be able too…

- Use appropriate terminology when describing the size of digital files
- Identify and compare the size of familiar digital media
- Solve small world problems that require reasoning about file sizes

Getting Started…

- As we embark on a new unit about Data and Digital Information we need to get familiar with terminology about data and different types of data files

Terminology – Byte

- Recall that a single character of ASCII text requires 8 bits.
- The technical term for 8 bits of data is a Byte
- A **byte** is the standard fundamental unit (or "chunk size")underlying most computing systems today.
- You may have heard "megabyte", "kilobyte", "gigabyte", etc. which are all different amounts of bytes.
- Even if we want to store 1-bit, you usually have to use a whole byte to do it.

Why is a Byte 8 bits?

- The 8-bit byte was not always standard. Computers used many different "byte" sizes over the course of history
- Since a lot of data and computer instructions were encoded in ASCII text, 8-bits became common, whence the 8-bit byte.
- 8-bit chunks meant a computer could "bite" off 8 bits at a time, we started calling it the "byte" (bite could be confused with bit)

Comparing sizes of plain text vs. MS Word doc.

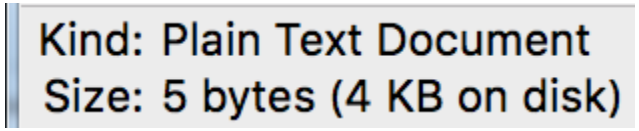
- Recall that we learned that besides simply the text, we often need to store information about formatting
- We might wonder how many bytes we need to store to include formatting
- If a single ASCII character is one byte then if we were to store the word “hello” in a plain ASCII text file in a computer, we would expect it to require 5 bytes (40 bits) of memory
- What about in Microsoft Word?

Comparing sizes of plain text vs. MS Word doc.

- Predict: How many more bytes will a Word doc require to store the word “hello” than a plain text document?
- Let’s check it out.

Text Files

- A 5-byte file is so small that some computers won't allocate a chunk of memory that small. For example you might see something like this:



- Which indicates that even though the file is 5 bytes, it's taking up 4 Kilobytes of memory on your computer.

Activity: Rapid Research – Bytes and File Sizes

- Modern data files typically measure in the thousands, millions, billions, or trillions of bytes.
- Let's get a little practice looking at files and how big they are.

Activity: Rapid Research – Bytes and File Sizes

- We are going to use the "Bytes and File Sizes – Activity Guide"
- Reference:
  - https://web.stanford.edu/class/cs101/bits-gigabytes.html
  - http://www.computerhope.com/issues/chspace.htm

Discrepancies:

- For computers, it's convenient to organize things in groups of powers of 2
  - "kilobyte" can mean 1024 or 1000 bytes
  - "megabyte" can mean 1,048,576 bytes or 1 million
- For this class, we don't need to worry about the exact number, but recognize that marketers vs. people in the field of computer science may see these terms differently

Wrap-up:

- As you can see, data files size can grow very quickly in size.
- There is a lot of data… and we want to send a lot of it over the Internet
- There is a problem though: we are limited by speed of the Internet connection (there is a physical limit to how fast you can transmit bytes"
- What if the data is big enough that even with fast internet connection, it takes an unreasonable amount of time? How do we solve this problem? (tomorrow we will learn about **compression**)

Review for test tomorrow!

- Go ahead and start looking over the review and let me know if you have questions.