Files and File I/O

What is a file?

What is a file?

We deal with files all the time. There are many kinds. What is the difference between them?

A file stores data

Files store data

What is data?

Why is it useful?

Ones and Zeros

At a low-level, to a computer everything is a number.

And those numbers are made up of ones and zeros.

From this point of view, “data” is just a sequence of ones and zeros!

What is a file?

To a computer, a file is just a sequence of ones and zeros.

Bits and Bytes

“Bit” is short for “binary digit” and is a 1 or 0.

Byte is a binary number consisting of 8 bits.

Binary

Base-2 counting system

Our regular human counting system is base ten:

1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, …

Binary is based two:

1, 10, 11, 100, 101, 110, 111, 1000, 1001, 1010, …

When we say “3,102” what does that mean?

How do you convert from decimal to binary?

And vice versa?

100101 = ?

512 = ?

Working with binary can be awkward

Hexadecimal

Base-16 counting system

Since we only have ten digits, hexadecimal needs to add 6 more. The letters A - F are used.

1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 1A, 1B, 1C, 1D, 1E, 1F, 20, 21 …

98, 99, 9A, 9B, 9C, 9D, 9E, 9F, A0, A1, A2

E0, E1, E2, …. EF, F0, F1, F2, F3, FF, 100 … ?

Why hexadecimal?

Hexadecimal

Because 16 is a power of 2, converting between binary and hexadecimal is very easy!

Four bits convert to one hexadecimal digit and vice versa

0000 -> 0   0001 -> 1

Hex editors

If all files are just sequences of bytes, why don’t they look that way?

Can we look at the bytes in a file?

But what do we do with a bunch of bytes?

Computers may like bytes, but humans want computers to handle text, colors, images, etc.

How do bytes turn into those things?

Encodings

The way (human) symbols and meanings are converted to bytes is called “encoding” and the reverse process is called “decoding”.

ASCII

One early encoding of characters (letters, numbers, punctuation etc.) still in wide use

One byte per character

Has some limitations

Unicode

Designed to overcome limitations of ASCII

Can encode any character from any human writing system

Multiple bytes per character

Other encodings

Graphics: BMP, PNG, JPEG

Audio: WAV, MP3, FLAC

Video: MP2, Quicktime, …

Lots of others!