COMP10001

WORKSHOP #10

o(╥﹏╥)o

Shevon Mendis
shevonm@unimelb.edu.au

Number Systems

- In essence, a number system is the way we represent numbers

- Common number systems:

Name	Base	Range	Conversion Function	Prefix	Example
Decimal	10	[0, … , 9]	`int`(num[,*base=10*])		`419`
Binary	2	[0, 1]	`bin()`	`0b`	`0b1000101`
Octal	8	[0, … , 7]	`oct()`	`0o`	`0o644`
Hexadecimal	16	[0, … , 15]	`hex()`	`0x`	`0x45`

- **Note:** For Hexadecimal, the characters `A, B, C, D, E, F` are used to represent `10, 11, 12, 13, 14, 15`
  - eg. `31 == 0x1F`

Number Systems

- Why is Binary useful?

  - Similar to how our *native* numbering system is Decimal, computers use Binary, so all the data stored is stored in Binary and processors are designed to perform calculations using those Binary values

- Why is Hexadecimal useful?

  - Higher information density: a single Hex character can represent 4 consecutive bits (ie. 2 Hex characters can represent 1 byte)
    - eg. `0b11011000 == 0xD8`

  - Often used to describe precise memory locations, as well as for specifying colours
  - Why do we use hexadecimal? - Niko Savas

Encoding Standards

- ASCII

  - **A**merican **S**tandard **C**ode for **I**nformation **I**nterchange
  - A standard that contains a unique mapping for every character in the **English** alphabet (alongside other characters like punctuation symbols and spaces, of course)

- Unicode

  - A standard for character encoding that provides a unique **code point** for a supernumerary of **graphemes**
  - Solved the problem that different languages had their own encoding standards, which were incompatible with one another and so, made it impossible to decode documents that contained multiple languages
  - Characters, Symbols and the Unicode Miracle - Computerphile

Text Document Encoding

- Fixed Length Encoders: All the characters will be represented using the same number of bits

  - ASCII
    - Uses 7 bits to represent a single character
    - Can encode up to 127 (2^7 - 1) characters
    - Characters are usually stored using a byte (8 bits) so there's a single redundant bit at the start

  - UTF-32 (Unicode Transformation Format 32-bit)
    - Uses 32 bits to represent a single character, so many more characters can be encoded
    - Wastes a tremendous amount of space since many representations will have redundant bits added to the front of it
      - eg. **A → 00000000000000000000000001000001**

◀ **Text Document Encoding** ▶

- Variable Length Encoders: Different characters may be represented using a different number of bits

  - ISO-8859
    - Adds a single bit to the front of an ASCII bit sequence, allowing an extra 128 characters to be represented (ISO-8859-1)
    - Since adding 1 extra bit doesn't allow a substantial number of extra orthographies to be represented, it comes in 16 variations (ISO-8859-1 to ISO-8859-16)

  - UTF-8
    - Uses 1 to 4 bytes to represent a single character, depending on how much space is needed for it
    - [How UTF-8 Unicode encoding works](#)

  - UTF-16
    - Uses 2 to 4 bytes to represent single character, depending on how much space is needed for it

ACM Code of Ethics

- Contribute to society and to human well-being, acknowledging that all people are stakeholders in computing

- Avoid harm

- Be honest and trustworthy

- Be fair and take action not to discriminate

- Respect the work required to produce new ideas, inventions, creative works, and computing artefacts

- Respect privacy

- Honour confidentiality

Dual Use

- In computing, *Dual Use* refers to technologies that can equally be used for good or malicious purposes

- Examples:
  - AI
    - Surveillance
    - Deep Fakes ([Deep Fake VFX - Fresh Prince of Bel-Air (ft. Azerzz)](#))
  - Nuclear technology

Announcements

- Project 2 due at 11.59PM on Wednesday, June 3rd

  - Try to avoid making the same mistakes you made in Project 1 (if any)
  - Submit on time (submitting at 11.59PM is considered late!)
  - *It wouldn't hurt to include meme refs or jokes within your submission*

- Final exam is set for Wednesday, June 17th (as per exam timetable)

  - Will be held on Grok, similar to the MST
  - A practice exam is available on Grok

(づ｡◕‿‿◕｡)づ **Thank You** (づ｡◕‿‿◕｡)づ

So long...