COMP10001
Foundations of Computing
Semester 1, 2021
Tutorial 11

Andrew Naughton

andrew.naughton@unimelb.edu.au

Outline

- ❖ Number Systems
- ❖ Encoding Standards
- ❖ Text Document Encoding
- ❖ ACM Code of Ethics
- ❖ Dual Use
- ❖ Exercises

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: In Hexadecimal, "A", "B", "C", "D", "E" are used for 10, 11, 12, 13, 14, 15, resp.

Number Systems

❖ Why is binary useful?
  ❖ Similar to how our native numbering system is Decimal, computers use Binary, so all data is stored in Binary and processors are built to perform calculations using Binary values

❖ Why is Hexadecimal useful?
  ❖ Higher information density: a single Hex character can represent 4 consecutive bits. Since a byte is 8 bits, 2 Hex characters can represent 1 byte.
  ❖ E.g. `0b11011000 == 0xD8`
  ❖ Used to describe precise memory locations, as well as for specifying colours

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 space)

❖ Unicode
  ❖ A standard that contains a unique mapping for every character in <u>every</u> language of the world
  ❖ Solves the problem that in the past, different languages had their own encoding standards, preventing multiple languages from appearing in the document

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 (i.e. $2^7 - 1$) 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 space, since many representations will have redundant bits added to the front

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
  ❖ Since adding 1 extra bit doesn't allow a substantial number of extra characters, 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

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

ACM Code of Ethics

❖ Having knowledge of programming is powerful; can augment human abilities, allowing people to deploy systems at scale and change the world in both small and large ways

❖ Without a sense of ethics, computing can be dangerous and damaging

❖ Association for Computing Machinery (ACM) have a Code of Ethics and Professional Conduct – next slide has some of them

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 which can be equally used for good and malicious purposes

❖ Many implementations of artificial intelligence (AI) and automation can be classified as dual use

❖ E.g.
  ❖ Surveillance
  ❖ Deepfake technology

Exercises