COMPSCI 1JC3
Introduction to Computational Thinking
Fall 2017

11 Data

William M. Farmer

Department of Computing and Software
McMaster University

November 24, 2017

McMaster University

Admin

- Final exam will be held on Fri., Dec. 8 at 12:30pm.
  - Will cover the material for the entire term.
  - 40 multiple choice questions.
  - One-stage format.
  - Review session in class on Mon., Dec. 4.
- Course evaluation.
  - Course discussion session today at 5:30 in ETB 230.
  - CS 1JC3 survey on Avenue.
  - Online course evaluations begin on Thu. at 10:00.
- Question and answer session on careers in computing on Wed., Dec. 6.
- Office hours: To see me please send me a note with times.
- Are there any questions?

Advice

- Constantly improve your communication skills!
  - Good oral and written communication skills are crucial for a successful career in almost an area.
  - Communication skills are a powerful tool for opening doors of opportunity.
- Develop a well-rounded education!
  - Programs in technical fields of study, like computing and engineering, provide a good education in the technical field but are often weak in the liberal arts.
  - Computing students should work hard to compensate for this imbalance in their education.

Review

1. Recursion and induction.
2. Little languages method.
3. Copy, modify, compare, and generalize method.
4. Private virtual networks (VPNs).
5. IP tunneling.

What Computers Do

- Computers store and manipulate information.
- The information is represented by various kinds of data.
- The behavior of a computer is controled by algorithms implemented as programs.

Kinds of Data

- Numeric data.
- Documents.
- Digital images.
- Digital audio.
- Digital video.
- Data organized for queries.

Digital Images

- Digital images are represented as a grid of pixels.
- The resolution of the image is the dimensions of the grid.
  - The higher the resolution the more detailed the image is.
- Each pixel is assigned a color.
- Representation formats:
  1. Raster image formats with various kind of compression.
  2. Vector graphics for representing scalable graphical images.

Digital Image (iClicker)

How many bits are needed to represent an uncompressed $1000x800$ resolution digital image with 24-bit color?

A. $1000 * 800$.

B. $1000 * 800 * 8$.

C. $\boxed{1000 * 800 * 24.}$

D. $1000 * 800 * 2^{24}$.

Models for Representing Color

- RGB model.
  - The RGB model is an additive model for representing color on a monitor screen.
  - A color is a sum of three colors (red, green, and blue).
  - A color is often represented as three 8-bit bytes.
  - No color is black and full color is white.
- CMYK model.
  - The CMYK model is a subtractive model for printing color using ink.
  - A color is produced by absorbing color with four inks:
    1. Cyan ink absorbs red.
    2. Magenta ink absorbs green.
    3. Yellow ink absorbs blue.
    4. Black ink absorbs all colors.
  - Complete absorption is black and no absorption is white.

RGB Color (iClicker)

Which of the following RGB color codes specifies the color gray?

A. `#C0C0C0.`
B. `#FF0000.`
C. `#FFFF00.`
D. `#FFFFFF.`

Digital Audio

- Digital audio is produced from a continuous sound wave by periodic sampling and discrete quantization.
- Sounds outside human perception are eliminated.
- Audio files are compressed to reduce size without significantly reducing sound quality.
- There are a number of audio formats.

Digital Video

- A digital video is a sequence of digital images.
- A frame rate of about 30 frames per second is needed to achieve smooth motion.
- A video file consists of a video track, audio track, and metadata.
- The video and audio tracks are compressed for storage and transmission, and decompressed before playing.
- There are a number of video formats.

File Formats

- There are two main categories of files:
  1. Text files composed of ASCII or unicode characters.
  2. Binary files composed of bits.
- Text files can be read and edited by humans unlike binary files.
- There are many different file formats.
- The extension (suffix) of a file name (e.g., `.pdf`) is used to identify the file's format.
- XML is a customizable file format that uses begin and end tags.

File Format (iClicker)

Which of the following is the extension for an audio file format?

A. `.avi`.
B. `.mp3`.
C. `.jpg`.
D. `.tex`.

Data Compression

- Compression is used to reduce data size, while decompression recovers the original data.
- Compression techniques:
  1. ZF77 algorithm for replacing repeated strings with references to earlier occurrences.
  2. Hoffman code in which the higher the frequency of a symbol the fewer bits are used to encode it.

Data Structures

- A data structure is a structured collection of values that is created and manipulated by a computer program.
- Examples:
  - Finite sequences of values:
    - Lists, arrays, records.
    - Stacks and queues.
    - Linked lists.
  - Algebraic data types:
    - Enumerated types.
    - Sum types.
    - Product types.
    - Inductive types.
  - Trees and graphs.
  - Objects (that contain data and operations).
  - Various kinds of tables including hash tables.

Data Bases

- A data base is an organized collection of data.
- The most common kind are relation data bases in which data is organized as a collection of relations.
- Data bases are designed to be modified and queried.
- The Structured Query Language (SQL) is the standard language for querying data bases.