Introduction to
Introduction to CS

Section 000: Eugene Agichtein

Your Instructor: Prof. Eugene Agichtein

- Sept 2006-: Assistant Professor in the Math & CS department
  - Affiliate Faculty, Linguistics, Computational Life Science
  - Affiliate Faculty, Web Science Program and Health Systems Institute @ Georgia Tech

- Summer 2007: Visiting Researcher at Yahoo! Research

- 2004 to 2006: Postdoctoral Researcher at Microsoft Research
  Text Mining, Search, and Navigation group, and MSN Search/Live

- 1998-2004: Ph.D. in Computer Science from Columbia University:
  dissertation on extracting structured relations from web-scale document repositories

Who are you, and why are you here?

• Please state <your name>
• Your year and major (if you have one)
• Fill in the blank:
  I think Computer Science is _________

What is Computer Science?
Let $AB$ and $CD$ be the two given numbers not relatively prime. It is required to find the greatest common measure of $AB$ and $CD$.

If now $CD$ measures $AB$, since it also measures itself, then $CD$ is a common measure of $CD$ and $AB$. And it is manifest that it is also the greatest, for no greater number than $CD$ measures $CD$.

Euclid’s Elements, Book VII, Proposition 2 (300BC)

By the word operation, we mean any process which alters the mutual relation of two or more things, be this relation of what kind it may. This is the most general definition, and would include all subjects in the universe... Supposing, for instance, that the fundamental relations of pitched sounds in the science of harmony and of musical composition were susceptible of such expression and adaptations, the engine might compose elaborate and scientific pieces of music of any degree of complexity or extent.

Ada Byron, 1843
What is the difference between Euclid and Ada?

“It depends on what your definition of ‘is’ is.”
Bill Gates (at Microsoft’s anti-trust trial)

Geometry vs. Computer Science

• Geometry (mathematics) is about *declarative* knowledge: “what is”
  If now $CD$ measures $AB$, since it also measures itself, then $CD$ is a common measure of $CD$ and $AB$

• Computer Science is about *imperative* knowledge: “how to”

  Computer Science has little to do with beige (or translucent blue) boxes called “computers” and is not a real science.
Computer Science

“How to” knowledge:
• Ways of describing information processes (computations)
  Language
• Ways of predicting properties of information processes
  Logic

What kinds of things do we want to predict?

Science, Engineering, Other?
Science?

- Understanding Nature through Observation
  - About *real* things like bowling balls, black holes, antimatter, electrons, comets, etc.

- Math and Computer Science are about *fake* things like numbers, graphs, functions, lists, etc.
  - Computer Science is a **useful tool** for *doing* real science, but not a real science

---

Engineering?

“Engineering is **design under constraint**... Engineering is synthetic - it strives to create what can be, but it is constrained by nature, by cost, by concerns of safety, reliability, environmental impact, manufacturability, maintainability and many other such 'ilities.'...”

William Wulf
Apollo Guidance Computer, 1969

1 Cubic Foot

Why did they need to fit the guidance computer in the rocket?

Measuring Computers

- 1 bit = smallest unit of information
  - True or False
  - 0 or 1
  - If we start with 2 possible choices, and get 1 bit, we can eliminate one of the choices
How much power?

- Apollo Computer: 30720 bits of changeable memory
- Lab machines have 512 MB (RAM)
  - 1 Megabyte = 1024 Kilobytes, 1 Kilobyte = 1024 Bytes, 1 Byte = 8 bits
  - 512 MB
    > (* 512 1024 1024 8)
    \( 4294967296 \) ~ 4.3 Billion bits
    > (round (/ (* 386 1024 1024 8) 30720))
    \( 139810 \) You have 105 404 times more power than AGC

You will understand this notation soon...but don’t worry if you don’t now.

If Apollo Guidance Computer power is 1 inch, you have 2.2 miles!

Computing Power 1969-2005
(in Apollo Control Computer Units)

Moore’s Law: computing power doubles every 18 months!
Constraints Computer Scientists Face

• Not like those for engineers:
  – Cost, weight, physics, etc.
  – If ~8 Million times what people had in 1969 isn’t enough for you, wait a year and you will have 20 Million times...

• More like those for Musicians and Poets:
  – Imagination and Creativity
  – Complexity of what we can understand

So, what is computer science?

• Science
  – No: it’s about fake things like numbers, not about observing and understanding nature

• Engineering
  – No: we don’t have to deal with engineering-type constraints

• Liberal Art?
Computer Science as Liberal Arts

Trivium (3 roads)
- Grammar: study of meaning in written expression
- Rhetoric: comprehension of discourse

Quadriovium (4 roads)
- Logic: argument for discovering truth
- Arithmetic: number study
- Music: number in time
- Astronomy

We will see all of these in this class!

Key Concepts in CS 170

- **Elementary computer architecture** - what is a computer, how does it work.
- **UNIX system** - how to use the UNIX system, commands to list and navigate directory hierarchy, edit and save files, etc.
- **Programming concepts**, such as parameter passing mechanisms, objects, data abstraction, inheritance, etc. etc.
- **Programming methodologies**, such as modular programming and object oriented programming.
- **Intro to data structures** - computer programs manage information, how do you manage information? Arrays
- **Some object-oriented programming concepts**, such as inheritance and interfaces
Key Tools

- **Java** programming language
- **UNIX** operating system
- **Eclipse** development environment
- **If time:** Javascript and HTML (Elementary web programming)

Key fact of the day: **The Class Webpage**

- [http://www.mathcs.emory.edu/~cs170000](http://www.mathcs.emory.edu/~cs170000)

- Information on the website is always more correct than in any handout
  – including this one

- Will try to avoid killing trees after today.
What is a Computer?

- Computer is a machine that performs operations (instructions)
- Instructions are simple (add 2 to a number)
- Computer program is a series of instructions
- This was not always so

History of the world: the ancients

Analytical Engine
(Babbage, 1830-1871)
Designed to evaluate any mathematical formula
- punch cards
- sequential control
- branching
- loops
History of the world: foundations (30s-40s)

- **Turing**
  - Theory of computability
  - Invented stored program computer
  - “Turing test”

- **John von Neumann architecture**
  - Single-memory computer architecture
  - Almost all modern computers use this paradigm

History: the mainframe era (50s-70s)

- High-level programming languages
- Fortran, Ada, Cobol, ...
- Data structures
- Sorting
- Searching
- Routing
- Graph theory
- Internet (ARPA-net)
- ...

---

25

26
History: The PC era: 80s

- MS-DOS
- Basic
- C, C++
- Xerox PARC/Macintosh
  - WIMP
- First hard disk (5MB)

---

History: the Web (90s)

- Distributed computing
- Mosaic, Netscape, IE
- Web search
  - Alta Vista, Excite...
Web 2.0: Now

And in this course...

"I think you should be more explicit here in step two."
Modern Computer Architecture

- Computer components: CPU, memory, hard disk, floppy disk, monitor, printer, and communication devices.

CPU

- Stands for central processing unit
- Retrieves instructions from memory and executes them
- Speed measured in GHz (gigahertz) (1 GHz = 1 billion of pulses per second)
- Modern CPU can have few cores
Memory (RAM)

- Stores data and program instructions for CPU
- Memory unit: byte (a sequence of 8 bits)
  - What is a bit? It's a binary digit taking value 0 or 1
  - Memory is a volatile storage
- Programs and data need to be loaded to main memory before execution by CPU

---

How is Data Stored in RAM

- Everything is coded as a series of bits
  - Why computers use only 0/1?
- Programmers need not to be concerned about the encoding and decoding of data
  - Performed automatically
Storage Devices

- Reminder: main memory is volatile
- Programs and data are permanently stored on storage devices or secondary memory
- Storage device types:
  - Disk drive
  - CD...

Hard Disks
First General Purpose Computer: ENIAC

Input and Output Devices

- I/O is the means by which a computer receives information from the outside world and sends result back.
- Output devices: monitor, printer
- Input devices: keyboard, mouse, joystick …
Computer Programs

- Computer program is a set of instructions
- Operating systems: core software of a computer
  - Windows (XP, Vista), Unix, Linux, MacOS, ...
- Application software: any software other than operating systems
  - Web browsers, word processors, games, ...

Operating Systems (OS)

- A program that manages and controls a computer’s activities
- Any other program (internet browser, word processor) is managed by OS
- The programs you write are also managed by OS

What are the tasks of OS?
1. Controlling and monitoring system activities
2. Allocating and assigning system resources
3. Scheduling operations
Program Languages

- Instructions are specified using a computer programming language
- Three types of programming languages:
  - Machine language
  - Assembly language
  - High-level language

Machine Language

- Set of primitive instructions built into every computer
- Instructions are in the form of binary code
  - Why don't we use those languages?

I wrote this awesome program: 10001110

No, no, no. It should be: 10000110

OK, hmm, sure, I’ll just...

Geek 1

Geek 2

Just a human being
Assembly Language

- Developed to make programming easy
- The computer cannot understand assembly language
  - Need to convert assembly code to machine code

```
ADDF3 R1, R2, R3
```

High-Level Language

- English-like and easy to learn and program
- Can be understood by just looking at the code

```
area = 5 * 5 * 3.1415;
```

- Required steps to produce byte code?
Compiling Source Code

- Source code
  - A program written in a high-level language
- Compiler
  - Used to translate the source program into a machine language program called an object program
- Linker
  - Used to link your object program with supporting libraries

Some Popular High-Level Languages

- COBOL (COmmon Business Oriented Language)
- FORTRAN (FORmula TRANslation)
- BASIC (Beginner All-purpose Symbolic Instructional Code)
- Pascal (named for Blaise Pascal)
- Ada (named for Ada Lovelace)
- C (whose developer designed B first)
- Visual Basic (Basic-like visual language developed by Microsoft)
- Delphi (Pascal-like visual language developed by Borland)
- C++ (an object-oriented language, based on C)
- C# (a Java-like language developed by Microsoft)
- Java (We will use this one!!!)
To be continued...

- Tuesday: lecture
- Tuesday: first lab!!!