A View of Computer Science & Engineering

Dan Lopresti
Associate Professor
Office PL 404B
dal9@lehigh.edu
Computer Science as a Discipline

Old view:
- Computing is for a few specialized applications.
- People adapt work habits to limitations of computers.
- Computer science as “care and feeding” of computing systems.

New view:
- Computing as hidden enabler (“ubiquitous computing”).
- Computers adapt to people's needs and work habits.
- Computer scientists work on urgent problems affecting society.

Computer science is not just computer programming!
Computer Science at Lehigh

- Artificial intelligence
  - Case-based reasoning
  - Machine learning
  - Intelligent agents
- Bioinformatics
- Biometrics & security
- Computer architecture
- Database systems
  - Text & data mining
  - Transaction & query processing
- Digital libraries & document analysis
- Embedded systems
- Enterprise information systems
- Graphics
- Human-computer interaction
  - Virtual environments
- Image processing
- Internet
  - Semantic web
  - Search
  - Peer-to-peer systems
- Machine vision
- Networking & distributed systems
- Network security
- Parallel processing
- Robotics
- Software engineering
- Ubiquitous & mobile computing

All of these areas are represented in our faculty.
“Biology easily has 500 years of exciting problems to work on.”
Donald Knuth (Stanford Professor & famous computer scientist)

By developing techniques for analyzing sequence data and related structures, we can attempt to understand genetic nature of diseases.

http://cmgm.stanford.edu/biochem218/
Complete set of chromosomes that determines an organism is known as its *genome*.

### GenBank Release 121.0 — December 15, 2000

<table>
<thead>
<tr>
<th>Species</th>
<th>Haploid genome size</th>
<th>Bases</th>
<th>Entries</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Homo sapiens</em></td>
<td>3,400,000,000</td>
<td>6,702,881,570</td>
<td>3,918,724</td>
</tr>
<tr>
<td><em>Mus musculus</em></td>
<td>3,454,200,000</td>
<td>1,291,602,139</td>
<td>2,456,194</td>
</tr>
<tr>
<td><em>Drosophila melanogaster</em></td>
<td>180,000,000</td>
<td>487,561,384</td>
<td>166,554</td>
</tr>
<tr>
<td><em>Arabidopsis thaliana</em></td>
<td>100,000,000</td>
<td>242,674,129</td>
<td>181,388</td>
</tr>
<tr>
<td><em>Caenorhabditis elegans</em></td>
<td>100,000,000</td>
<td>203,544,197</td>
<td>114,553</td>
</tr>
<tr>
<td><em>Tetraodon nigroviridis</em></td>
<td>350,000,000</td>
<td>165,539,271</td>
<td>188,993</td>
</tr>
<tr>
<td><em>Oryza sativa</em></td>
<td>400,000,000</td>
<td>125,948,974</td>
<td>151,411</td>
</tr>
<tr>
<td><em>Rattus norvegicus</em></td>
<td>2,900,000,000</td>
<td>106,344,366</td>
<td>218,598</td>
</tr>
<tr>
<td><em>Bos taurus</em></td>
<td>3,651,500,000</td>
<td>71,215,626</td>
<td>159,473</td>
</tr>
<tr>
<td><em>Glycine max</em></td>
<td>1,115,000,000</td>
<td>62,817,102</td>
<td>141,802</td>
</tr>
<tr>
<td><em>Medicago truncatula</em></td>
<td>400,000,000</td>
<td>50,991,920</td>
<td>104,535</td>
</tr>
<tr>
<td><em>Trypanosoma brucei</em></td>
<td>35,000,000</td>
<td>49,855,996</td>
<td>91,334</td>
</tr>
<tr>
<td><em>Lycopersicon esculentum</em></td>
<td>655,000,000</td>
<td>49,415,566</td>
<td>97,112</td>
</tr>
<tr>
<td><em>Giardia intestinalis</em></td>
<td>12,000,000</td>
<td>47,639,714</td>
<td>54,328</td>
</tr>
<tr>
<td><em>Strongylocentrotus purpur</em></td>
<td>900,000,000</td>
<td>47,590,936</td>
<td>77,532</td>
</tr>
<tr>
<td><em>Entamoeba histolytica</em></td>
<td>—</td>
<td>44,522,016</td>
<td>49,938</td>
</tr>
<tr>
<td>* Hordeum vulgare*</td>
<td>—</td>
<td>44,489,692</td>
<td>57,779</td>
</tr>
<tr>
<td><em>Danio rerio</em></td>
<td>1,900,000,000</td>
<td>40,906,902</td>
<td>83,726</td>
</tr>
<tr>
<td><em>Zea mays</em></td>
<td>5,000,000,000</td>
<td>36,885,212</td>
<td>77,506</td>
</tr>
<tr>
<td><em>Saccharomyces cerevisiae</em></td>
<td>12,067,280</td>
<td>32,779,082</td>
<td>18,361</td>
</tr>
</tbody>
</table>

http://www.cbs.dtu.dk/databases/DOGS/
http://www.nsrl.ttu.edu/tmot1/mus_musc.htm
http://www.oardc.ohio-state.edu/seedid/single.asp?strID=324
Bioinformatics

Genomes are determined using a technique known as shotgun sequencing.

Computer scientists have played an important role in developing algorithms for assembling such data.

It's kind of like putting together a jigsaw puzzle with millions of pieces (a lot of which are “blue sky”).

http://occawlonline.pearsoned.com/bookbind/pubbooks/bc_mcampbell_genomics_1/medialib/method/shotgun.html
Bioinformatics

Mouse and Human Genetic Similarities


Courtesy Lisa Stubbs
Oak Ridge National Laboratory
New CSE Course in Bioinformatics

Introduced in Spring 2004. We study algorithms for:

- Sequence comparison & alignment (pairwise & multiple).
- Sequence assembly (shotgun sequencing).
- Physical mapping of DNA.
- Constructing phylogenetic (evolutionary) trees.
- Computing genome rearrangements.
- RNA and protein structure prediction.
- DNA microarray analysis.
- DNA computing.

Materials @ http://www.cse.lehigh.edu/~lopresti/courses.html
Data is becoming more portable (PDA's, cell phones, laptops, etc.) and theft is a growing concern.

Why aren't passwords enough?
- Very easy to “crack.”
- Thief can just disassemble and reverse-engineer device.

Two-pronged solution:
- Use biometrics in place of (or in addition to) passwords.
- Use secure data structure to encrypt information.
Using Biometrics to Protect Data

- Cryptographic key broken into shares and mixed with random data.
- Features extracted from user's speech or handwriting.
- Only input from true user will select correct shares to yield proper key.

Key shares

Feature extraction

Secure data structure

Impostor

True user

A View of Computer Science & Engineering • Lopresti
February 2005 • Slide 10

A328qv3r8...

98affnuqtr23...

Four score and...
Using Biometrics to Protect Data

Work with grad student Jarret Raim:
- Examine effectiveness.
- Quantify number of bits.
- Identify potential attacks.

Biometrics may be vulnerable:
- Study generative models.
- If successful, many current systems called into doubt.

Use our experience to improve biometrics, increase security.
Digital Libraries

The Historical New York Times Project:

George Washington Papers (Library of Congress):

Lehigh Digital Bridges:
Digital Libraries

Cornell and University of Michigan scanned over 900,000 pages documenting American social history from the antebellum period through reconstruction (1815 - 1926).

A search for the term “modem” which was first used in 1950's:

View the 10 matches in 9 books.
Ooops ...

... not “m o d e m” but “m o d e r n”!

These are OCR errors, not true hits. More work is needed ...
Protecting Online Services

The Internet has become a vehicle for distributing valuable content. But malicious programs ("bots") attempt to exploit online services intended for human users.

Idea: create a pattern recognition task that is easy for humans to solve, but hard for machines.
Protecting Online Services

Yahoo! method for protecting free email service. User must solve simple character recognition task:
Visual Tests

Currently, most such tests exploit gap in reading ability between humans and machines when confronted with degraded images of text.

Luckily, we recently hired Professor Henry Baird, an expert on optical character recognition and an originator of this research area.

Second International Workshop on Human Interactive Proofs will take place at Lehigh this Spring (co-chaired by Baird and Lopresti).
Spoken Language Tests

Bell Labs test comparing human vs. machine performance:

- Cell phone simulation (many other cases also studied).
- Humans nearly always much better than machine.
- Still open questions on how to use this.
Computer Science & Engineering

- Combines rich history with energetic new faculty.
- Builds on Lehigh tradition of excellence.
- Skills we teach can be applied across disciplines.
- Offers many opportunities for undergraduates, including:
  - Wide range of courses to take (good even for non-majors).
  - Chances to get involved with research projects.

These slides @ http://www.cse.lehigh.edu/~lopresti/talks.html
Study Break and “Show & Tell”

On Wednesday, March 23, CSE Department will host special Study Break and Show & Tell. All students invited, especially freshmen.

- Room – PL 466
- Study Break @ 4:00 pm – lots of good food (some even healthy).
- Show & Tell @ 4:30 pm – Professor John Spletzter will talk about his work in robotics.

Stop by and meet faculty and students, hear about courses, ask questions, etc.