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.

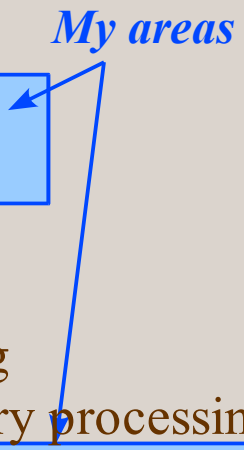
New view:

- 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

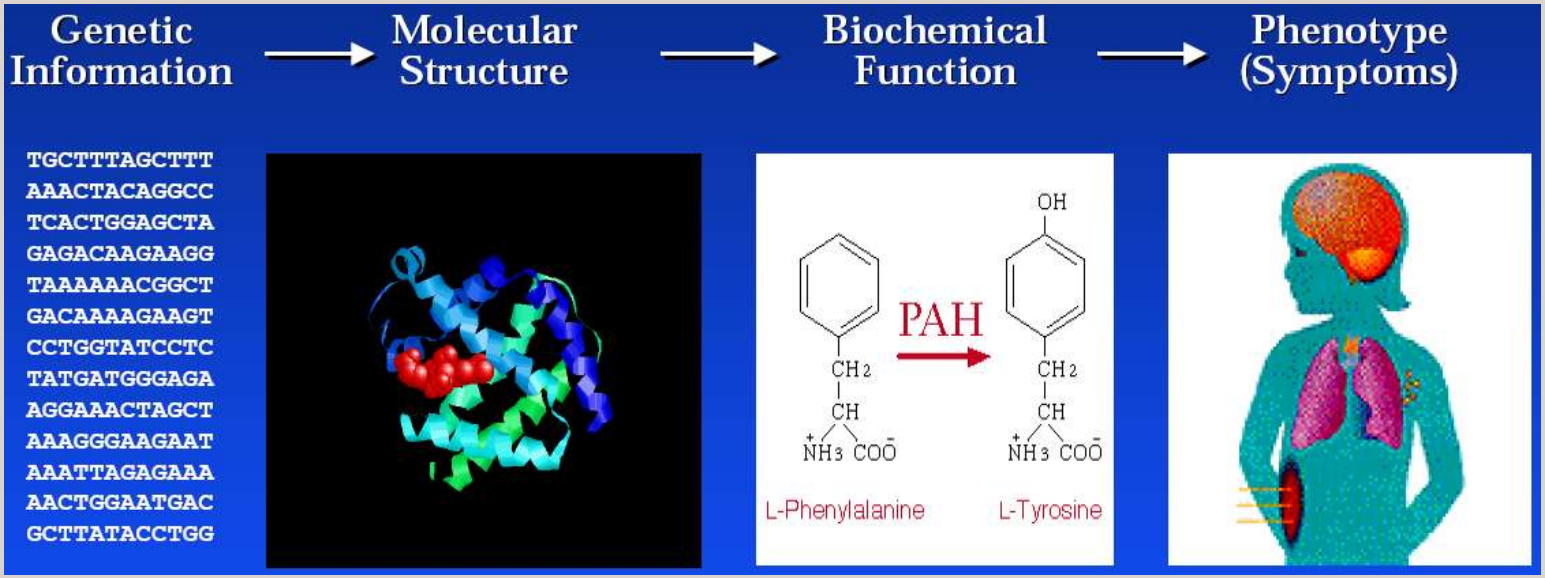
- Artificial intelligence
  - Case-based reasoning
  - Machine learning
  - Intelligent agents
- Bioinformatics
- Biometrics & security
- Computer architecture
- Database systems
  - Text & data mining
  - Transaction & query processing
- Digital libraries & document analysis
- Embedded systems
- Enterprise 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.

Bioinformatics

"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/

Bioinformatics

Complete set of chromosomes that determines an organism is known as its *genome*.

Mus musculus

GenBank Release 121.0 — December 15, 2000			
Species	Haploid genome size	Bases	Entries
Homo sapiens	3,400,000,000	6,702,881,570	3,918,724
Mus musculus	3,454,200,000	1,291,602,139	2,456,194
Drosophila melanogaster	180,000,000	487,561,384	166,554
Arabidopsis thaliana	100,000,000	242,674,129	181,388
Caenorhabditis elegans	100,000,000	203,544,197	114,553
Tetraodon nigroviridis	350,000,000	165,539,271	188,993
Oryza sativa	400,000,000	125,948,974	151,411
Rattus norvegicus	2,900,000,000	106,344,366	218,598
Bos taurus	3,651,500,000	71,215,626	159,473
Glycine max	1,115,000,000	62,817,102	141,802
Medicago truncatula	400,000,000	50,991,920	104,535
Trypanosoma brucei	35,000,000	49,855,996	91,334
Lycopersicon esculentum	655,000,000	49,415,566	97,112
Giardia intestinalis	12,000,000	47,639,714	54,328
Strongylocentrotus purpur	900,000,000	47,590,936	77,532
Entamoeba histolytica	—	44,522,016	49,938
Hordeum vulgare	—	44,489,692	57,779
Danio rerio	1,900,000,000	40,906,902	83,726
Zea mays	5,000,000,000	36,885,212	77,506
Saccharomyces cerevisiae	12,067,280	32,779,082	18,361

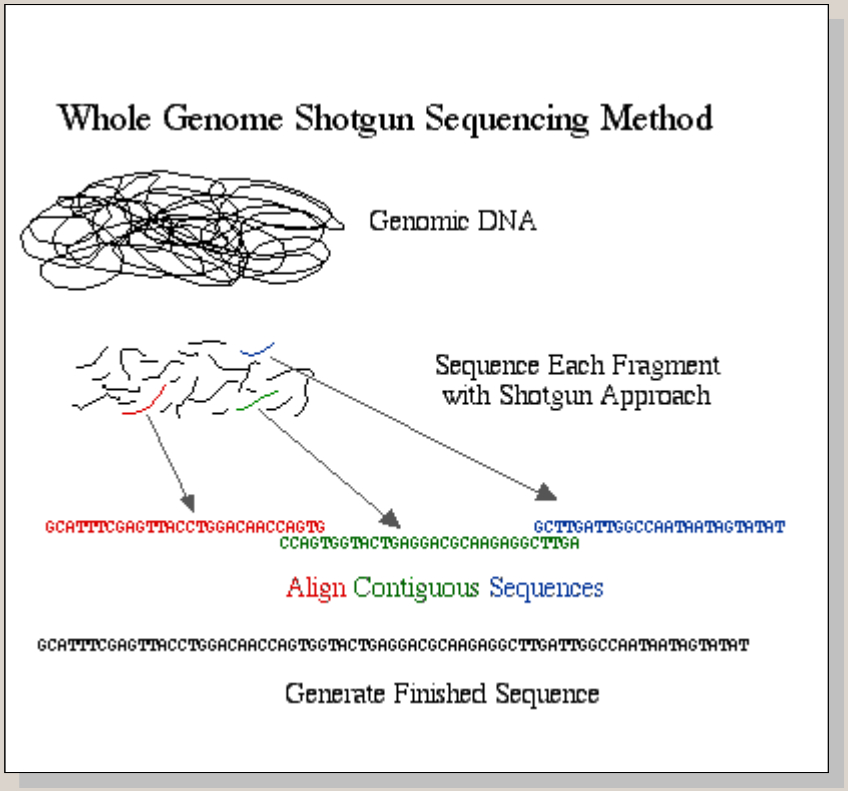
http://www.cbs.dtu.dk/databases/DOGS/
http://www.nsrl.ttu.edu/tmot1/mus_musc.htm
http://www.oardc.ohio-state.edu/seedid/single.asp?strID=324

Bioinformatics

Genomes are determined using a technique known as *shotgun sequencing*.

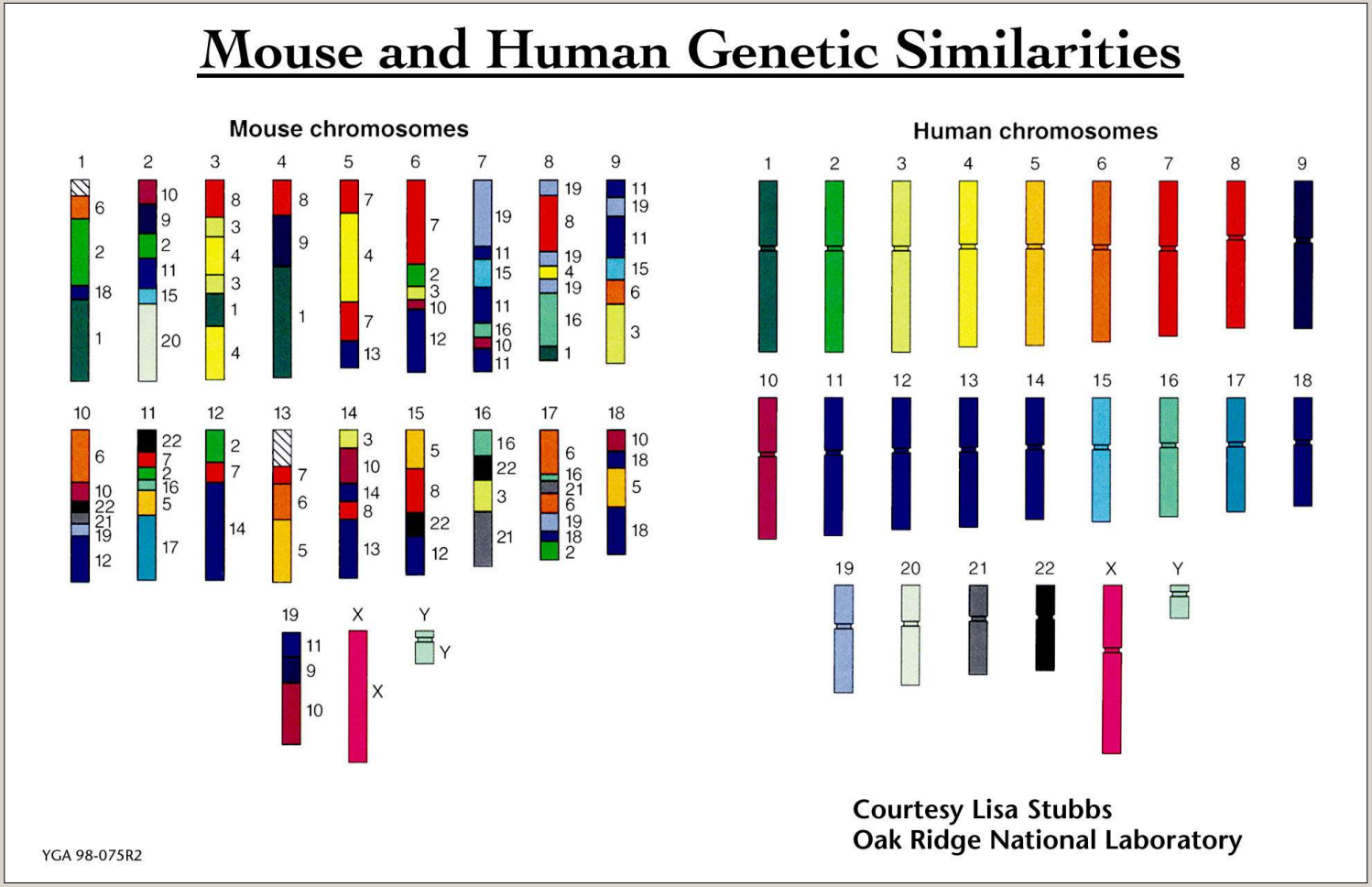
Computer scientists have played an important role in developing algorithms for assembling such data.

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



http://www.ornl.gov/sci/techresources/Human_Genome/graphics/slides/ttmousehuman.shtml

New CSE Course in Bioinformatics

Introduced in Spring 2004. We studied 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

Protecting Mobile Data

Data is becoming more portable (PDA's, cell phones, laptops, etc.) and theft is a growing concern.

Why aren't passwords enough?

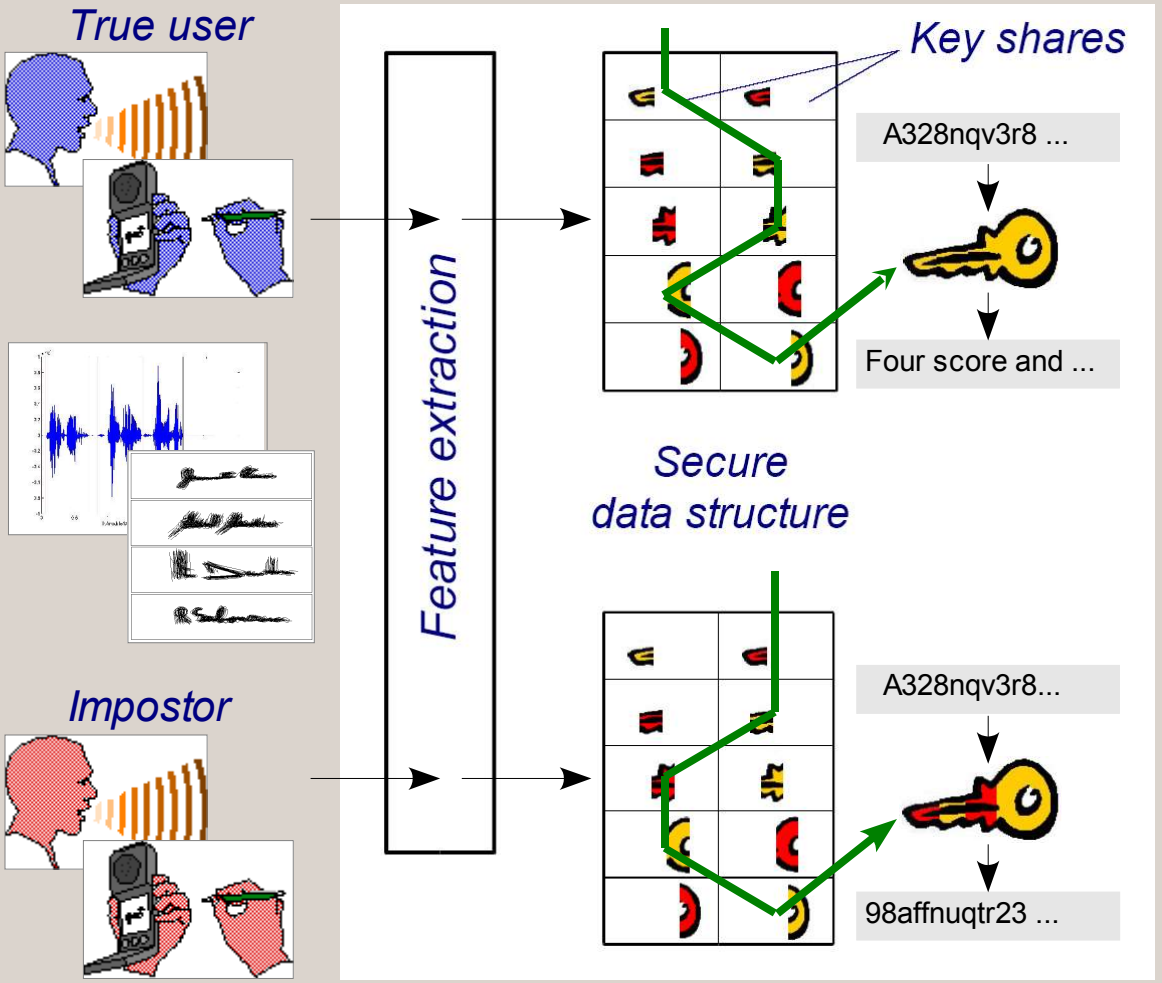
- Very easy to “crack.”
- Thief can just disassemble and reverse-engineer device.

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.



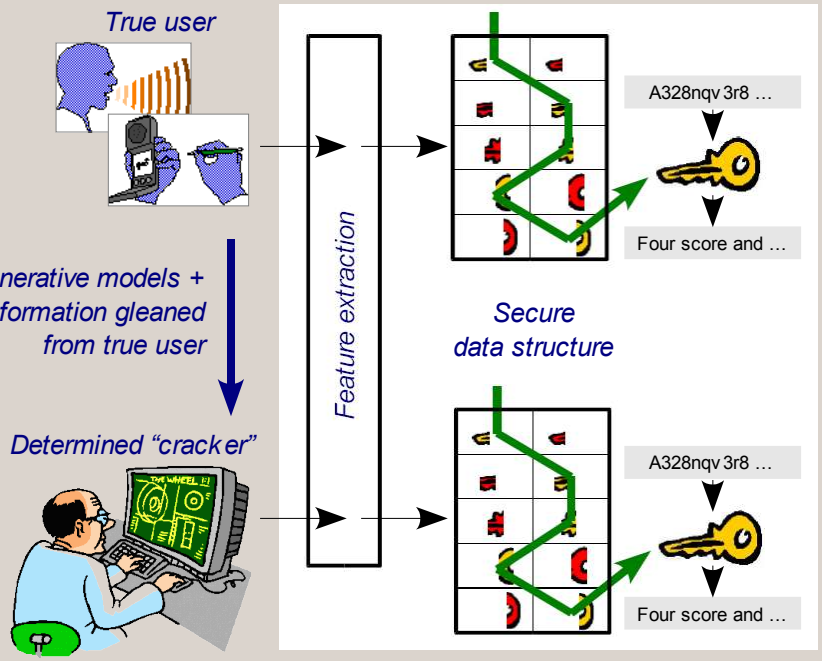
Using Biometrics to Protect Data

Work with grad student Jarret Raim:

- Examine effectiveness.
- Quantify number of bits.
- Identify potential attacks.

Biometrics may be vulnerable:

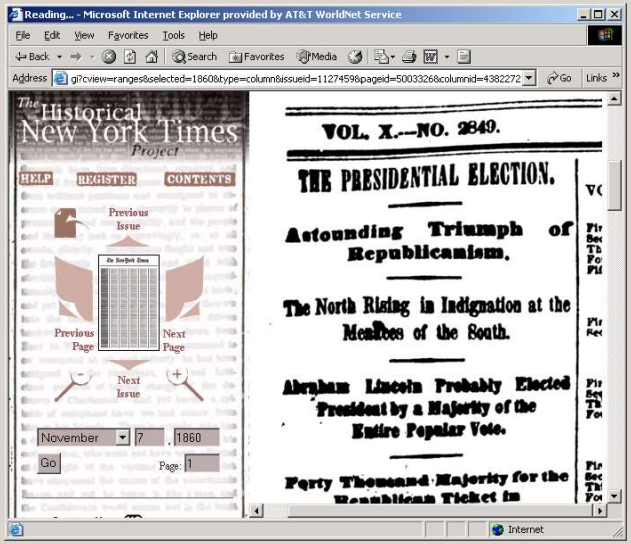
- Study generative models.
- If successful, many current systems called into doubt.



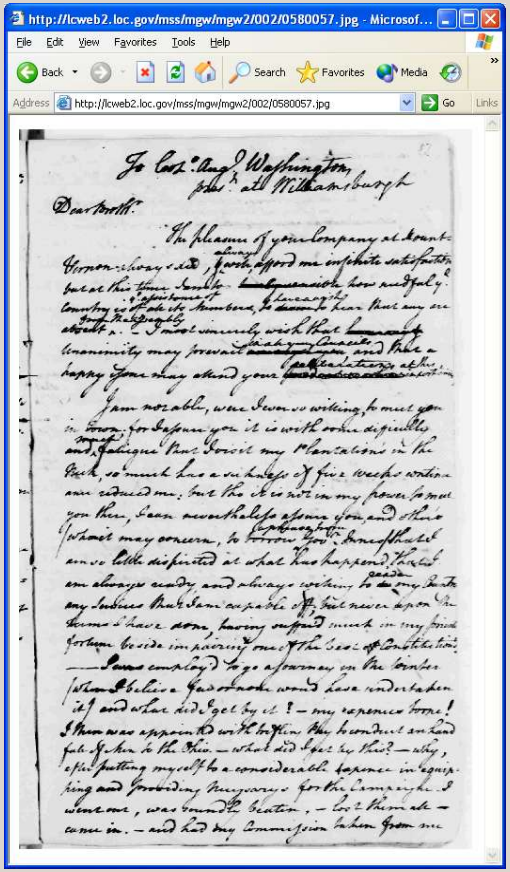
Use our experience to improve biometrics, increase security.

Digital Libraries

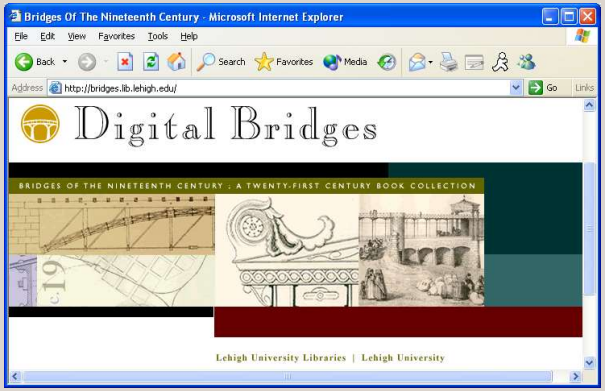
The Historical New York Times Project:



George Washington Papers (Library of Congress):



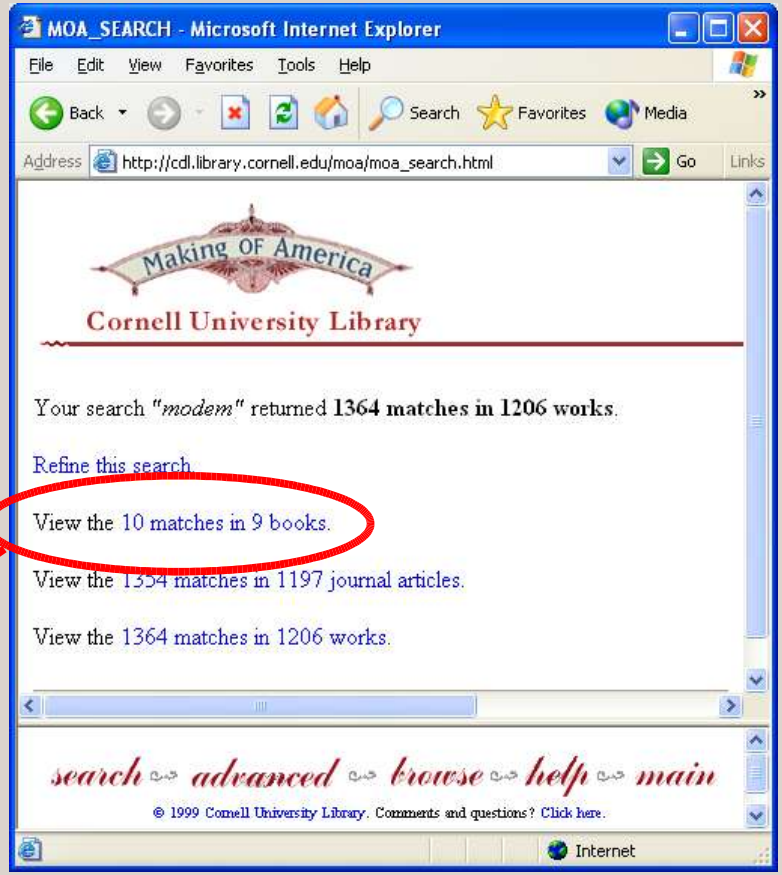
Lehigh Digital Bridges:



Digital Libraries

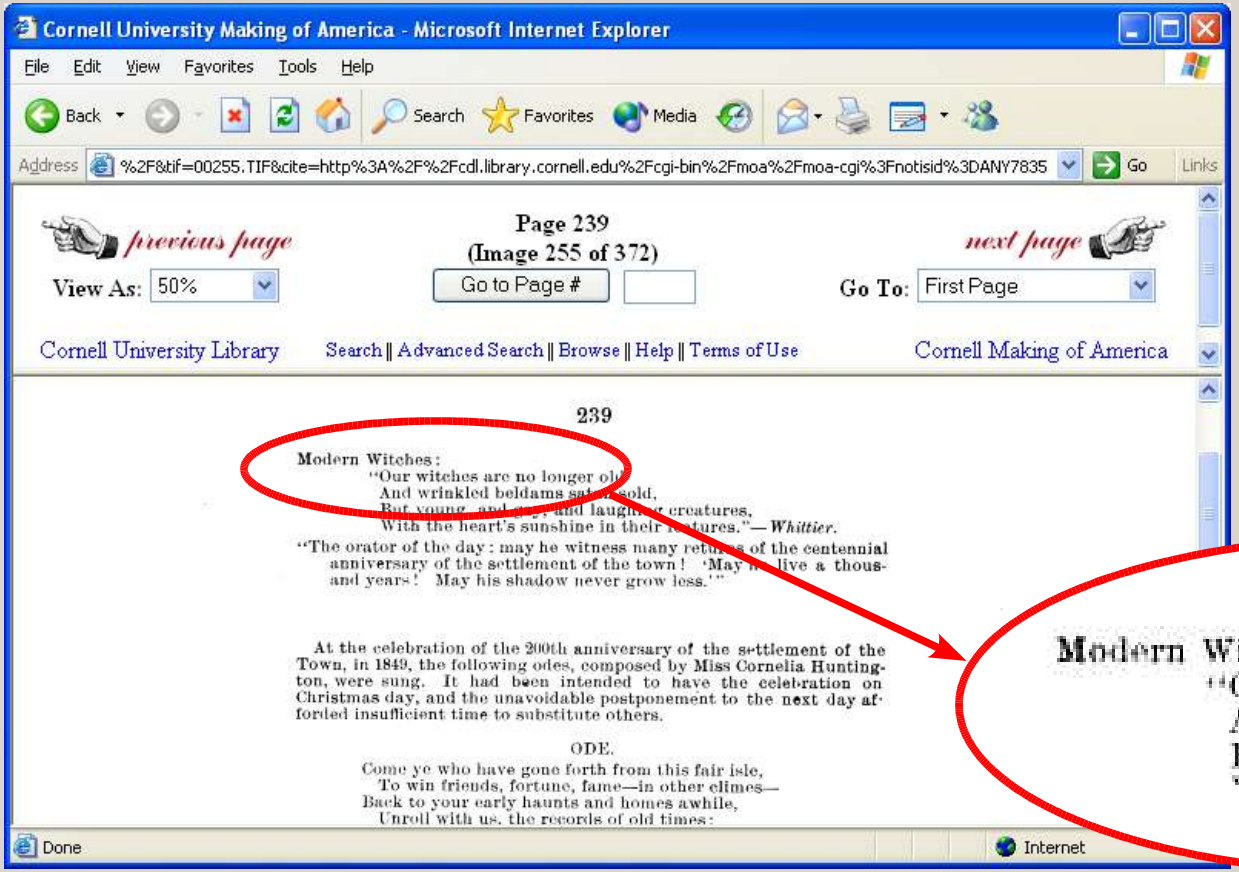
Cornell and University of Michigan scanned over 900,000 pages documenting American social history from the antebellum period through reconstruction (1815 - 1926).

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



View the 10 matches in 9 books.

Ooops ...



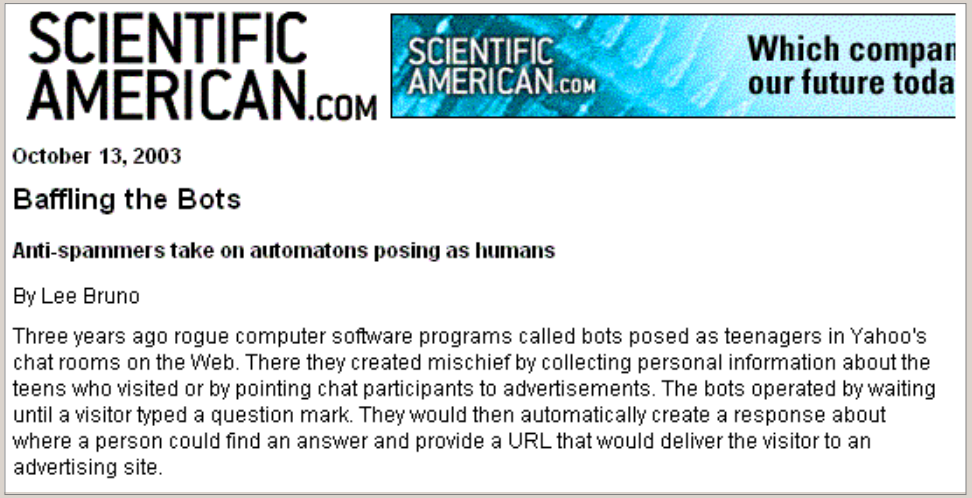
... not “m o d e m”
but “m o d e r n”!

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

Protecting Online Services

The Internet has become 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.



SCIENTIFIC AMERICAN.COM

October 13, 2003

Baffling the Bots

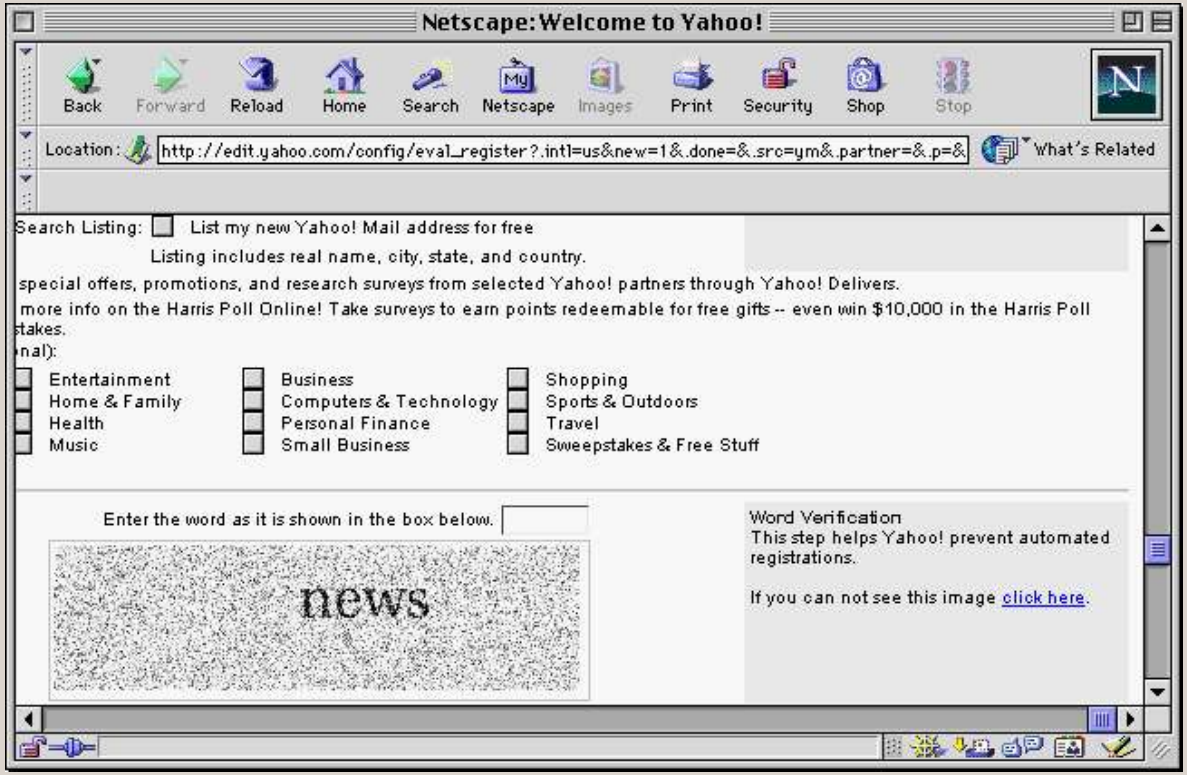
Anti-spammers take on automatons posing as humans

By Lee Bruno

Three years ago rogue computer software programs called bots posed as teenagers in Yahoo's chat rooms on the Web. There they created mischief by collecting personal information about the teens who visited or by pointing chat participants to advertisements. The bots operated by waiting until a visitor typed a question mark. They would then automatically create a response about where a person could find an answer and provide a URL that would deliver the visitor to an advertising site.

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.

EZ-Gimpy

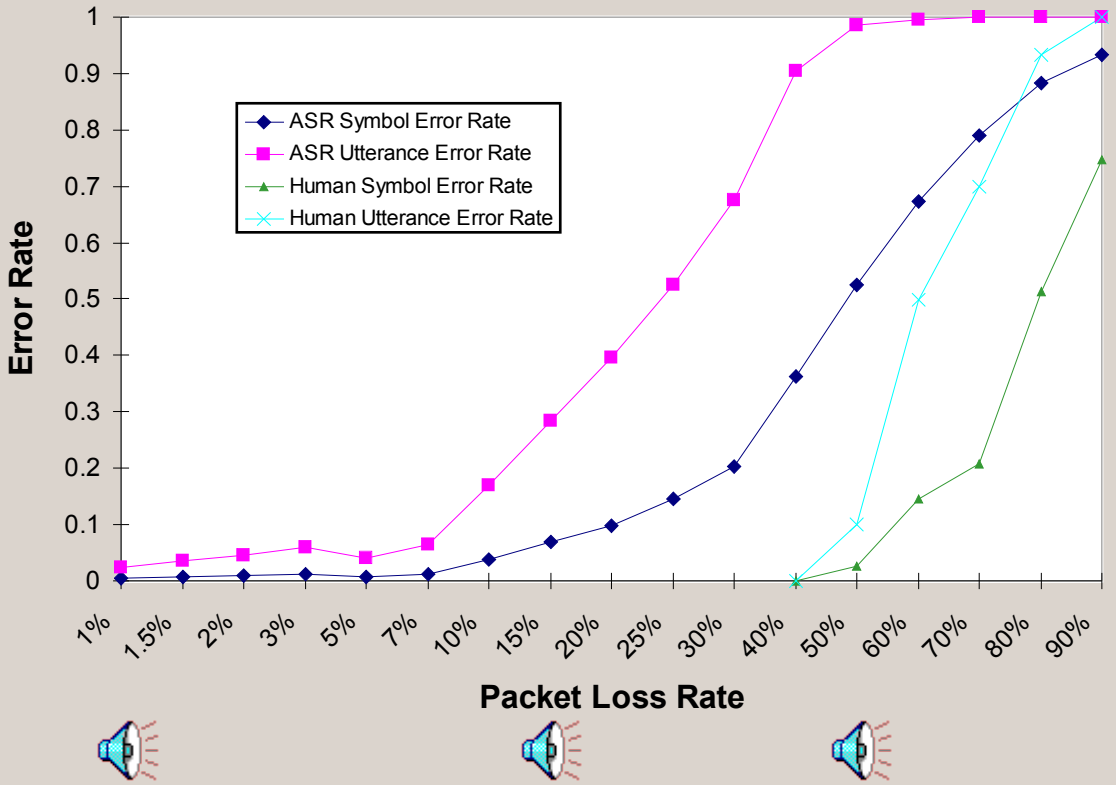
PessimalPrint

BaffleText

Second International Workshop on Human Interactive Proofs will take place at Lehigh next 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