

Computer Science & Engineering @ Lehigh

February 2007



Dan Lopresti
Associate Professor
Office PL 404B
dal9@lehigh.edu

Computer Science & Engineering

Pervades business and modern society:

- Communication, entertainment, manufacturing, information processing, finance, biotech, cybersecurity, national defense, etc.
- Advances in Computer Science & Engineering have enormous impact on fundamental problems in engineering, the sciences, business, public policy, the arts, etc.

**Computer Science & Engineering is at
the core of the Information Age!**

Fields CSE Grads Work In

- Design

- testing
- manufacturing

- Laser-guided

- surgery
- weapons

- Medicine

- genomics
- laparoscopic
- MRI

- Transportation

- airplanes (fly by wire)
- automobiles

- Computer controls

- devices & processes
- nanotechnology

- Simulation & training

- astronauts
- soldiers



- Entertainment

- media, devices
- synthesis
- delivery network
- interactive creation

- Communication

- telecom, Internet, etc.
- devices, methodology
- distribution

- Publishing

- tools
- media, distribution

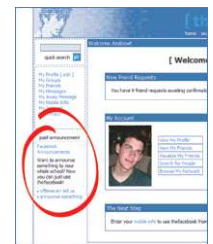
- Knowledge industry

- Social networking and services

- Financial

- Business strategy

- All forms of engineering ...



Hiring of CSE Majors

They're Hiring in Techland

By Spencer Ante

BusinessWeek online

January 2006

The tech job market has sprung back to life, and this year could be the best one since 2000

In the summer of 2002, amid the depths of the dot-com bust and the rise of global outsourcing, William Davis did something that seemed kind of crazy: He became a graduate **student in computer science** ... Last summer, Davis' investment paid off

... "As the memory of the tech bust fades, we seem to be getting better and better job growth," says Zandi.

...**MORE TECHIES**. Corporate America is spending money on technology again, so the hiring is relatively broad-based.

...Google (GOOG), Microsoft (MSFT), Accenture (ACN), Amazon (AMZN), Advanced Micro Devices (AMD), Altera (ALTR), Infosys (INFY), and Citrix Systems (CTXS) **all say they plan to hire more techies in the U.S. in 2006**. Many successful startups, such as OfficeTiger and NetSuite, continue to expand their workforces as well.

Starting Salaries



January 2006

1. Computer systems software engineer -- \$81,140*

- Computer systems software engineers work to coordinate a company's computer needs and maintain its computer systems. They may also set up a company's intranets to ease communication between the various departments. Most jobs require a bachelor's degree in computer science or computer information systems.

2. Computer applications software engineer -- \$76,310

- Computer applications software engineers use programming languages such as C++ and Java to design, construct and maintain general computer applications software. Most jobs require at least a bachelor's degree, but some more complex jobs require a graduate degree.

3. Biomedical engineer -- \$70,520

4. Physician assistant -- \$69,250

5. Environmental engineer -- \$67,620

6. Computer systems analyst -- \$67,520

- Computer systems analysts help an organization get the most for their technology investment dollars by solving computer problems and planning and developing new computer systems. Educational requirements vary by the employer and job complexity, ranging from a two-year degree to a graduate degree, and may include continuing education and certification.

7. Database administrator -- \$61,950

- Database administrators ensure system performance by setting up computer databases, testing and coordinating modifications to computer systems, identifying user requirements and adding new users to the system. Employers prefer candidates with technical degrees, but the specific level of education and type of training required depends on the complexity of the job and employers' needs.

8. Physical therapist -- \$61,560

9. Network systems and data communication analyst -- \$61,250

- Network systems and data communication analysts are responsible for keeping electronic communications like Internet, voice mail and e-mail up and running. They spend much of their days testing and evaluating systems including local area networks (LANs), wide area networks (WANs) and intranets.
- Depending on employer and complexity of the job, educational requirements range from an associate's degree to a computer-related bachelor's degree.

10. Hydrologist -- \$60,880

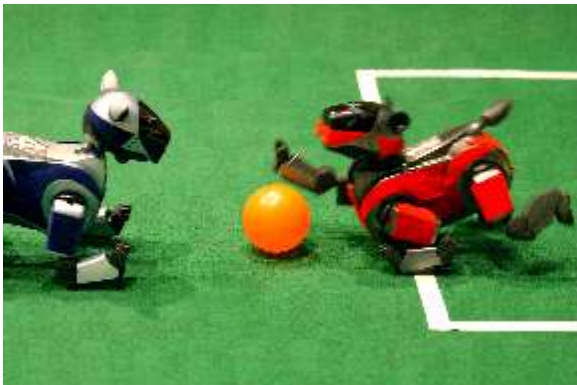
Source: <http://www.cnn.com/2006/US/Careers/01/26/cb.top.jobs.pay/index.html>

5 of top 10 slots are CSE-related.

Computer Science & Engineering

Computer Science & Engineering is much more than programming:

- How can you make a computer think (e.g., recognize words in speech or images, plan, move...)?
- How do you make believable opponents in video games?
- How do you ensure privacy and security for online transactions?



**In Computer Science & Engineering
you think creatively and
solve problems innovatively!**

Computer Scientists & Engineers

innovators

engineers

architects

inventors

thinkers

problem solvers

Integrators

the most successful
programmers game creators

leaders

team members

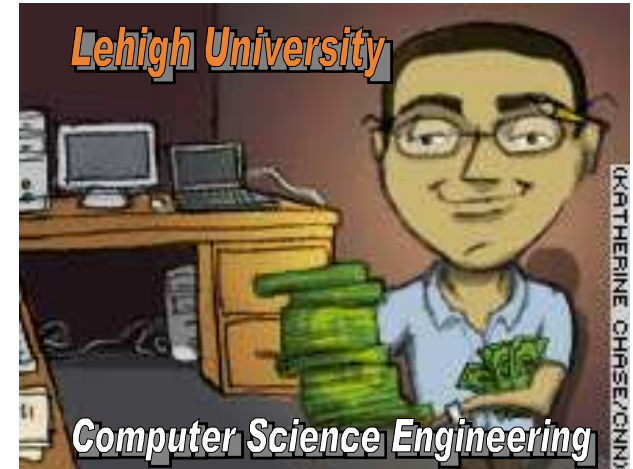
animators

entrepreneurs

designers

company builders

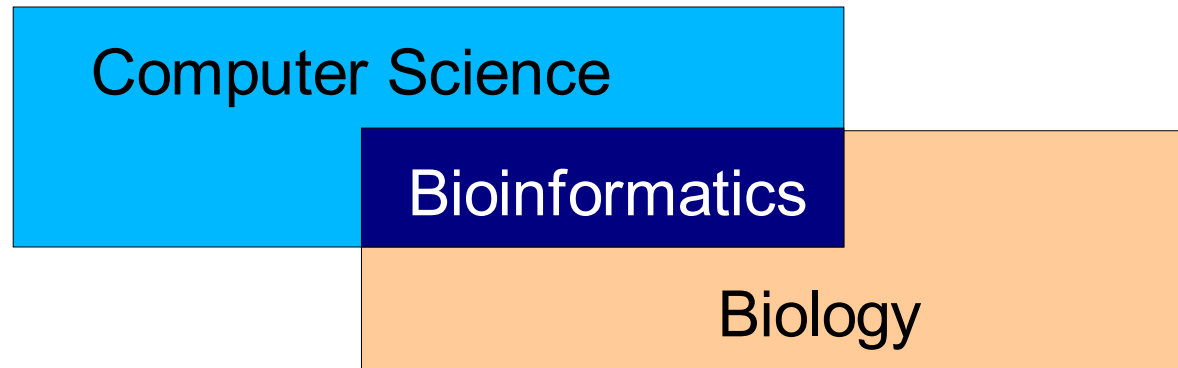
Collaborators



**For some, it's the challenge;
for others, it's a sense of fulfillment**

Bioinformatics

What is bioinformatics? Application of techniques from computer science to problems from biology.

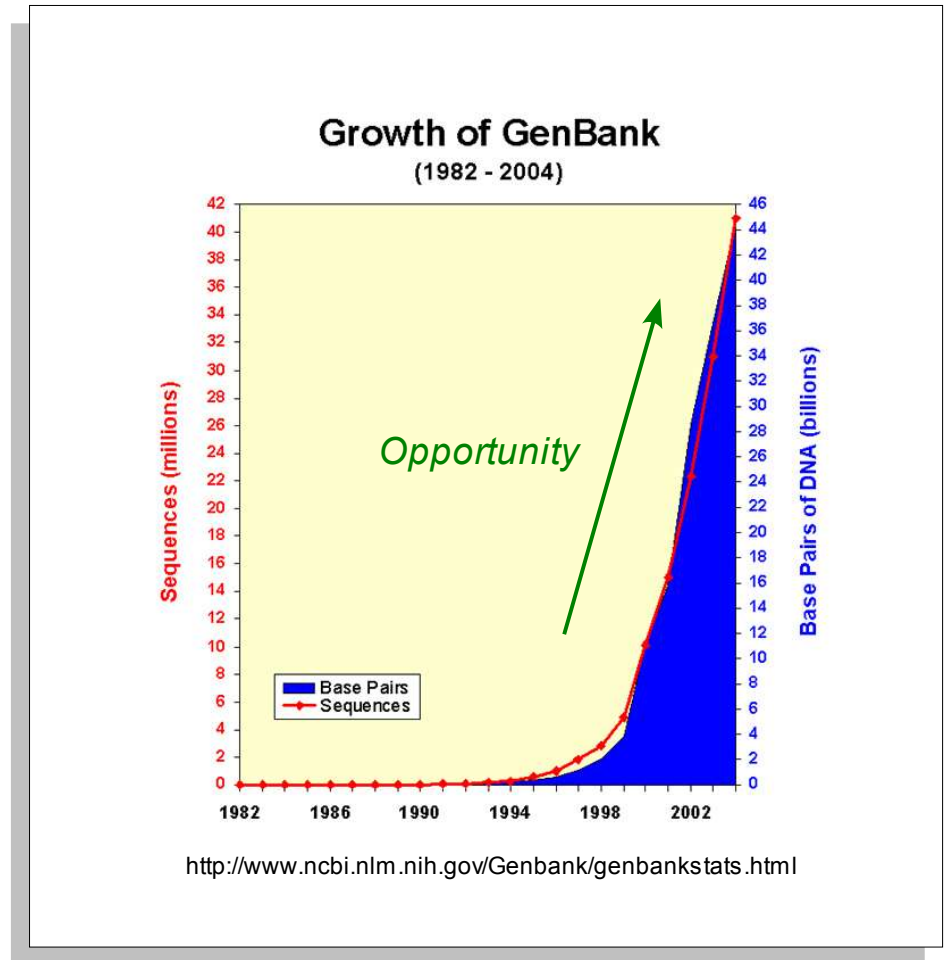


Why is it interesting?

- Important problems.
- Massive quantities of data.
- Desperate need for efficient solutions.
- Success is rewarded.

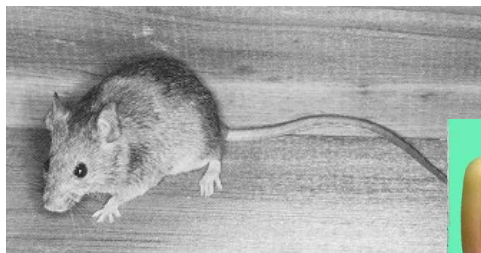
Motivation

- Genetic identity of most organisms is encoded in long molecules made up of four basic units: the nucleic acids *adenine*, *cytosine*, *guanine*, and *thymine*.
- To first approximation, DNA is language over 4 character alphabet, {A, C, G, T}.
- Complete set of chromosomes that determines an organism is known as its *genome*.



Bioinformatics

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



Mus musculus



Us →

←

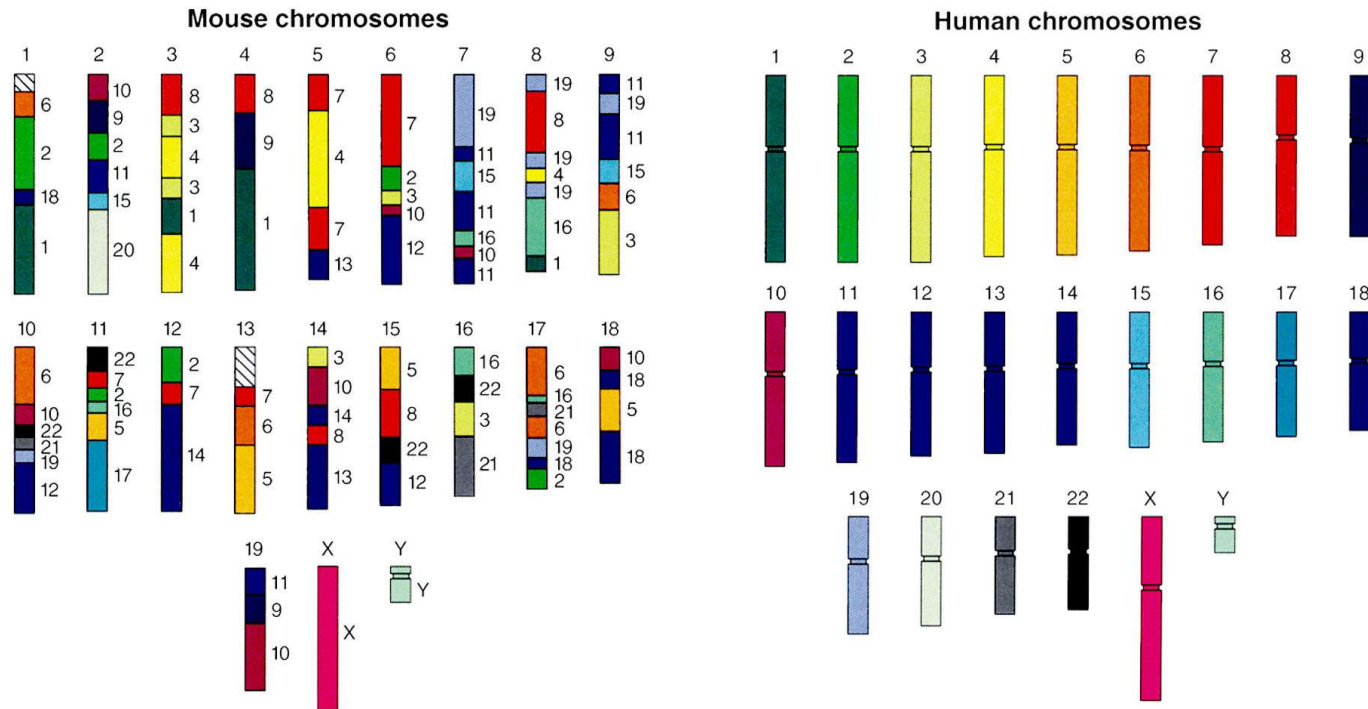
GenBank Release 121.0 — December 15, 2000

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.nslr.ttu.edu/tmot1/mus_musc.htm
<http://www.oardc.ohio-state.edu/seedid/single.asp?strID=324>

Comparative Genomics

Mouse and Human Genetic Similarities



Courtesy Lisa Stubbs
Oak Ridge National Laboratory

YGA 98-075R2

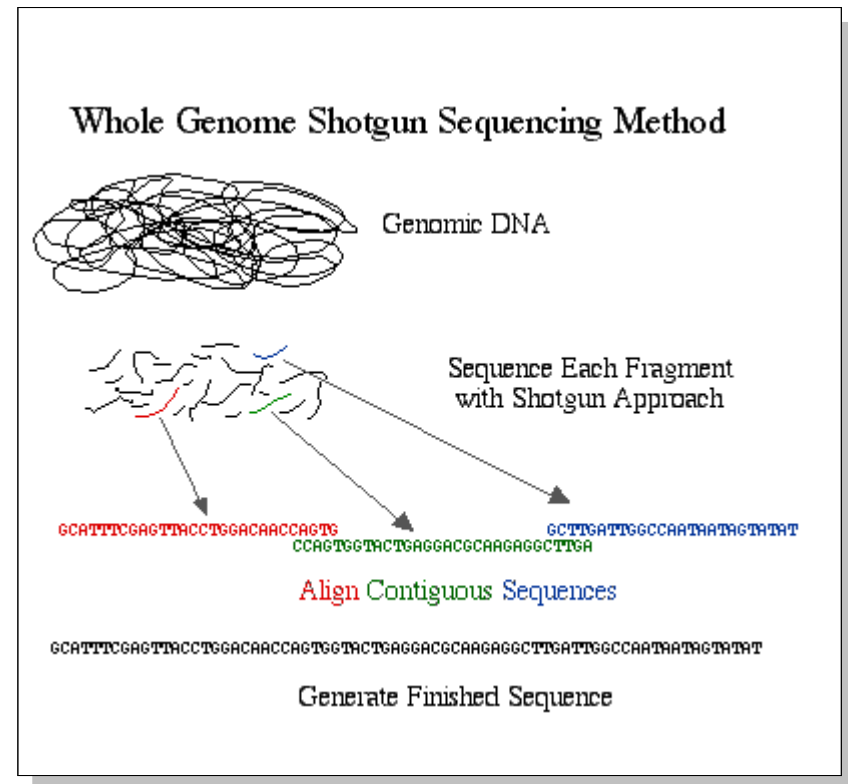
http://www.ornl.gov/sci/techresources/Human_Genome/graphics/slides/ttmousehuman.shtml

Sequencing a Genome

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”).

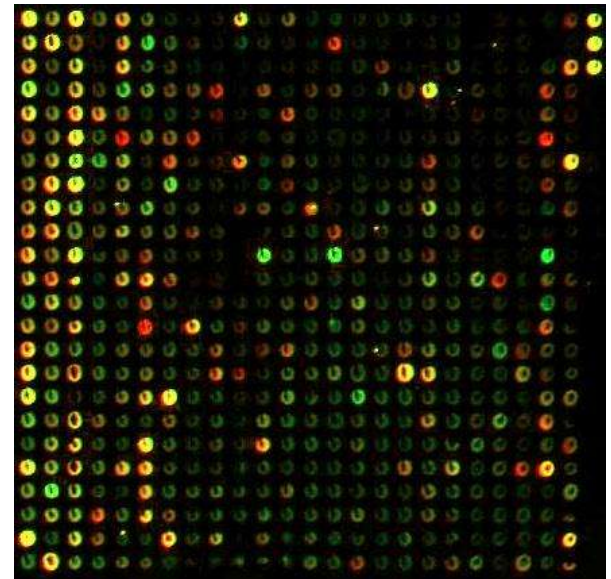


http://ocawlonline.pearsoned.com/bookbind/pubbooks/bc_mcampbell_genomics_1/medialib/method/shotgun.html

DNA Microarrays

- Allows simultaneous measurement of the level of transcription for every gene in a genome (gene expression).
- Differential expression, changes over time.
- Single microarray can test ~10k genes.
- Data obtained faster than can be processed.
- Want to find genes that behave similarly.
- Must find ways to uncover patterns.

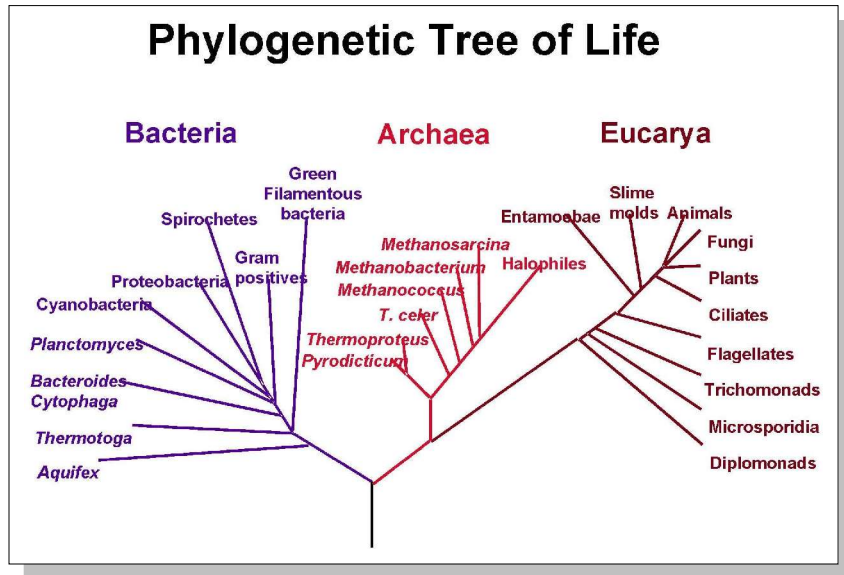
green = repressed
red = induced



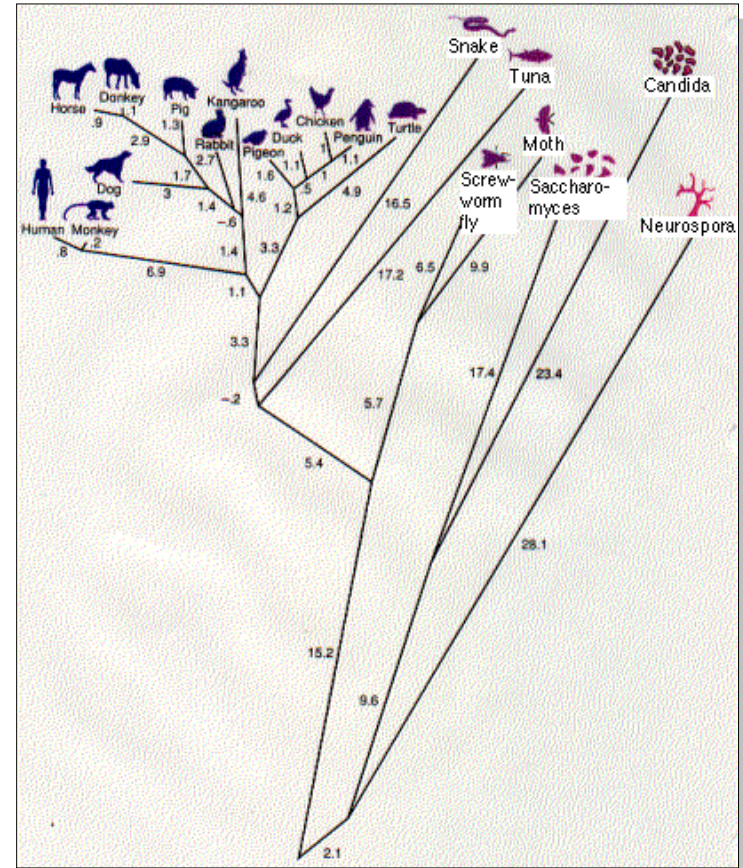
**Analysis of this data
is a CSE problem.**

The “Tree of Life”

Scientists build phylogenetic trees in an attempt to understand evolutionary relationships.



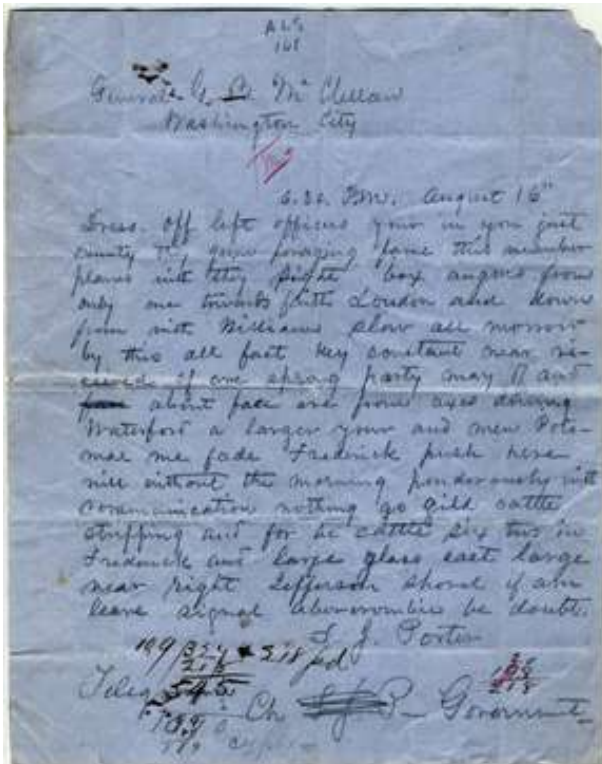
Note: these trees are “best guesses” and certainly contain some errors!



http://en.wikipedia.org/wiki/Phylogenetic_tree
<http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/T/Taxonomy.html>

Breaking a Civil War Secret Code

The Civil War letter ...



<http://digital.lib.lehigh.edu/remain>

... encrypted – not yet broken.

The players ...



**Major General
Fitz-John Porter**
author of letter, blamed for Union
loss at Second Bull Run, court-
martialed in 1863



**Major General
George McClellan**
recipient of letter



Anson Stager
inventor of cypher system and
later an early leader in U.S.
telecommunications

Breaking a Civil War Secret Code

The news story ...

"Lehigh team works to crack Civil War code"

Lehigh team works to crack Civil War code

A group of Lehigh students, professors and staff members recently worked to de-code an old letter between two Civil War generals.

A team of Lehigh librarians, professors and students spent the past semester engaged in unraveling a Civil War-era mystery by trying to crack an old code buried in a letter from a Union Army leader about the Battle of Bull Run.

The letter, which was the property of Lehigh's Special Collections, was written by Major General Fitz John Porter to General G.B. McClellan in Washington, D.C. and may possibly have contained Porter's explanation of the actions he took - or failed to take - on the brink of the battle.

Monica Najar, assistant professor of history, joined forces with Dan Lopresti, associate professor of computer science and engineering, to conduct independent studies with their students. Najar worked directly with Adam Ressler, a junior in Lehigh's College of Business and Economics, to approach the project from a historical perspective and develop a rich spatial and social/historical context for the letter to identify key terms and locations.

Lopresti worked with Ben Wu '05 in employing computer programs to test multiple word-order possibilities.

The letter from Major General Fitz John Porter to General G.B. McClellan in Washington, D.C., is part of Lehigh's Special Collections.

... ending yet to be written.

The current software tool ...

off and down received from county from
left of and near augurs they with
officers may Loudon constant box gum Williams
your party faith key sight foraging slow
in spring towards fact they lame all
you are one all with this morrow
just of only this planes number by

Enumerate Routings

Columns: 7 Lines: 7

Skip Text Strings: all all of of this this

Require Text Strings: received from this morrow with officers

Seed: 5039,127

Routing: D1 D6 U5 U3 D7 U4 U2

Total Routings: 645120

Generated Routings: 645120 (100.0000%)

Qualified Routings: 96 (0.015% total) (0.015% so far)

Revert Next Stop OK

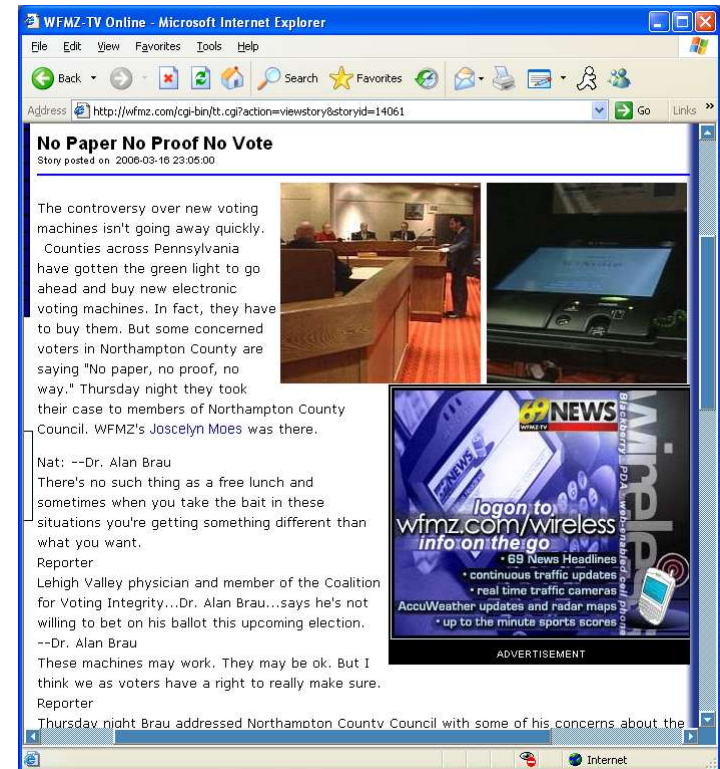
Record Auto Count Auto Save

... promising, but needs work.

Hint: I'm looking for help ...

Better Electronic Voting

E-voting has generated enormous controversy recently ...



Maryland votes “yes” for paper trail ... while Pennsylvania votes “no.”

Better Electronic Voting

E-voting: what's the right answer?

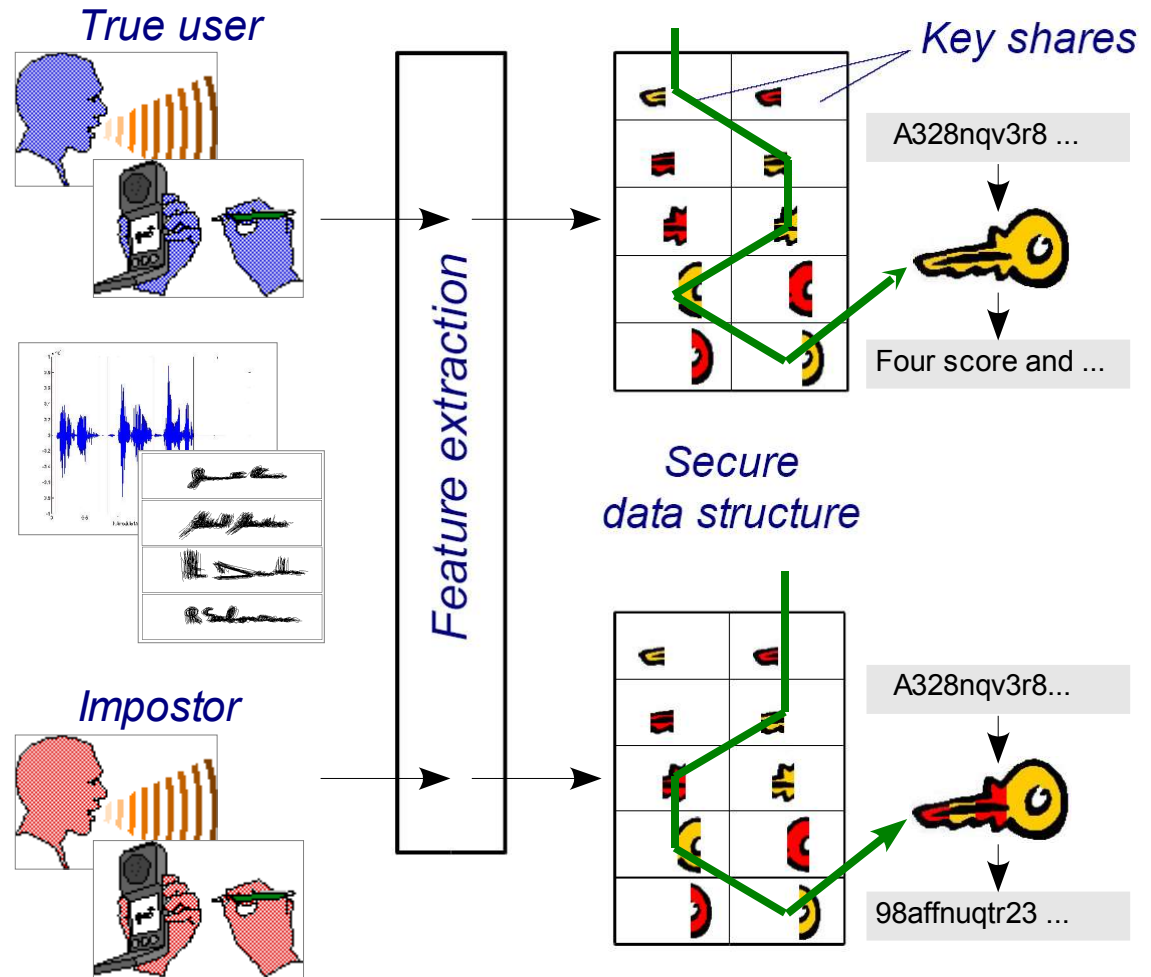
- Take a critical look at all aspects of the problem.
- Examine both security and usability issues.
- Build a prototype of an e-voting system that includes a reliable Voter Verified Paper Audit Trail (VVPAT).
- Some say it's impossible: we disagree.
- Of fundamental importance because our democracy depends on fair and transparent elections.



Diebold e-voting system

Evaluating Biometric Security

- 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.



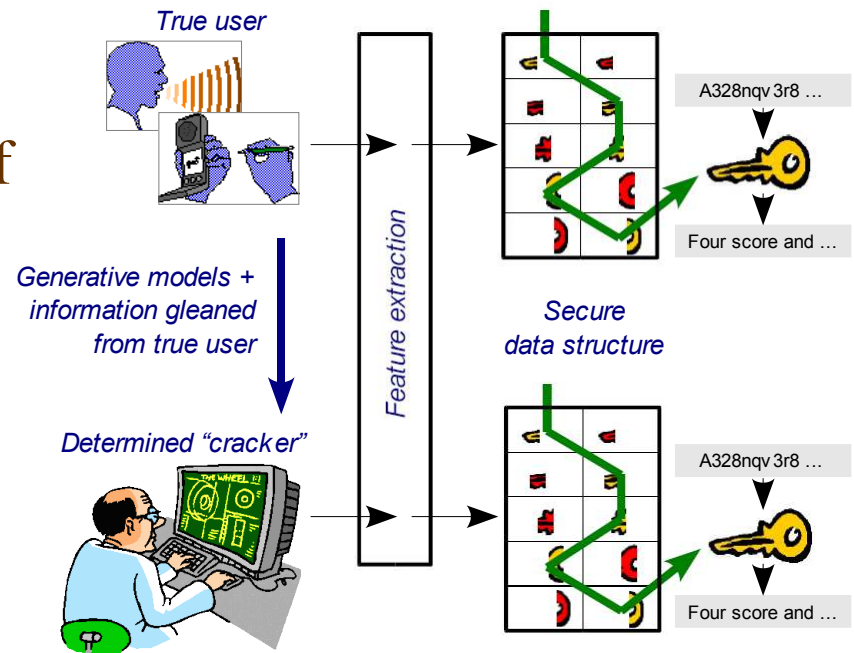
Evaluating Biometric Security

Biometrics may be vulnerable to attacks using generative models.

- Some current systems at risk.
- Results for handwriting show machine can equal performance of skilled human forger.

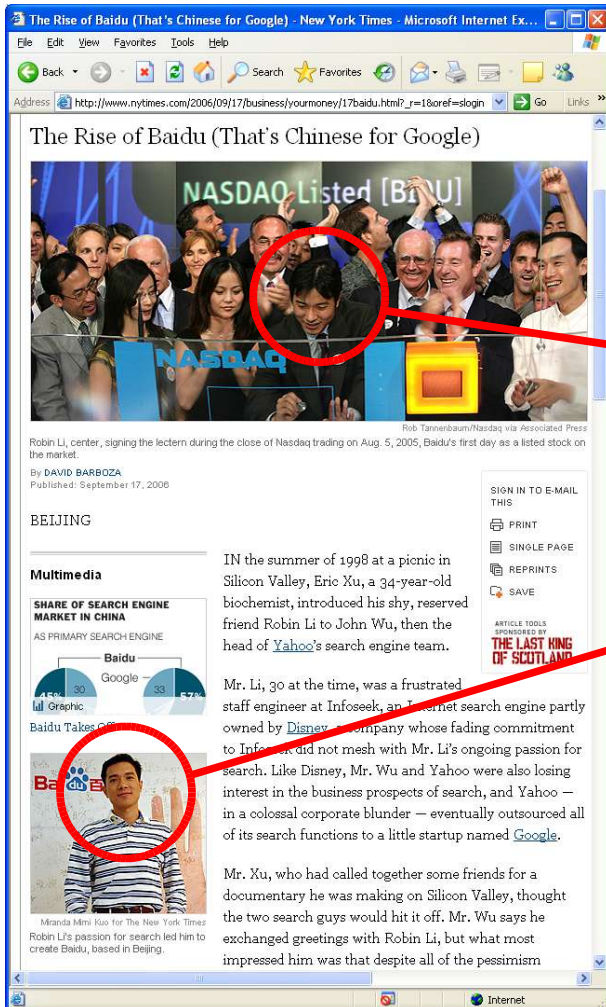


Same idea used in Mission Impossible 3.



Use our experience to improve biometrics, increase security.

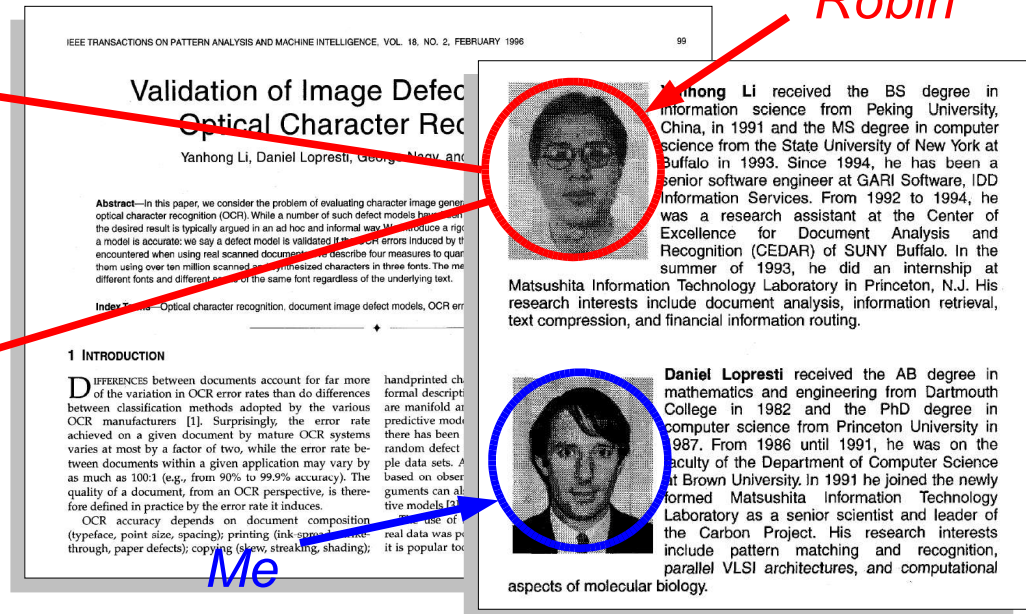
New Field = Golden Opportunities



Robin (Yanhong) Li was a student intern working with me in the mid-1990's at a research lab in Princeton.

He's now the billionaire founder of Baidu (the Chinese version of Google).

Robin



No guarantee you'll become a billionaire if you major in CSE. But in what other fields is this even possible?

From the New York Times, Sunday Sept. 17, 2006.

Alternatives Abound

Computer Science & Engineering at Lehigh includes:

- B.A. in Computer Science in CAS.
- B.S. in Computer Science in CAS.
- B.S. in Computer Engineering in RCEAS (along with ECE Dept).
- B.S. in Computer Science in RCEAS.
- B.S. in Computer Science & Business (dual degrees, with CBE).
- Minor in Computer Science (available to all).

Suggested “Tracks” #1

● Artificial Intelligence

- CSE 327 Artificial Intelligence Theory and Practice
- CSE 326 Pattern Recognition
- CSE 335 Topics in Intelligent Decision Support Systems
- CSE 347 Data Mining
- CSE 348 AI Game Programming
- CSE 360 Introduction to Mobile Robotics
- CSE 368 AI Programming

● Bioinformatics

- CSE 308 Bioinformatics
- CSE 241 Database Systems
- CSE 326 Pattern Recognition
- CSE 347 Data Mining
- BIOS 41 Biology Core I: Cellular and Molecular
- BIOS 115 Biology Core II: Genetics

● Computing Principles

- CSE 241 Database Systems
- CSE 302 Compiler Design
- CSE 327 Artificial Intelligence Theory and Practice
- CSE 375 Topics in Parallel Computing
- CSE 376 Parallel Algorithms

● Hardware-Software

- ECE 81 Principles of Electrical Engineering
- CSE 209 Assembly Language Programming
- CSE 271 Programming in C and Unix
- ECE 319 Digital System Design
- CSE 336 Embedded Systems
- CSE 363 Network Systems
- CSE 375 Topics in Parallel Computing

<http://www3.lehigh.edu/engineering/cse/academics/undergrad/TRACKS.asp>



Suggested “Tracks” #2

● Information Management

- CSE 241 Database Systems
- CSE 335 Topics in Intelligent Decision Support Systems
- CSE 345 WWW Search Engines
- CSE 347 Data Mining
- CSE 352 Information Technology for Commerce

● Interactive Multimedia Systems

- CSE 313 Computer Graphics
- CSE 197 Computer Game Design
- CSE 331 User Interface Systems and Techniques
- CSE 332 Multimedia Design and Development
- CSE 348 AI Game Programming

● Systems and Networks

- CSE 271 Programming in C and Unix
- CSE 265 System and Network Administration
- CSE 336 Embedded Ssystems
- CSE 342 Fundamentals of Internetworking
- CSE 343 Network Security
- CSE 345 WWW Search Engines
- CSE 363 Network Systems

● Software Systems

- CSE 271 Programming in C and Unix
- CSE 366 Object-Oriented Programming
- CSE 209 Assembly Language Programming
- CSE 302 Compiler Design
- CSE 332 Multimedia Design and Development
- CSE 376 Parallel Algorithms

<http://www3.lehigh.edu/engineering/cse/academics/undergrad/TRACKS.asp>



Last but not least ...

Why attend Lehigh?

Because this is a great place to be a student!

Thank you!