What does this position entail?

- Do you want to build quantitative models millions of people will use, based on data from the world's largest online laboratory? Are you passionate about formulating relevant questions and producing solutions to initially ill-defined problems? Do the challenges and opportunities of terabytes of data excite you? Can you think abstractly and apply your ideas to the real world? Can you contribute to the big picture and are not afraid to handle the details?
- We are looking for people with the right blend of vision, intellectual curiosity, and hands-on skills, who want to be part of a highly visible, entrepreneurial team

http://www.ph.tn.tudelft.nl/PRInfo/jobs/msg00185.html

What is this about?

- Ideal candidates will have a track record of creating innovative solutions, and typically a Ph.D. in computer science, physics, statistics, or electrical engineering. Significant research experience is desired in fields including active learning, probabilistic graphical models and Bayesian networks, data mining and visualization, Web search and information retrieval, judgment and decision making, consumer modeling, and behavioral economics.
- What is data mining? What is machine learning?

My recommendations at Amazon

1. Mozart's Magnificent Voyage: Tales Of The Dream Children
~ Various Artists (Composer), et al
Average Customer Review:
Release Date: October 13, 1998
Our Price: $11.99 Used & new from $10.08
Add to cart | Add to Wish List
See related items
Why was I recommended this?
Rate this item | I own it | Not interested

2. Mozart's Magic Fantasy: A Journey Through 'The Magic Flute'
~ Classical Kids
Average Customer Review:
Release Date: April 11, 1995
Our Price: $11.99 Used & new from $5.00
Add to cart | Add to Wish List
See related items
Why was I recommended this?
Rate this item | I own it | Not interested

3. Emergence: The Connected Lives of Ants, Brains, Cities, and Software
by Steven Johnson
Average Customer Review:
Publication Date: September 10, 2002
Our Price: $11.20 Used & new from $4.95
Add to cart | Add to Wish List
See related items
Why was I recommended this?

And again…

8. Structural Bioinformatics (Methods of Biochemical Analysis, V. 44)
by Philip E. Bourne (Editor), Helge Weissig (Editor)
Average Customer Review:
Publication Date: February 7, 2003
Our Price: $71.30 Used & new from $50.00
Add to cart | Add to Wish List
See related items
Why was I recommended this?
Rate this item | I own it | Not interested

9. Beginning Perl for Bioinformatics
by James Tisdall
Average Customer Review:
Publication Date: October 15, 2001
Our Price: $26.37 Used & new from $11.99
Add to cart | Add to Wish List
See related items
Why was I recommended this?
Rate this item | I own it | Not interested

10. Mastering Perl for Bioinformatics
by James D. Tisdall
Average Customer Review:
Publication Date: June 2003
Our Price: $26.37 Used & new from $25.35
Add to cart | Add to Wish List
See related items
Why was I recommended this?
Rate this item | I own it | Not interested

Finally, …

What is the Internet?

- The Internet was originally designed as an "overlay" network running on top of existing phone and other networks. It is based on a small set of software protocols that direct routers inside the network to forward data from source to destination, while applications run on the Internet to rapidly scale into a critical global service. However, this success now makes it difficult to create and test new ways of protecting it from abuses, or from implementing innovative applications and services.

http://www.intel.com/labs/features/idf09041.htm

How does the Internet work?

- Differences between the Internet and phone networks
  - Dedicated circuits/routes
  - Distributed, end-to-end
- Where is the intelligence?
  - Not in the network, per se, in the design and the ends
  - End-to-end Arguments in System Design
- Success of email, web, etc., relies on not building intelligence into the network
  - What about overlay networks?
  - What about PlanetLab?

What can be programmed?

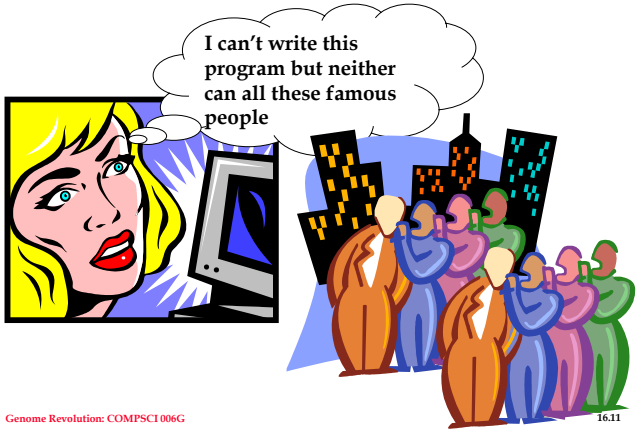
- What class of problems can be *solved*?
  - G5, 1000Mhz Pentium III, Cray, pencil?
  - Alan Turing proved some things, hypothesized others
    - Halting problem, Church-Turing thesis
- What class of problems can be *solved efficiently*?
  - Problems with no practical solution
    - What does practical mean?
  - Problems for which we can't find a practical solution
    - Solving one solves them all
    - Would you rather be rich or famous?

Schedule students, minimize conflicts

- Given student requests, available teachers
  - write a program that schedules classes
  - Minimize conflicts
- Add a GUI too
  - Web interface
  - …
  - …

One better scenario

Another possible scenario



The halting problem: writing `doesHalt`

```
public class ProgramUtils
    /**
     * Returns true if progname halts on input,
     * otherwise returns false (progname loops)
     */
    public static boolean doesHalt(String progname,
                                   String input){
    }
}
```

- A compiler is a program that reads other programs as input
  - Can a word counting program count its own words?
- The `doesHalt` method might simulate, analyze, …
  - One program/function that works for *any* program/input

How to tell if Foo stops on 123 456

```
public static void main(String[] args) {
    String prog = "Foo.java";
    String input = "123 456"
    if (ProgramUtils.doesHalt(prog,input)){
        System.out.println(prog+" stops");
    }
    else {
        System.out.println(prog+" 4ever");
    }
}
```

- Can user enter name of program? Input?
  - What's the problem with this program?

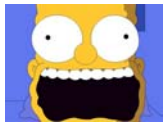
Consider the class *Confuse.java*

```
public static void main(String[] args){
    String prog = "Foo.java";
    if (ProgramUtils.doesHalt(prog,prog)) {
        while (true) {
            // do nothing forever
        }
    }
}
```

- We want to show writing `doesHalt` is impossible
  - Proof by contradiction:
  - Assume possible, show impossible situation results
- Can a program read a program? Itself?

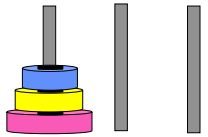
What's a meta catalog? Top 10 sites?

- Consider a website of interesting sites
  - Does the website list itself? Is this a problem?
- Consider a website that lists every useless website
  - Would this be a useful resource?
  - Does the website list itself?
- What about a site of all the sites that list themselves?
  - What about sites that don't list themselves? *nolist.com*



Not impossible, but impractical

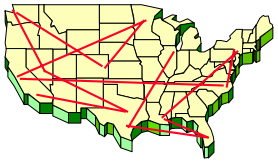
- Towers of Hanoi
  - How long to move n disks?



- What combination of switches turns the light on?
  - Try all combinations, how many are there?
  - Is there a better way?

Travelling Salesperson

- Visit every city exactly once
- Minimize cost of travel or distance
- Is there a tour for under $2,000 ? less than 6,000 miles?
- Is close good enough?
  - Within 10% of optimal
  - Within 50% of optimal
  - …



Try all paths, from every starting point -- how long does this take?

a, b, c, d, e, f, g
b, a, c, d, e, f, g …

Are hard problems easy?

- P = easy problems, NP = "hard" problems
  - P means solvable in polynomial time
    - Difference between N, N^2, N^{10} ?
  - NP means non-deterministic, polynomial time
    - *guess a solution and verify it efficiently*
- Question: P = NP ?
  - if yes, a whole class of difficult problems , the NP-complete problems, can be solved efficiently
  - if no, none of the hard problems can be solved efficiently
  - showing the first problem was NP complete was an exercise in intellectual bootstrapping, satisfiability/Cook/(1971)

Theory and Practice

- Number theory: pure mathematics
  - How many prime numbers are there?
  - How do we factor?
  - How do we determine primeness?
- Computer Science
  - Primality is "easy"
  - Factoring is "hard"
  - Encryption is possible

public-key cryptography

randomized primality testing

Computer Science in a Nutshell