Is Research For Me?

CRAW Undergraduate Workshop
Grace Hopper Celebration of Women in Computing 2010
September 29, 2010
Atlanta, GA

Can we “solve a research problem” in less than one hour?

• Is this task impossible?
  – Pretty rare to solve a problem that fast
• Is research challenging and frustrating?
  – Some mathematically challenging
  – Some involve figuring out how to put together things you already know
  – Can take a long time, reach dead ends
  – What if someone has already solved your research problem

Can we “solve a research problem” in less than one hour? (cont)

• Is research exciting and rewarding?
  – Solving a problem noone else has solved before feels pretty good.
  – You may invent something new and become famous
  – You may improve on a previous problem
    • solve faster than before
    • That could lead to faster software → happier customers
  – You may save lives!

Where do you find a research problem?

• ?
There are lots of research areas in CS

- Cars that drive themselves
- Underwater robots that reassemble themselves
- Predicting drug resistance from mutations
- Terrain modeling for flood forecasting
- False name manipulation in social networks
- Improving the diagnosis and treatment of disease using high-throughput clinical data
- Self organizing traffic lights
- Controlling a robotic wheelchair with your thoughts
- On-demand utility computing (“in the cloud”)
- etc.

Can you find too many research problems?

- Focus on an area
- Narrow your search down
- Don’t work on just one problem
- Still look at other areas – Why?

Now we will examine a research problem from Start to Finish

- Research problem that I solved when I was a graduate student
- We’ll do some problem solving
- We’ll do some thinking about the research process
- Laptops and phones off, no googling for solutions, let’s use our brains for the next hour

Should we work alone?

- Is research done alone?
- STOC 2010 (Symposium on Theory of Computing)
  - 78 papers, 235 authors, 3.01 authors per paper
- SIGCOMM 2010 (Communications and Networks)
  - 33 papers, 162 authors, 4.9 authors per paper
- CVPR 2010 (Computer Vision and Pattern Recognition)
  - 466 papers, 1509 authors, 3.24 authors per paper
- Let’s work in groups
First Pick an Area

• Area we picked: Computational Geometry

• This area has led to improvements in computer graphics software and computer-aided design and manufacturing (CAD)

• Visual (geometric shapes) and mathematical

Example: Convex Hull problems

Graphics leads to careers in Animation

Problem: Dynamic Maintenance of Maximal Points in a Plane

• Points in the x-y plane
• We will calculate which points are maximal
• As the points come and go, we want to be able to quickly list out the maximal points (dynamically)

• We will look at something a little simpler
We have decided on a problem to work on, now where do we start?

• ?

Definition of Maximal Points in a Plane

A point $p_i = (x_i, y_i)$ in the x-y plane is maximal

if there is no point $p_j = (x_j, y_j)$

such that $x_j > x_i$ and $y_j > y_i$

Points in the (x,y) plane, which are maximal points?

Maximal Points – form a staircase or m-contour
Simpler Problem – Find the maximal points statically first.

• Given a set of points, find all the maximal points

• Write an algorithm

• Analysis: If you have n points, how long does it take to find all the maximal points?

A Smaller Dynamic Problem:

• Assume you have N points. You want to be able to do two operations dynamically
  • 1) Insertion operation: Given a new point, you want to decide if this point is a maximal point or not.
  • 2) List out the maximal points

• What kind of data structure do you use to store the maximal points?
• What is the algorithm for insertion and listing out the maximal points?
• What is the analysis of these operations? Insert all N points into your structure.
Problem: Dynamic Maintenance of Maximal Points

- Operations:
  - Insert a new point
  - Delete a point
  - List the points that are maximal

- How do deletions make it harder?

Deleting point 9

Algorithms? Data Structures?

- ?

Algorithms? Data Structures? (cont)
Presenting the Dynamic Contour Search Tree

- Here is the data structure I invented to solve this problem

- The points are in the leaves sorted by x-value
- Non-leaves contain several values, max y-value
- Links link the m-contour points together

Circular Linked list for transfer and next_down pointers

M-Contour after inserting point 9

M-Contour before and after deletion
Comparison with Other Approaches

- Dynamic Contour Search Tree (Frederickson and Rodger, 1990)
  - Insertion $O(\log n)$, Deletion $O((\log n)^2)$
  - List maximal points $O(m)$
- Overmars and Van Leeuwen, 1981 (balanced binary tree with queues at the non-leaf nodes)
  - Insertion $O((\log n)^2)$, Deletion $O((\log n)^2)$
  - List maximal points $O(m)$
- Willard and Leuker, 1985
  - Insertion $O(\log n)$, Deletions $O(\log n)$, but no easy way to list out maximal points

What was our contribution?

- We shaved off a “$\log n$” factor in deletions
- Did we save any lives? I don’t think so.

What has happened since?

Go to ACM Portal to see who cited our paper

There are six citations to our paper

CITED BY


杉山隆之, Dynamic maintenance of maximal 2-point sets, Proceedings of the Tenth Annual Symposium on Computational Geometry, 0.45-0.49, June 1994, Nice, France, Stony Brook, New York, United States

Hong-Yi Huang, Chih-Chiang Yu, Jiun-Feng Wang, Optimal Algorithms for the Interval Location Problem with Range Constraints on Length and Average, IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB), v.5 n.2, p. 281-290, April 2008
Results since then...

- Kapoor improved on our result
- Cheng and Janardan used our data structure to solve another problem
- Hsieh et al are looking at the maximal points as part of solving a problem for computational biology

Our research example is complete

- We just experienced several months of research followed by years later
- How do you get started in research?

Take Advantage of Research Opportunities

- [www.cra-w.org/undergraduate](http://www.cra-w.org/undergraduate)
  - Several research programs
    - Summer experience at another university
    - Academic year and summer at your institution
- Students have websites about their experience

DREU Student Web page

Welcome to my DREU website!

read all about my summer in the DREU program at Carnegie Mellon's [CUPS Lab in Pittsburgh](http://www.cra-w.org/undergraduate).
The Project

An Investigation Of Facebook Grouping

Final Report

Motivation

Recently, online social networking sites have become some of the most popular activities on the internet. Currently with more than 500 million users, Facebook is huge. However, concerns about the privacy of users have increased. And even though Facebook has improved the user interface for privacy settings, there is still room for more user control within an individual's friend network. The average number of friends that users have is currently 130 and rising. It is not uncommon for somebody to have more than 300 Friends, and many people have more than 600 Friends. This leads to the question, how many of your Facebook “Friends” are really your friends? Is there an easy way to interact with only certain groups of your Friends at different times?

To this purpose, Facebook has the Friend Lists feature which allows users to group their friends in different lists, which they can then use to control their privacy settings. Unfortunately, the interface to make or edit a Friend List currently requires users to click through all of their Friends. This becomes a very difficult task for users that have more than say, 100 Friends, which is more than the average user.

Information about the student

curious about me?

Yael Mayer

I am currently a Computer Science major at Harvey Mudd College. My interests are currently very broad. I enjoyed the RC1 and Privacy Research I did during my DREU summer, but I am also interested in more theoretical computer science. I love interdisciplinary projects and think that the most exciting projects happen where many fields meet. My plan is to see a many TED talks as possible and figure out what my favorite stuff is.

Harvey Mudd College

Computer Science

Rising Senior, 2011

myayer@home.edu

Other things I like

Music: I love music. I play the violin and I’d love to learn how to play the ukulele this summer. I plan to pick up the guitar as well. I love basketball and I do some climbing and backpacking. Any outdoor stuff, really. I love it.

Information about the research team

meet the team:

PhD Students

Patrick Bailey, COS

Sanjana Komanduri, COS

Rich Shah, COS

Kami Veres, CSD

Chris brave Liao, EPP

Research Scientist

Yung Wang, Ph.D., C4Lab

Undergraduates

Greg Narciso, University of Pittsburgh

Robin Brewer, University of Maryland, College Park (DREU)

Yael Mayer, Harvey Mudd College (DREU)

Daily Journal

blogging...

Week 1: May 23 – May 29

I met with Patrick and Loren, who gave me an idea of the project I will be working on. They pointed me to some papers on privacy in social networking sites (SNS), as well as some research the C4U published recently.

I looked up more background research on Facebook staistics, social networking sites, and how people tend to group friends online.

We started brainstorming questions to think about how they would like to have their friends grouped on SNS, specifically Facebook. To be able to carry out interviews myself, I had to complete an IRB certification.

I will also be helping Rich and Sanjana on another project studying password requirements. I also went up to LA this past weekend to study the requirements of some major websites to get background information on the current state of password requirements.

Week 2: May 30 – Jun 5

Patrick, Robin and I finished a first draft of the questions that we would like to use in our surveys during the interviews. We continued to look for related research. We broadened our research to psychological and sociological papers studying social networks, as well as different papers on social network vocabularies.

We also worked on a draft of everything that we want to say and ask during the interviews. It is harder than I had realized to come up with good
Other Research Opportunities

• NSF REU programs
  www.nsf.gov/crssprgm/reu
  – Summer programs at other institutions
  – Deadlines in February or early March

• The complete slides for this talk will be put on the Grace Hopper 2010 site and also here:
  www.cs.duke.edu/~rodger under talks