Francis Galton
If you guess the right number of jellybeans, you may win a prize.

http://www.flickr.com/photos/jlmiller/35599670
Requirements for “Wise” Crowds

Diversity of Opinion
Independence
Decentralization
Aggregation

Monday, January 24, 2011
When computers were human...
NASA Clickworkers (2000)

Instructions: Please click on four points around the

* Show the craters I've marked.

You can start clicking on the next crater — undo the last one, if it's wrong.

When you're done:

1. I've marked some of the craters, but can't finish.
2. I've marked every crater larger than this.

2. Select which one of the following three classes (Fresh crater, degraded crater, "ghost" crater) best describes that crater. Then press "Send".

- Fresh crater
  Displays a sharp rim, distinctive ejecta blanket, and well-preserved interior features (if any). Central peaks count as interior features, but craters within the crater do not.

- Degraded crater
  Surrounding ejecta blanket is gone. Interior features are largely or totally obliterated. Rim is rounded or  flattened.

- "Ghost" crater
  Crater is faintly visible through overlying debris.
Search for Steve Fosset (2007)
<table>
<thead>
<tr>
<th>Requester</th>
<th>HIT Expiration Date</th>
<th>Reward</th>
<th>HITs Available</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dolores Labs</td>
<td>Jan 31, 2011 (6 days 23 hours)</td>
<td>$0.02</td>
<td>9763</td>
</tr>
<tr>
<td>Oscar Smith</td>
<td>Jan 24, 2011 (1 hour 59 minutes)</td>
<td>$0.02</td>
<td>77</td>
</tr>
<tr>
<td>Oscar Smith</td>
<td>Jan 24, 2011 (1 hour 59 minutes)</td>
<td>$0.02</td>
<td>34</td>
</tr>
<tr>
<td>nimrod shoam</td>
<td>Jan 27, 2011 (2 days 23 hours)</td>
<td>$0.10</td>
<td>1</td>
</tr>
<tr>
<td>Andy Guy</td>
<td>Feb 3, 2011 (1 week 2 days)</td>
<td>$2.10</td>
<td>1</td>
</tr>
<tr>
<td>techlist</td>
<td>Jan 24, 2011 (2 hours 45 minutes)</td>
<td>$0.01</td>
<td></td>
</tr>
</tbody>
</table>
Business Cards Solved

step 1
point & capture

step 2
real people

step 3
contact done

GET STARTED

Simple
Accurate
Reliable
CashAudit Alert!

Date/Time: July 18, 2010 at 11:44 PM
ALERT: Cash On Counter (Code 3)
View Video

Date/Time: July 18, 2010 at 11:46 PM
ALERT: Cash Not Put In POS (Code 4)
View Video

This email automatically generated by ReTel Technologies and sent to mgr_2438@___.com

To change your email settings and distribution lists, please click here.
most of their images shareable content.

Other's images, Pixazza connect experts through
tagging service, any tagging functionality.
you send
samasource
a project

samasource
breaks it down
into microwork

work is allocated to
our service partners

women, youth, and
refugees complete
work

samasource
compiles work and
assures quality

your project gets
delivered and helps
reduce poverty

samasource.org
<table>
<thead>
<tr>
<th>What's the craziest thing your cat has ever done?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Requester:</strong> Bjoern Hartmann</td>
</tr>
<tr>
<td><strong>HIT Expiration Date:</strong> Jul 27 2008, 07:33 AM PDT</td>
</tr>
<tr>
<td><strong>Reward:</strong> $0.08</td>
</tr>
<tr>
<td><strong>Assignments Requested:</strong> 30</td>
</tr>
<tr>
<td><strong>Assignments Pending Review:</strong> 0</td>
</tr>
<tr>
<td><strong>Reviewed Assignments:</strong> 30</td>
</tr>
<tr>
<td><strong>Remaining Assignments:</strong> 0</td>
</tr>
<tr>
<td><strong>Remaining Time:</strong> Expired</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sketch a cat</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Requester:</strong> Bjoern Hartmann</td>
</tr>
<tr>
<td><strong>HIT Expiration Date:</strong> Aug 14 2008, 08:24 AM PDT</td>
</tr>
<tr>
<td><strong>Reward:</strong> $0.05</td>
</tr>
<tr>
<td><strong>Assignments Requested:</strong> 50</td>
</tr>
<tr>
<td><strong>Assignments Pending Review:</strong> 0</td>
</tr>
<tr>
<td><strong>Reviewed Assignments:</strong> 50</td>
</tr>
<tr>
<td><strong>Remaining Assignments:</strong> 0</td>
</tr>
<tr>
<td><strong>Remaining Time:</strong> Expired</td>
</tr>
</tbody>
</table>
Amazing but True Cat Stories
by Folks on the Internet
My cat has only one eye and he can open anything. He is a master at undoing the lid to his cat food container. Anything that screws on or latches he can get open. I have tried every kind of container and within 1 day he has it figured out. He can open any door. He loves to open the refrigerator and let the dog eat all she wants. He has even opened the front door for a visitor and they were shocked when the door opened and he was the only one sitting there. He can open the sliding door and any cabinet he feels the need to open.
<table>
<thead>
<tr>
<th>Rank</th>
<th>Votes</th>
<th>Answers</th>
<th>Views</th>
<th>Title</th>
<th>Tags</th>
<th>Age</th>
<th>Poster</th>
<th>Views</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>2</td>
<td>26</td>
<td>Help with Javascript to C# converter</td>
<td>c#, javascript, converter</td>
<td>13m ago</td>
<td>griegs</td>
<td>6,893</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>6</td>
<td>76</td>
<td>How to comment/uncomment in HTML code</td>
<td>html, coding-style, coding, comment</td>
<td>19m ago</td>
<td>Jeff Knecht</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>45</td>
<td>Contents of a static library</td>
<td>c++, c, library, gcc</td>
<td>40m ago</td>
<td>Jonathan Leffler</td>
<td>61.3k</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>2</td>
<td>24</td>
<td>Most Efficient way to 'look up' Keywords</td>
<td>c++, search, efficiency, lexer</td>
<td>2m ago</td>
<td>ergosys</td>
<td>3,567</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>3</td>
<td>11</td>
<td>unknown css effects in moz and webkit</td>
<td>css, webkit, effects, tips-and-tricks, mozilla</td>
<td>2m ago</td>
<td>Pandiya Chendur</td>
<td>6,317</td>
</tr>
</tbody>
</table>
Craft Brewery Start-up Needs Your Help!!

Entries most recent first
The Dark Side of Crowdsourcing

Surveillance

Alienation & Deskilling

Moral Valence

Race to the Bottom

We should be careful to create a world we actually want to live in. (Stu Card)
Iranian protesters appearing in widely disseminated online photos from the ongoing post-election demonstrations in Iran, are now being targeted on a website of the Islamic Revolutionary Guard Corps.

The website [...] shows images of 20 people with red circles drawn around their faces claiming without evidence that they have been involved in creating “chaos” in Tehran.

Citizens are invited to call or email if they can identify the people on the photos. Gerdab also claims that two of the people depicted have already been arrested. The site provides no further information about any of the depicted people.

globalvoices.org
Actual Sighting Videos - BorderWatchSM Archives

Virtual Stake Outs - Live Border Cameras

Camera 1
This is a known drug traffic area. If you see people crossing from the left to right over these tracks please report this activity.

Camera 2
This area is known of illegal drug activity. If you see people moving in this area please report activity.

Camera 3
This area is a high crime area. If you see suspicious activity in this area please report. When focused on the river, please report subjects crossing via raft or swimming. If focused on land, look for subjects on foot moving towards right and may approach a vehicle parked along this area please report it immediately.

Camera 4
If you see people along river and or subjects in the water please report this activity. If camera focused on land, you are looking for individuals in the thick.

Virtual Texas DeputySM Log In

What We Do
The Texas Border Sheriff's Coalition (TBSC) has joined BlueServoSM in a public-private partnership to deploy the Virtual Community Watch, an innovative real-time surveillance program designed to empower the public to proactively participate in fighting border crime. The TBSC BlueServoSM Virtual Community WatchSM is a network of cameras and sensors along the Texas-Mexico border that feeds live streaming video to www.BlueServo.net. Users will log in to the BlueServoSM website and directly monitor suspicious criminal activity along the border via this virtual footageSM.
Tradeoffs

Free or paid?

Friends, community or strangers?

Anyone (worldwide) or experts only?

Work product open or closed?

Simple or complex tasks?
New forms of (information) production:
Micro-task markets
Friendsourcing
Online social participation
Winner-takes-all competitions

Which of these are crowdsourcing?
Which of these are valuable research domains?
Uses of MTurk-like systems in CS

Participant pool for user studies, experiments
Source of annotations for computer vision and NLP
HCl: worker pools hardwired into UIs
Research in crowd programming models, systems
Task markets as subjects of study: Economics, demographics
As a driver of international development
Uses of MTurk-like systems in CS

Participant pool for user studies, experiments
Source of annotations for computer vision and NLP

HCI: worker pools hardwired into UIs

Research in crowd programming models, systems
Task markets as subjects of study: Economics, demographics

As a driver of international development
VizWiz
Bigham, UIST 2010 (Best Paper Award)

Mobile service that aids blind users with “visual questions” in near-realtime

http://www.cs.rochester.edu/u/jbigham/vizwiz/video/
“Double-tap to take a photo.”

“Double-tap to begin recording your question and again to stop.”

“The first answer is ‘The right side,’ the second answer is …”

“Which can is the corn?”

Local Client
Remote Services and Worker Interface

Database -
Web Server -
Speech Recognition -

quikTurkit

Monday, January 24, 2011
Techniques for reducing answer latency:
Recruit workers on application startup (or maintain pool)
Keep workers busy with previously completed tasks, then move new tasks to top of queue
Soylent

A Word Processor with a Crowd Inside

Michael S. Bernstein  msbernst@csail.mit.edu

Greg Little, Robert C. Miller, David R. Karger, David Crowell, Katrina Panovich

Bjoern Hartmann

Mark Ackerman
ideas = []
for (var i = 0; i < 5; i++) {
    idea = mturk.prompt("What’s fun to see in New York City? Ideas so far: "+ ideas.join(",",")
    ideas.push(idea)
}

ideas.sort(function (a, b) {
    v = mturk.vote("Which is better?", [a, b])
    return v == a ? -1 : 1
})
Assumption: local computation is cheap & fast; Mturk calls are expensive & slow

Memoize Mturk call results across executions

Crash script on any long-latency Mturk call that has not yet completed; automatically or manually rerun

Benefits:
iterative programming
retroactive print-line debugging
• Please transcribe as many words as you can.
• Put a * in front of words you are unsure about.

Iteration 4: TV is* festival ___ was *two *me ____ , *but ___ *is ___ ___ TV ____ . I *two ___ tv ___ ___ ___ *festival , ___ I ___ ___ is* ___ it ___ *festival .

Iteration 6: TV is supposed to be bad for you , but I ____ watching some TV *shows . I think some TV shows are *really *advertising , and I ____ ____ is good for the ____

Iteration 12: TV is supposed to be bad for you , but I am watching some TV shows . I think some TV shows are really entertaining , and I think it is good to be entertained .
Ongoing Research

**Research goal:** determine the boundaries of what can be achieved by crowdsourcing and peer production systems.

- Discover decomposition, aggregation, coordination and control primitives
- Survey task specification approaches (e.g., declarative programming, visual dataflow, query languages, ...)
- Develop **crowd programming abstractions and interfaces:**
  - Online management: *execution control* and *feedback*
  - Socio-technical equivalents of prog. lang. and software engineering:
    - **Type checking:** Predicting if your crowd program will perform
    - **Exceptions:** How can crowd workers report problems?
    - **Debugging & Testing:** How would one debug crowd programs?
Biggest Socio-Technical Challenge: Platforms & Testbeds

- How can we test new crowd work platforms at appropriate scale?
  *Test alternative markets, incentives and mechanisms (e.g., Aardvark, Quora)*

- Which questions should research tackle? Which will be answered by startups? *(students vs VC funding)*
Pragmatics
Organizers

Bjoern Hartmann (EECS)
Tapan Parikh (iSchool)
Mike Franklin (EECS)
Anand Kulkarni (IEOR)
Kuang Chen (EECS)
Administrivia

**Meeting times:** M 3:30-5pm, 373 Soda

**Office Hours:**
- Bjoern: W 2:30-3:30pm, 629 Soda
- Tap: F 10-11am, 303B South Hall
Units

Take the course for 1 unit if you only want to read papers (and post responses)

Take the course for 2-4 units if you also want to complete a research project

3 hours of work week per units outside of class

We have project ideas...
Course wiki:

bit.ly/cal-crowd-course
Main Page

Contents

1 CS298-52: Crowdsourcing Seminar
2 Format
3 Preliminary Schedule
4 Potential additional topics

CS298-52: Crowdsourcing Seminar

- UC Berkeley, Spring 2011
- Mondays, 3:30-5:00, 373 Soda Hall
- 1-3 units, CCN 27226
- Organizers: Profs. Bjoern Hartmann, Mike Franklin, Tapan Parikh; Graduate Students Anand Kulkarni & Kuang Chen

Crowdsourcing and human computation have rapidly become interesting topics of research in Computer Science. In this reading seminar, we will survey and discuss the latest literature, with the following overarching themes:

- Algorithms, patterns and systems for crowd programming
- Organizational behavior and economic aspects of crowdsourcing
- Participatory sensing & crowdsourcing for development
- Demographics & labor issues of crowdsourcing
- Applications in active learning and AI

If you want to read and discuss papers only, sign up for 1 unit. You may take this course for up to three units if you are also working on a research project related to crowdsourcing as part of the class. Students are encouraged to collaborate on projects and choose projects that can be finished and submitted as papers to UIST 2011, the ACM Symposium on User Interface Software and Technology.

Questions? Email: crowdsourcing-seminar-staff@lists.eecs.berkeley.edu

Format

For each class period, students will submit short reading responses online by midnight on the day before class. Reading responses are submitted directly to this wiki.

In addition to commentaries, students will be asked to lead one class discussion. The discussant should read all student commentaries before class and...
Staff email:

crowdsourcing-seminar-staff@lists.eecs.berkeley.edu
Class Format

You will read 2 papers per week from the research literature and post comments before class on the wiki.

1-3 students will lead discussion each class.

A class project is optional.
Accessing Readings

All readings are linked from the schedule on the main wiki page.

Most readings are in the ACM digital library. On campus access is easy; from off-campus, you need to use a library proxy.

Access private readings with:
\[ l: \text{cs298} \quad / \quad p: \text{cs298Reading} \]
Submit **by Monday, 12:01 am (i.e., well before class), on the class wiki.**

To avoid locking/merging problems, please compose your response offline, then paste it in.
**Preliminary Schedule**

**Jan 24:** *Introduction, Project Opportunities*  
*In-class Reading 1:* [Crowd Control](#). Leah Hoffman, Crowd Control, Communications of the ACM Vol. 52 No. 3, Pages 16-17  
*In-class Reading 2:* [Five Rules of the New Labor Pool](#). Jeff Howe, Wired 14.06  
**Assignment** due by Friday Jan 28, 5pm: [Create Wiki Account](#)

**Jan 28:** *No class, but items due*  
**Due:** Create Wiki Account  
**Due:** Submit Class Petition

**Jan 31:** *Surveys and Taxonomies*  
**Paper 1:** [A Taxonomy of Distributed Human Computation](#). Alexander J. Quinn, Benjamin B. Bederson. UMD Tech Report 2009-23.  
**Assignment** due before class on Feb 7: [Create and Perform HITs on Mechanical Turk](#)

**Feb 7:** *Applications in HCI*  
**Paper 1:** [VizWiz: nearly real-time answers to visual questions](#). Jeffrey P. Bigham, Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, C. Miller, Robin Miller, Aubrey Tatarowicz, Brandyn White, Samuel White, and Tom Yeh. UIST 2010.  
**Paper 2:** [Soylent: a word processor with a crowd inside](#). Michael S. Bernstein, Greg Little, Robert C. Miller, Björn Hartmann, Mark S. Ackerman, David R. Karger, David Crowell, Katrina Panovich. UIST 2010.  
**Due:** Create and Perform HITs on Mechanical Turk  
**Due:** for project students: Project Ideas  
**Assignment** due before class on Feb 28: [Program HITs with the Mechanical Turk API](#)
Surveys and Taxonomies

Discussant's Slides and Materials

*Discussant: Upload your slides and link them here.*

Reading Responses

*Students: Submit your reading responses here. I have created a sample entry below.*

Bjoern Hartmann

My reading response here...

This page was last modified on 6 January 2011, at 03:19.

This page has been accessed 26 times.
Surveys and Taxonomies

== Discussant's Slides and Materials ==
"Discussant: Upload your slides and link to them here."

== Reading Responses ==
"Students: Submit your reading responses here. I have created a sample entry below."

---Bjoern Hartmann---
My reading response here...
What to write

One-sentence summary of the main point.

1-3 paragraphs of interpretation:
Why is/isn’t the paper important?
What new results, techniques or methodologies does it offer?
How does the paper relate to today’s technologies?
Does the paper relate to your own work?
What about blind spots?
Reading Responses: Tips

Read & write comments in one sitting

Turn off your cell phone, email, twitter, etc.

Go somewhere undisturbed  
(the library works well)
In-class discussion

You will lead at least one class discussion:
Summarize papers
Guide discussion & class activities

Submit preferences as part of your petition by Friday.
Assignments

Become familiar with current platforms:
Post HITs, work on HITs on MTurk
Post HITs through an API

Check-in for research project.

Important deadline: 4/22 UIST & HCOMP
Crowd Control

Using crowdsourcing applications, humans around the world are transcribing audio files, conducting market research, and labeling data, for work or pleasure.

Leah Hoffmann

Communications of the ACM
Vol. 52 No. 3, Pages 16-17
10.1145/1467247.1467254

Though computers have outstripped us in arithmetic and chess, there are still plenty of areas where the human mind excels, such as visual cognition and language processing. And if one mind is good, as the proverb goes, two—or two thousand—are often better. That insight, and its consequences, drew worldwide interest with the 2004 publication of James Surowiecki's best-selling The Wisdom of Crowds, which argued that a large group of people are superior at certain types of rational tasks than individuals or even experts.

Now researchers are turning to computers to help us take advantage of our own cognitive abilities and of the wisdom of crowds. Through a distributed problem-solving process