Crowd research

Michael Bernstein
Stanford University
http://crowdresearch.stanford.edu
Join us now at: crowdresearch.slack.com in #meetings

Email crowdresearch@cs.stanford.edu if you need an invite.
How might computing connect large groups to tackle bigger, harder problems than they could complete in isolation?
Today: small tasks, many people

• Combine many paid non-expert opinions
  
• e.g., text shortening
• e.g., image labeling
• e.g., data collection
### End Credits Identification

**Requester:** Video Data Gurus

**HIT Expiration Date:** May 23, 2015 (15 hours 59 minutes)

**Time Allotted:** 6 minutes

**Reward:** $0.01

**HITS Available:** 36

---

### Is this in the receipt

**Requester:** Vishwanath Kumar

**HIT Expiration Date:** Aug 1, 2015 (9 weeks 6 days)

**Time Allotted:** 60 minutes

**Reward:** $0.03

**HITS Available:** 1

---

### Are These Locations The Same?

**Requester:** Venue Quality

**HIT Expiration Date:** May 26, 2015 (3 days 2 hours)

**Time Allotted:** 5 minutes

**Reward:** $0.01

**HITS Available:** 2

---

### Product Search Relevance Side By Side - US (Closed)-Sat May 23 03:15:08 PDT 2015

**Requester:** Amazon Requester Inc.

**HIT Expiration Date:** Jun 2, 2015 (1 week 2 days)

**Time Allotted:** 10 minutes

**Reward:** $0.00

**HITS Available:** 1459
The status quo needs help.

• Today’s platforms are notoriously bad at...
  • ensuring high-quality results
  • producing fair wages and respect for workers
  • making it easy to author effective tasks
  • collaboration and discussion
  • and much more
What might we create if we knew that our children would become crowd workers?
our goal: create a new standard for crowdsourcing platforms
Designing a new future

- We will follow a user-centered research trajectory:
  - Empathy and needfinding
  - Brainstorming and ideation
  - Rapid prototyping and implementation
  - Launch
  - Evaluation

- …but you don’t need to be an expert at all of these!
Outcomes

- A new marketplace that we’re all proud to support and use
- Papers to top-tier academic venues — with you as coauthor
- A chance to learn with me as we experiment with new forms of research at scale
Alpha platform + UIST poster paper on July 9

Our first publication!
My current assumptions

- We will do our research in the open
- The platform will be open source
- Authorship will be determined by participation
Crowd worker fellowships

• If you are a crowd worker participating (e.g., from Amazon Mechanical Turk), we want to help defray the costs you spend not working on normal tasks.

• We have a fellowship budget that can support you for $50/month of participation as long as your team submits effective milestones. This isn’t complete payment, but we hope it helps.
The story so far

http://hci.st/crowdkickoff
Needfinding

- What struggles do workers and requesters face?
  - We tried to make money on Amazon Mechanical Turk
  - We tried to do tasks on Amazon Mechanical Turk and oDesk
- We interviewed major requesters and workers
Our two core problems

• Trust
  • Will the other person do what they said they will?

• Power
  • Who determines whether a task is good, or the work is good?
Foundations and features

- **Foundations** will likely be the major contributions of our platform, and set the stage. We start here.

- **Features** will make the platform better, but can’t solve problems of trust and power. We add these later.
Micro+macrotask market

- “Could the same marketplace scale from 2 to N people not just labeling images, but also Photoshopping my vacation photos or mixing my new song?”

- Maintain the submission approach from microtask markets, which is focused on two to hundreds of replications, but find ways to make it accessible to both microtask and expert work
Micro+macrotask market

• Adam Marcus: “Why would a programmer want to join a crowd work platform? The draw of AMT right now is that you can find jobs quickly. Can we enable that same benefit for programmers?”
Micro+macrotask market

- How we do this is still an open research question!
Challenges

• How do we ensure high-quality results? Do you let an expert work for hours and submit? That seems risky. Should there be intermediate feedback mechanisms?

• How do you trust that someone is an expert?

• Does this look like AMT where any expert certified in an area can accept the task? Or like an oDesk negotiation?
Micro+macrotask: suggestion

- All tasks can be taken up without negotiation by anyone who qualifies, and worked on immediately.
- For all task submissions on our marketplace, we require at least one milestone.
- That milestone serves as a checkpoint:
  - If it's a microtask, it can be after 5% of tasks are complete.
  - If it's a macrotask, it might be a description of what they should do first.
Micro+macrotask: suggestion

• The results of that milestone can be used to select specific workers to qualify to work on the rest of the task, or just to launch the rest of the tasks with no qualification.

• The requester can add as many milestones along the way as they want; we suggest one every couple days.
Input and output transducers

- Tasks get vetted or improved by people on the platform immediately after getting submitted, and before workers are exposed to them. Results are likewise vetted and tweaked.
Input and output transducers

• Also an open research question!
Challenges

- Cost: who pays for this? In other words, can this be done without hugely increasing the cost of crowdsourcing?

- Speed: is it possible to do this quickly enough to give near-immediate feedback to requesters? Like, 2–4 minutes?
  - Spamgirl: “The #1 thing that requesters love about AMT from her recent survey of requesters, is that the moment that I post tasks, they start getting done"
Challenges

• From Edwin: What happens when I have a task that I know is hard but I want workers to just try their best and submit? I’m OK with it being subjective, but the panel would just reject my task, which would have been frustrating.

• From Edwin: Could this help deal with people feeling bad when rejecting work? Maybe we need a new metaphor, like revision.
I/O transducers: suggestion

- Suggestion for input transducer:
  - While the task is in the first milestone stage, workers can leave feedback on the design of the task publically.
  - That feedback gets returned to the requester when the milestone completes.
  - The requester can use that feedback to iterate on the task before launching.
I/O transducers: suggestion

• Suggestion for output transducer:
  • By default, the platform checks a box that sends all work for review to a worker who is one (or two) levels more advanced than the worker before publishing it back to the requester.
  • It is done by publishing a new task back to the marketplace with the right qualifications.
  • This default adds cost and time, but addresses quality control.

• (Can we do this without adding significant time/cost?)
Reputation and expertise

• Instead of everyone getting 4.5 stars, can we make it in my best interest to rate honestly?
  • e.g., if I rate them highly, I’m more likely to get them as a worker in the future

• Could “guilds” of experts in a topic area determine whether you’re a Level 3 or Level 5 Photoshop expert?
## Workers on AMT

### Which feature will best support a fair and effective reputation system? (2w)

<table>
<thead>
<tr>
<th>Ideas</th>
<th>Score (0 - 100)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indicate when a worker/ requester has joined the system: the calendar date or the number of days, weeks, months, and/or years since inception.</td>
<td>75</td>
</tr>
<tr>
<td>Workers get promoted with good work reviews.</td>
<td>68</td>
</tr>
<tr>
<td>Traditional approach: workers and requesters can rate each other on a five-star scale after work is completed</td>
<td>65</td>
</tr>
<tr>
<td>Provide ability to test up to various levels of a skill set. (i.e. Level 5 Photoshop, Level 2 Python)</td>
<td>64</td>
</tr>
<tr>
<td>Reputation based on job rating (5 stars), private feedback, and platform participation (work evaluation, dispute moderation, forum interaction)</td>
<td>62</td>
</tr>
<tr>
<td>In addition to level based on skill, reputation also includes levels by types of tasks, content and knowledge domains.</td>
<td>62</td>
</tr>
</tbody>
</table>
## Requesters

### Which feature will best support a fair and effective reputation system? (3r)

<table>
<thead>
<tr>
<th>Ideas</th>
<th>Score (0 - 100)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guilds: more senior workers with same skills review your work periodically and promote you (e.g., Level 5 C++ programmer)</td>
<td>90</td>
</tr>
<tr>
<td>Rate worker on your desire to work with them on future projects.</td>
<td>89</td>
</tr>
<tr>
<td>Workers involved in work evaluation and work dispute moderation get higher reputation and/or pay.</td>
<td>86</td>
</tr>
<tr>
<td>Guilds of experts and levels within each guild (e.g., Level 3 C++). When you want to be promoted, you apply and the guild reviews your work.</td>
<td>83</td>
</tr>
<tr>
<td>Badges for high-performing and poor-performing workers.</td>
<td>82</td>
</tr>
</tbody>
</table>
Which feature will best support a fair and effective reputation system? (1R)

<table>
<thead>
<tr>
<th>Ideas</th>
<th>Score (0 - 100)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Workers involved in work evaluation and work dispute moderation get higher reputation and/or pay.</td>
<td>89</td>
</tr>
<tr>
<td>Reputation based on public/private/anonymous client/worker feedback. The system would create a reputation tied to privileges/incentives</td>
<td>80</td>
</tr>
<tr>
<td>Offer workers option to include professional online profiles such as GitHub repositories or LinkedIn entries</td>
<td>77</td>
</tr>
<tr>
<td>Guilds: more senior workers with same skills review your work periodically and promote you (e.g., Level 5 C++ programmer)</td>
<td>75</td>
</tr>
<tr>
<td>Incentive to rate honestly: if I give a high rating, the system makes it likely I get that person in the future. Low ratings → less likely.</td>
<td>75</td>
</tr>
</tbody>
</table>
Open governance

• Leadership shared by requesters, workers, (researchers?)

• Policy changes can be worked out by this group
Challenges

- Is it direct voting on everything? Or representative democracy?

- How exactly will this work?

- Can the research group have a hand here?

- If there are changes that require engineering effort, who executes that? Us? Other volunteers?
Research engineering

• Research in HCl involves bringing together many skills: there is no “research” vs. “engineering”.
• We design, we build, we test, we iterate.
• Let’s show you what we have so far!
How will this work?
Milestones

- Goal: each week, assign yourself to one or more milestones in the Crowd Research Trello board
  - Each milestone should have a member of the group who is leading it, known as the Directly Responsible Individual ("DRI") — you’re the DRI by default!
  - Each milestone or board will also have another person who can help you and mentor you. Ping them on Slack!
Milestones

- The Getting Started board is there now, and more will arrive soon.
- Milestones range from introductory skills like getting the code working to contributions like...
  - building a new feature
  - redesigning a user interface
  - brainstorming ideas for one of our four milestones
  - paper prototyping and storyboarding
  - …and more
Milestones

- Milestones are due each Thursday at 11:59pm Pacific Daylight Time

- Between Saturday’s morning meeting and Thursday, collaborate with everyone who has signed up on the same task to get it done!
  - Make a Slack channel to meet up
Milestones

- If you’re not done by Thursday night, our cleanup bot will remove you from the milestone on Trello.
- You can always sign yourself back up for it Friday morning!
- But deadlines make sure that we don’t drag our feet.

- Remember: we succeed or fail at this together!
Example milestones this week

• Do at least one of these…

• Learn about crowdsourcing
  • Go make $1 on Mechanical Turk
  • Go post a task on Mechanical Turk (we can PayPal you $ if you need)
  • Watch our panel where we interviewed big requesters and workers
  • Hello World with our code
  • Read related work
  • Help us name the platform
  • Many open GitHub issues to contribute to the code
Want a team?

• You are welcome to work in a team, or solo
• If you’d like to find a team, introduce yourself in #team-formation
Weekly team meeting

• Google Hangouts on Saturday morning @ 9am PDT
• Just like this!

• We will discuss hard open problems, figure out our next steps and milestones
• Discussion on Slack #meetings just like now
• Also: some folks each week will join us live on the hangout
Code on GitHub: http://hci.st/crowdgithub
Getting started with code

- http://hci.st/hellocrowd
- Meet up in #research-engineering

- The wiki page above will get you onboarded to your first pull request! Don’t worry about whether you know the right programming languages yet, we can help you follow it in #research-engineering