Prototyping

Michael Bernstein
Slides adapted from Scott Klemmer
Today

- How we prototype
- Milestone 5
- Overview of crowdsourcing research
Prototyping

It’s not the same as coding.
People thought Jeff Hawkins was crazy when they saw him taking notes, checking appointments, and synchronizing a small block of wood with his PC, pretending all the while that the block was a handheld computer.

“If I wanted to check the calendar I'd take it out and press the wooden button.”


Image Courtesy http://en.wikipedia.org/wiki/PalmPilot
Prototype: rapid approximation of a design idea used to gather feedback

Images Courtesy Bjoern Hartmann
Criteria for a prototype

- Constructed rapidly!
- Should not be required to be complete
- Should be easy to change
- Gets to retire
Versatile
Quick
Powerful

Storyboarding isn’t about “pretty pictures”. It’s about communicating ideas.
Let's check out places in SF...

1. Show interactions
2. Develop a setting
3. Show satisfactions

&a finaly, be creative! You don't need to be an artist to get a point across.

adapted with permission from Amal Dar Aziz, Guide to Storyboarding, http://hci.st/story
Storyboards are... a story!

- Setting
  - People involved
  - Environment
  - Task being accomplished
- Sequence
  - What steps are involved?
  - What leads someone to use the app?
- Satisfaction
  - What's motivates people to use this system?
  - What does it enable people to accomplish?

adapted with permission from Amal Dar Aziz, Guide to Storyboarding, http://hci.st/story
Paper prototyping

Lifealytics

Discover what makes you happy

BEGIN
How do you choose?

- What’s your question? What’s the biggest unknown thing about your idea?
  - How it looks?
    - Paper prototype, Balsamiq, proto.io, Photoshop, Sketch 3, sketchboard.io, Powerpoint, Keynote
  - How people will interact?
    - Storyboard, video, Powerpoint(?)
  - How is it embedded peoples’ everyday lives?
    - Storyboard, video prototype
Milestone 5

Ideation and prototyping
Our goals now

- We are going to be deepening our exploration for a subset of the ideas.
- We’ll use this deepening to lead us to new ideas, as well as select directions that we’re going to focus on.
- Our task is to figure out which approaches will be most fruitful.
Prototype two ideas

• Pick a category from last week, and choose two ideas in that category that you find inspiring (or riff to create a new one).

• Create prototypes for both ideas:
  • If it’s a user interface idea: make a low-fidelity mockup (e.g., using Balsamiq, Justinmind, or Canva)
  • If it’s a social interaction idea: make three “social storyboards”: storyboards to showcase different ways the interpersonal interaction could play out
Crowdsourcing research

To prompt your ideas…
Can the whole be greater than the sum of the parts?

- Can technology guide large groups of people to tackle bigger, harder problems than they could in isolation?
- Help large groups come together to act…
  - At an expert level,
  - On complex tasks,
  - At a high level of quality.
Early crowdsourcing research

[Little et al., HCOMP 2009]

Two distributed workers work independently, and a third verifier adjudicates their responses

You (misspelled) (several) (words). Please spellcheck your work next time. I also notice a few grammatical mistakes. Overall your writing style is a bit too phoney. You do make some good (points), but they got lost amidst the (writing). (signature)
Early crowdsourcing research
[Grier 2007]

Two distributed workers work independently, and a third verifier adjudicates their responses

1760
British Nautical Almanac
Neil Maskelyne
Work distributed via mail
Two people doing the same task in the same way will make the same errors.
Mathematical Tables Project

- WPA project, begun 1938
- Calculated tables of mathematical functions
- Employed 450 human computers
- The origin of the term *computer*
Etymology

• Crowdsourcing term coined by Jeff Howe, 2006 in Wired

• “Taking [...] a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call.”
Success: games with a purpose

Label every image on the internet using a game

[von Ahn and Dabbish, CHI ’06]
Success: scientific collaboration

- FoldIt: protein-folding game
- Amateur scientists have found protein configurations that eluded scientists for years
More successes

Largest encyclopedia in history

Kasparov vs. the world

NASA Clickworkers

Disaster reporting

Collaborative math proofs

DARPA Red Balloon Challenge
Paid Crowdsourcing

- Pay small amounts of money for short tasks
- Amazon Mechanical Turk: Roughly five million tasks completed per year at 1-5¢ each [Ipeirotis 2010]

<table>
<thead>
<tr>
<th>Task</th>
<th>Reward</th>
</tr>
</thead>
<tbody>
<tr>
<td>Label an image</td>
<td>$0.02</td>
</tr>
<tr>
<td>Transcribe audio clip</td>
<td>$0.05</td>
</tr>
</tbody>
</table>

- Population: 40% U.S., 40% India, 20% elsewhere
- Population is broad and replicates many well-known studies
Major topics of research

Crowd algorithms
[Little et al., HCOMP 2009]

Incentives and Quality
[Mason and Watts, HCOMP 2009]
[Dow et al., CSCW 2012]

Crowd-powered systems
[Bernstein et al., UIST 2010]
[Bigham et al., UIST 2010]

AI for HCOMP
[Dai, Mausam & Weld, AAAI 2010]

Complex Work
[Kittur et al., UIST 2011]
Crowdsourcing algorithms
Goal: guide crowds as they work

- Designing crowdsourcing algorithms is often like designing a user interface that will keep a user “in bounds” on your application
- Challenges
  - Taking unexpected action
  - Trying too hard
  - Trying not hard enough
Crowdsourcing algorithm

- A generalized version of a workflow

- Iterative algorithms [Little et al. 2009]
  - Hand off from one worker to the next

- Most crowdsourcing processes are more parallel, but less interesting algorithmically
Crowdsourcing algorithms

- Open-ended editing: Find-Fix-Verify [Bernstein et al., UIST ’10]
- Graph search [Parameswaran et al., VLDB ’11]
- Clustering [Chilton et al., CHI ’13]
- and many more...

- When write an algorithm?
  If you tried this in a straightforward way, would crowds fail? Why?
Incentives and quality
Incentives

- Does paying more produce better work?
  - More work, but not higher-quality work
    [Mason and Watts, HCOMP ’09]

- Does feedback produce better work?
  - Self-assessment and expert assessment both improve the quality of work
    [Dow, Kulkarni, Klemmer and Hartmann, CSCW ’11]
Incentives
[Shaw, Horton and Chen, CSCW '11]

- Which of these approaches improve quality?
  - Comparison to other workers
  - Normative claims: “it’s important that you try hard”
  - Solidarity: your team gets a bonus if you are right
  - Humanization: “thanks for working; I’m Aaron.”
  - Reward or punish accuracy with money
  - Reward or punish agreement with money
  - Bayesian truth serum: predict others’ responses
  - Bet payment on the accuracy of your responses
Incentives

[Shaw, Horton and Chen, CSCW ’11]

- Which of these approaches improve quality?
  - Comparison to other workers
  - Normative claims: “it’s important that you try hard”
  - Solidarity: your team gets a bonus if you are right
  - Humanization: “thanks for working; I’m Aaron.”
  - Reward or punish accuracy with money
  - Reward or punish agreement with money
  - Bayesian truth serum: predict others’ responses
  - Bet payment on the accuracy of your responses
Motivations
[Antin and Shaw, CHI ’12]

• Ask workers: “I am motivated to do HITs on Mechanical Turk…”
  • To kill time
  • To make extra money
  • For fun
  • Because it gives me a sense of purpose

• List experiment: vary which reasons appear in the list, and ask how many reasons the participant agrees with
  • This technique counters social desirability bias
Motivations
[Antin and Shaw, CHI ’12]

· US workers
  · 40% overreporting of money as a reason to work

· India-based workers
  · 142% underreporting of killing time and 60% underreporting fun as reasons
  · Money was not over- or under-reported
Communitysourcing

Engaging Local Crowds to Perform Expert Work Via Physical Kiosks

Kurtis Heimerl, Brian Gawalt, Kuang Chen
Tapan Parikh, Björn Hartmann
University of California, Berkeley
Judging quality explicitly

- Gold standard judgments [Le et al., SIGIR CSE ’10]
  - Include questions with known answers
  - Performance on these “gold standard” questions is used to filter work
- Get Another Label [Sheng, Provost, Ipeirotis, KDD ’08]
  - Estimate the correct answer and worker quality jointly
  - Try it! https://github.com/ipeirotis/Get-Another-Label
Judging quality implicitly

[Rzeszotarski and Kittur, UIST ’12]

- Observe low-level behaviors
  - Clicks
  - Backspaces
  - Scrolling
  - Timing delays
- SVMs on these behaviors predict work quality
- Limitation: models must be built for each task
Crowd-powered systems
Why do it?

- Embed crowd intelligence inside of user interfaces and applications we use today
Automatic clustering generally helps separate different kinds of records that need to be edited differently, but it isn’t perfect. Sometimes it creates more clusters than needed, because the differences in structure aren’t important to the user’s particular editing task. For example, if the user only needs to edit near the end of each line, then differences at the start of the line are largely irrelevant, and it isn’t necessary to split based on those differences. Conversely, sometimes the clustering isn’t fine enough, leaving heterogeneous clusters that must be edited one line at a time. One solution to this problem would be to let the user rearrange the clustering manually, perhaps using drag-and-drop to merge and split clusters. Clustering and selection generalization would also be improved by recognizing common text structure like URLs, filenames, email addresses, dates, times, etc.
VizWiz
[Bigham et al., UIST ’10]

- Visual question answering for the blind

- 1 to 2 minute responses by keeping workers on fake tasks until needed
Crowd-powered databases

- Database with open-world assumptions:
  SELECT * FROM ice_cream_flavors
- Several university flavors
  - Berkeley: CrowdDB [Franklin et al., SIGMOD ’11]
  - MIT: Qurk [Marcus et al., CIDR ’11]
  - Stanford: Deco [Parameswaran et al. ’11]
- Tackling many important optimization questions: e.g., joins, ranking, sorting
Realtime crowdsourcing

[Bernstein et al., UIST '11]
Realtime crowdsourcing

- Realtime captioning using shotgun gene sequencing techniques
Artificial intelligence for crowds
TurKontrol: AIs guiding crowds
[Dai, Mausam and Weld, AAAI ’10]

- Workflow planning as a decision-theoretic optimization problem
- Trade off quality vs. number of workers required
  - POMDP to decide: do we need a vote? do we need more voters? do we need more improvement?
Complex work
Conflict and coordination

- What happens to collaboration costs as Wikipedia grows? [Kittur, Suh, Pendleton, and Chi, CHI '07]

Amount of direct work on articles goes down, and activity on coordination pages goes up
Conflict and coordination

- As more editors join, which kinds of coordination techniques succeed? [Kittur and Kraut, CSCW ’08]
  - Explicit: participation in talk pages
  - Implicit: set direction by making edits

More editors only improves article quality only with implicit coordination — a few take on a disproportionate amount of work.
CrowdForge
[Kittur et al., UIST ’11]

• Crowdsourcing as a map-reduce process
• To write a wikipedia page, partition on topics, map to find facts and then reduce into a paragraph
Turkomatic
[Kulkarni, Can, and Hartmann, CSCW '12]

- Let the workers decide on task design
- Is a task too complicated for $D$? If so, ask for sub-tasks and recurse. If not, do it yourself.

- Creating a blog with content:
Careers in crowd work
[Kittur et al., 2013]

- More and more people are engaging in online paid work: programmers, singers, designers, artists, …
- Would you feel comfortable with your best friend, or your own child, becoming a full-time crowd worker?
- How could we get to that point? What would it take?
  - Education
  - Career advancement
  - Reputation