Crowdsourcing Discussion 2/7

VizWiz: Nearly Real-time Answers to Visual Questions

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.
VizWiz

• “automatic and human-powered services to answer general visual questions for people with visual impairments.”

• Lets blind people use mobile phone to:
  1. Take a photo
  2. Speak a question
  3. Receive multiple spoken answers
Motivation

- Current technology uses **automatic** approaches to help blind people access visual information

- Problems: Error-prone, limited in scope, expensive
  - Ex: OCR cannot read graphic labels, handwritten menu, street sign
Motivation

Solution: Ask real people

• Can phrase questions naturally
  - “What is the price of the cheapest salad?” vs. OCR reading the entire menu

• Feedback
  - Real people can guide blind people to take better photos

• Focus on blind people’s needs, not current technology
Human-Powered Services

- Problem: Response latency
- Solution: quikTurkit (and some tricks)
  - “first attempt to get work done by web-based workers in nearly real-time”
  - maintain a pool of workers to answer questions
“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 -

Server

quikTurkit
quikTurkit

- “requesters create their own web site on which Mechanical Turk workers answer questions.”

- “answers are posted directly to the requester’s web site, which allows [them] ... to be returned before an entire HIT is complete.”

- “workers are required to answer multiple previously-asked questions to keep workers around long enough to possibly answer new questions”
Wiki Comments

• “it seems somewhat of a waste to have workers answering other questions until real queries come up. Given that latency is a general problem with MTurk, I wouldn't be surprised if new crowdsourcing platforms were developed to address this issue directly, maybe using social networks.”

• “I was also surprised that the authors didn't try to verify the responses...surely providing incorrect/spammy answers back to the blind person is less desirable than providing an answer in which you have more confidence after some small delay?”
Response from Jeffrey Bigham

- “I agree that strategies like signaling to workers when work is available may be more cost effective than keeping them busy-waiting, but busy-waiting ends up being cost effective if you have enough users. The additional complexity of the signaling system may not be worth it in the end.”

- “Our mechanism for dealing with answer quality was to present multiple answers to users. Most strategies you might consider to ensure answer quality end up delaying the answer -- for instance, waiting for other users to verify. We decided to rely on the user to make sense of the answers, especially given that answers were correct the majority of the time. We actually saw zero malicious answers.”
Discussion Question #1

quikTurkit achieves “nearly real-time” answers from real people.

What are some other scenarios where quikTurkit might empower automatic applications?

Discuss in teams of 2-4
Please answer the question about the photo below:

Audio Playing - Turn Up Your Volume!

Answer the question:

Unable to answer the question?
If you are unable to answer the question, please try your best to describe the image. If you have trouble describing the image, include instructions on how to make the picture better - is the image too blurry, too dark, or is there another problem?

Instructions
Type text from the image on the left in the box above. An automatic tool has tried to recognize the text, but it often makes mistakes which you should fix. Move your mouse over the image to see a magnified view.

This HIT is designed to quickly and accurately answer questions about photographs and describe audio. We pay you to answer the questions quickly, but will manually review all answers provided for accuracy within 24 hours. Please provide concise answers quickly.
Display Site

1. tableau a feuilles mobiles
2. Priorities Research Agenda
3. research agenda
4. Volume!
5. its telling about research
Discussion Question #2

The authors seem to focus on interface design for the blind users...but what about the interface for the crowdsourced workers?

Do you think improving this interface can lower latency? What are possible improvements for the current interface?
Deployment

• Results: 86.6% of first answers “correct”
• Average of 133.3s latency for first answer
• Problems: Photos too dark or too blurry and thus unanswerable.
  - VizWiz 2.0 detects and alerts users if photo is too dark or blurry
VizWiz: LocateIt

- Combine VizWiz with computer vision to help blind people locate objects

- Results: Similar in time, but less accurate (compared to barcode scanner)
Future Work

• Expand worker pools beyond Mechanical Turk (e.g. social network)

• Reduce cost by using game, volunteers, friends

• Improve interface to make photo-taking easier for blind people

• Combine automatic approaches to improve delay
Discussion

• What are the strengths and shortcomings of this paper?
• Is VizWiz a compelling system?
• How credible is the evaluation?
Wiki Comments

• “VizWiz's most useful contribution seems to be their ‘quickTurkit’ model”

• “how much of this is a function of Amazon Mechanical Turk, and how much is a function of the general idea of crowdsourcing”
Wiki Comments

• “Locatelt demonstrates a specific instance of what would be cool to explore further: question rewriting!”

• “Both papers evaluate the cost of their systems by comparing them to non-crowdsourced solutions for the problems. This is a reasonable practice, and especially effective in domains like VizWiz where there are expensive existing tools for the problem.”
Response from Jeffrey Bigham

“Looking back on VizWiz, I think what was interesting was that it showed (i) you can do crowdsourcing in something close to real-time (regardless of how you do it) and (ii) you can use resulting "deployable Wizard-of-Oz" prototypes to learn more about your population.”