

Internet Enabled Human Computation

CSE 454
Daniel Weld



Built in 1770 by Wolfgang von Kempelen



To do

- **Challenge - Mechanisms for deterring vandals**
 - Reputation
 - Gold standard answers
 - Randomized redundancy
- **Balloon challenge**
- **More on foldit**
- **Game design, plateaus & levels**

A screenshot of the Amazon Mechanical Turk website. The page features a navigation bar with 'Your Account', 'HTIs', and 'Qualifications'. The main content area includes a yellow banner stating 'Mechanical Turk is a marketplace for work' and '127,021 HITs available'. Below this, there are two columns: 'Make Money by working on HITs' and 'Get Results from Mechanical Turk Workers'. The 'Make Money' section lists benefits for workers like 'Can work from home' and 'Choose your own work hours'. The 'Get Results' section lists benefits for requesters like 'Have access to a global, on-demand, 24 x 7 workforce'. At the bottom, there are three circular icons representing 'Find an interesting task', 'Work', and 'Earn money'.

Crowdsourcing

“a neologistic compound of **Crowd** and **Outsourcing** for the act of taking tasks traditionally performed by an employee or contractor, and outsourcing them to a group of people or community, through an “open call” to a large group of people (a crowd) asking for contributions”

---[Wikipedia]

amazonmechanical turk
Artificial Artificial Intelligence

PowerSet

A screenshot of a Mechanical Turk HIT page. The page shows a search bar for HITs, a timer set to 60 minutes, and a list of HITs. One HIT is highlighted with a red circle around the 'Qualifications' section. The qualification is 'PowerSet Knowledge' with a requirement: 'PowerSet Knowledge semantic resources qualification' is greater than 9. A red arrow points from the word 'PowerSet' in the section header to the 'PowerSet' in the qualification name.

Find Substitutable Words

In the sentence below, what words or phrases could replace the **bolded** word without changing the meaning? F

Example:

In most countries **children** are required by law to attend school.

You might enter:

kid
youngster
pupil
young person

Try to enter single words or short phrases like "water bottle" or "post office." You are encouraged to use the "a station".

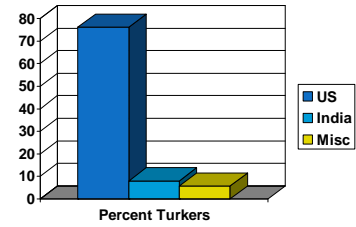
Avoid descriptive phrases, e.g. "a container you drink out of," or "a place you mail things from" unless you abs. Further, tell us how easy or difficult it is to assign one of several possible meanings for the **bolded** word in the

Your sentence is: The term silver dollar is often used for any large white metal coin issued by the United States with a **face** value of one dollar ; although purists insist that a dollar is not silver unless it contains some of that metal .

Enter *one term* per box.

\$0.05

Turker Demographics



March, 2008
(Panos Ipeirotis)

Fast & Cheap, but is it Good?

[Snow et al. EMNLP-08]

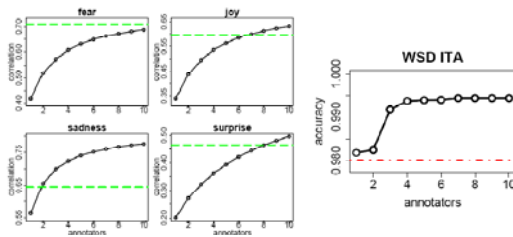
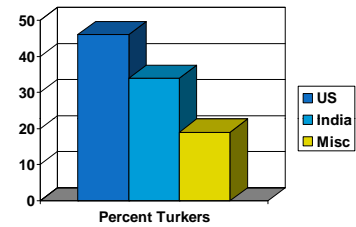


Figure 1: Non-expert correlation for affect recognition

Turker Demographics



February, 2010
(Panos Ipeirotis)

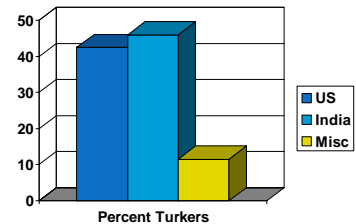
How Cheap + Fast?

[Snow et al. EMNLP-08]

In our experiment we ask for 10 annotations each of the full 30 word pairs, at an offered price of **\$0.02 for each set of 30 annotations** (or, equivalently, at the rate of 1500 annotations per USD). The most surprising aspect of this study was the speed with which it was completed; the task of 300 annotations was completed by 10 annotators in less than 11 minutes ...

1724 annotations / hour.

Turker Demographics

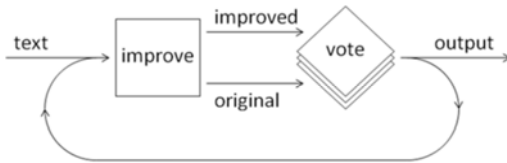


May, 2010
(Crowdfower)

<http://blog.crowdfower.com/2010/05/amazon-mechanical-turk-survey/>

Complex Jobs

- TurKit [Little 09]
- Casting Words



Iterative Improvement

Version 7



A close-up photograph of the following items:

A CASIO multi-function, solar-powered scientific calculator.

A blue ball point pen with a blue rubber grip and the tip extended.

British coins, two of £1 value, three of 20p value and one of 1p value.

Seems to be a theme illustration for a brochure or document cover treating finance – probably personal finance.”

TurKit [Little et al. 09]

- **Determine a fixed allowance**
 - Money spent in a problem
- **Each improvement iteration**
 - Ask two workers to vote
 - A third is asked if the first two disagree
 - Keep the artifact by majority vote

12/1/2010

14

Limitation: Workflow is Fixed

- **Number of iterations is determined**
 - By the allowance
 - Not by the quality of the answers or the workers
- **Number of votes / iter is almost fixed**
 - Not based on the difficulty of the job

12/1/2010

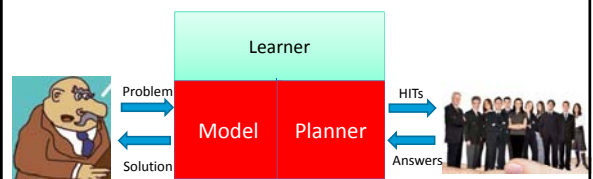
17

Iterative Improvement



A partial view of a pocket calculator together with some coins and a pen. ?

TurKontrol [Dai AAA10]



Input

- a picture
- an initial description

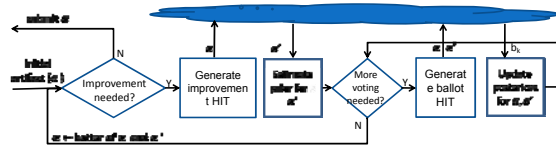
Output

- a high quality description

12/1/2010

18

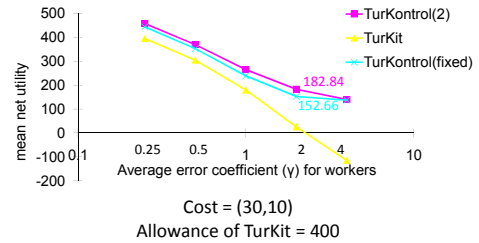
TurKontrol Workflow



12/1/2010

19

Comparison with Fixed Workflows



12/1/2010

22

Evaluation Measures

• Quality measure

- Quality improvement probability (QIP)
- An artifact has QIP q
 - $1 - \Pr(\text{an average worker improves the artifact})$
 - Never exactly known
 - Can be estimated by a random variable Q

• Utility function

- $U(q)$

12/1/2010

20

How Motivate People to Help?

- Money

Control Problem is a POMDP

- **States**
 - Artifact pair (a, a') with QIPs (q, q')
- **Belief space**
 - QIP distribution $(f_Q, f_{Q'})$
- **Actions**
 - Create a ballot HIT
 - Receives y/n ballot answers (observations)
 - Create an Improvement HIT
 - Transition to a new state (a, q)
 - Submit the best artifact
 - Go to a goal state

12/1/2010

21

DARPA Network Challenge

\$40k

10 Moored Weather Balloons



10am ET Saturday 12/5/09

Winner

MIT Red Balloon Challenge Team

All 10 Balloons – 8:52

Also notable:

Groundspeak Geocachers

7 Balloons – 6:02

<https://networkchallenge.darpa.mil/ProjectReport.pdf>

How Motivate People to Help?

- Money
- Altruism
- Esteem
- Self-Interest
- Fun

Successful Tools

- Marketing + media broadcast strategies to get team members
- Recursive, incentivized recruiting of networks to build team
- Extraction of reported locs from open iNet sources (eg Twitter)
- Automated means of extracting data, e.g. Twitter crawler
- Deployment of automatic reporting capability, e.g. iPhone apps
- Dispatching team members as spotters to confirm
- Website design that motivates, encourages recruitment, or allows easy, secure reporting
- Search engine rank optimization of website

Altruism



WIKIPEDIA
The Free Encyclopedia

Self-Esteem

Customer Reviews

3,314 Reviews

5 stars	(2,579)
4 stars	(419)
3 stars	(279)
2 stars	(66)
1 star	(75)

Average Customer Rating

★★★★☆ (3.314 out of 5)

Most Helpful Customer Reviews

515 of 581 people found the following review helpful

★★★★★ A stunning and thoroughly satisfy!

By **T. Burger** (Chicago) - [See all my reviews](#)

TOP 100 REVIEWER REAL NAME USER VOICE

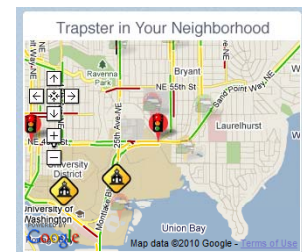
Recursive Incentivizing



Figure 5: Recruitment network hierarchy

- method that reached
- almost 5,400 individuals in approximately 36 hours. The ingenuity of the recruiting method was
- that the incentive to join the effort was transferred undiminished with each

trapster



Download Trapster For Free!



Collaborative Geomapping

- State Troopers Reaction to Trapster
- Motivation & Vandalism Control
- Other Applications
 - North Korea Uncovered (Google Earth)
 - DARPA Network Challenge

StackOverflow



Self-Interest



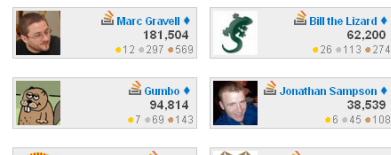
StackOverflow

Who Are You?

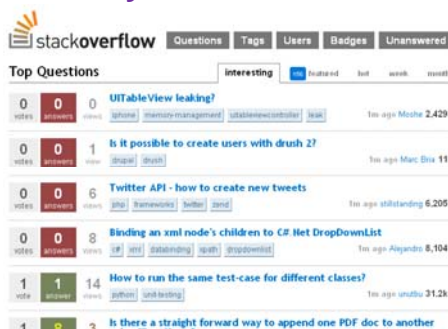
Mostly, we're software engineers who love building great software – just I

We're Also the Community

Site moderators, elected [May 19th, 2009](#) and [February 8th, 2010](#):



Hybrid Models



StackOverflow Optional Reputation

- Answer voted up +10
- Question voted up + 5
- Answer accepted +15 (+2 to acceptor)
- Post voted down - 2 (-1 to voter)

Max 30 votes / user / day

Reputation → Privileges

- 15 vote up
 - 15 flag offensive
 - 50 leave comments
 - 100 edit community wiki posts
 - 125 vote down (costs 1 rep)
 - 500 retag questions
 - 1000 create new tags
 - 2000 edit other people's posts
- Etc...

ACCESSIBILITY

LESS THAN 10% OF THE WEB IS ACCESSIBLE TO THE VISUALLY IMPAIRED

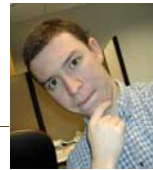
REASON: MOST IMAGES DON'T HAVE A CAPTION

Slides by Luis von Ahn

Motivating People

- Money
- Fun

LABELING IMAGES WITH WORDS



FACE
MAN
SUPER SEXY

STILL A COMPLETELY OPEN PROBLEM

Slides by Luis von Ahn

IMAGE SEARCH ON THE WEB



USES FILENAMES
AND HTML TEXT

Slides by Luis von Ahn

DESIDERATA

A METHOD THAT CAN LABEL
ALL IMAGES ON THE WEB
FAST AND CHEAP

Slides by Luis von Ahn

THE ESP GAME

TWO-PLAYER ONLINE GAME

PARTNERS DON'T KNOW EACH OTHER
AND CAN'T COMMUNICATE

OBJECT OF THE GAME:
TYPE THE SAME WORD

THE ONLY THING IN COMMON IS
AN IMAGE

Slides by Luis von Ahn

THE ESP GAME IS FUN

3.2 MILLION LABELS WITH 22,000 PLAYERS

MANY PEOPLE PLAY OVER 20 HOURS A
WEEK

Slides by Luis von Ahn

THE ESP GAME

PLAYER 1



GUESSING: CAR
GUESSING: HAT
GUESSING: KID
SUCCESS!
YOU AGREE ON CAR

PLAYER 2



GUESSING: BOY
GUESSING: CAR
SUCCESS!
YOU AGREE ON CAR

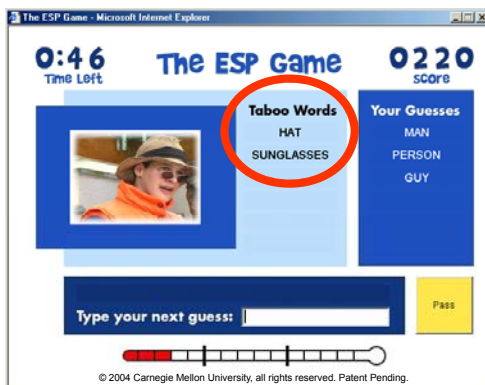
Slides by Luis von Ahn

LABELING THE ENTIRE WEB

5000 PEOPLE PLAYING SIMULTANEOUSLY CAN
LABEL ALL IMAGES ON GOOGLE IN 30 DAYS!

INDIVIDUAL GAMES IN YAHOO! AND MSN
AVERAGE OVER 10,000 PLAYERS AT A TIME

Slides by Luis von Ahn



Slides by Luis von Ahn

9 BILLION MAN-HOURS OF
SOLITAIRE WERE PLAYED IN 2003

EMPIRE STATE BUILDING
7 MILLION MAN-HOURS
(6.8 HOURS OF SOLITAIRE)

PANAMA CANAL
20 MILLION MAN-HOURS
(LESS THAN A DAY OF SOLITAIRE)

Slides by Luis von Ahn

GWAP

- Problem?

30 Photo Seed with Holes



PhotoCity Reconstructing the World in 3D Bringing Games with a Purpose Indoors

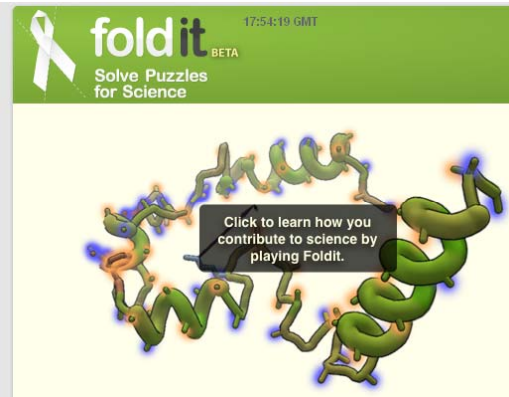


Mobile App



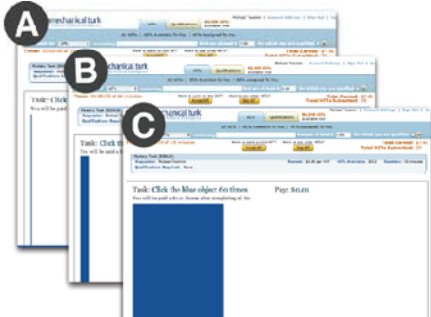
(a) iPhone

PhotoCity Gameplay



Hybrid Models Revisited

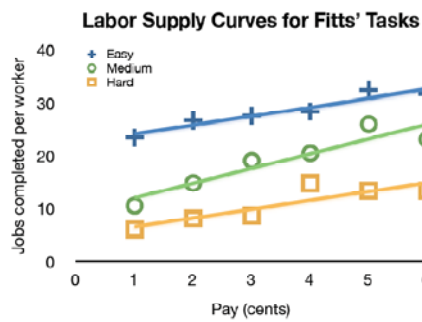
Effect of Pay on Job Completion



Hybrids

- What else could you add to a MT Task?
 - Leaderboards
 - Raffles
 - ????

Hybrid Models Revisited



Motivation

- Money
- Altruism
- Esteem
- Self-Interest
- Fun

Hybrid Models Revisited

