

# Crowdsourcing & Human Computation



Matt Lease

School of Information

University of Texas at Austin

[ml@ischool.utexas.edu](mailto:ml@ischool.utexas.edu)



# “Amazon Remembers”



## 1 See something you want to remember

When you notice an item to remember, tap “Remembers” in the Amazon App.



2

## Snap photo & send

Your iPhone camera will open. Take a photo of the item and it will be sent to Amazon.



3

## See reminders

Your photos & any similar products that Amazon finds are stored in the app and on Amazon.com.



**J. Pontin. *Artificial Intelligence, With Help From the Humans*. NY Times (March 25, 2007)**

# Amazon Mechanical Turk (MTurk)

## Make Money by working on HITs

HITs - *Human Intelligence Tasks* - are individual tasks that you work on. [Find HITs now.](#)

**As a Mechanical Turk Worker you:**

- Can work from home
- Choose your own work hours
- Get paid for doing good work

**Find an interesting task** → **Work** → **Earn money**

[Find HITs Now](#)

## Get Results from Mechanical Turk Workers

Ask workers to complete HITs - *Human Intelligence Tasks* - and get results using Mechanical Turk. [Get started.](#)

**As a Mechanical Turk Requester you:**

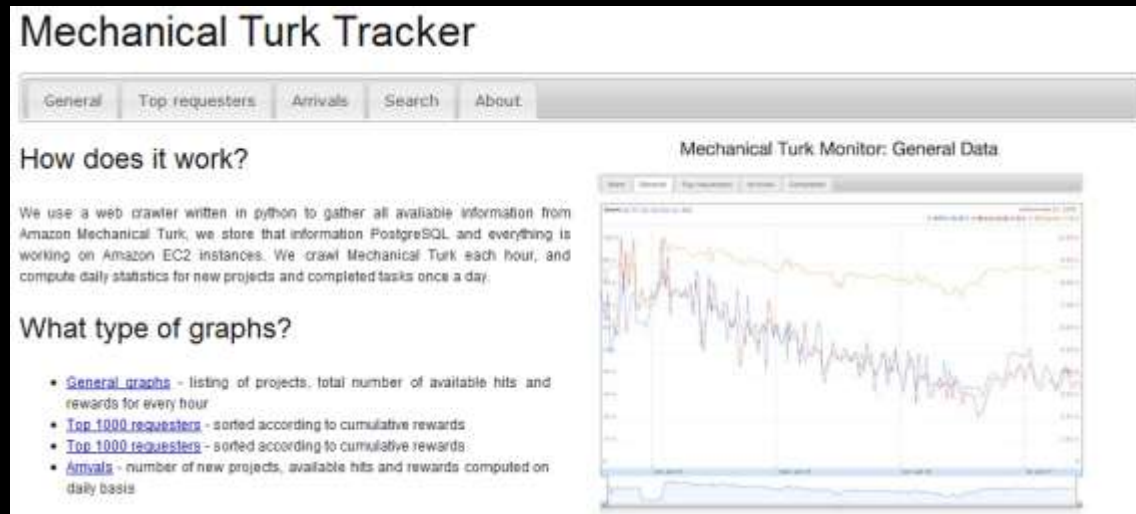
- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of HITs completed in minutes
- Pay only when you're satisfied with the results

**Fund your account** → **Load your tasks** → **Get results**

[Get Started](#)

- “Micro-task” crowdsourcing marketplace
- On-demand, scalable, real-time workforce
- Online since 2005 (and still in “beta”)
- Programmer’s API & “Dashboard” GUI

# From Outsourcing to Crowdsourcing



<http://www.mturk-tracker.com> (P. Ipeirotis'10)

From 1/09 – 4/10, 7M HITs from 10K requestors worth \$500,000 USD (significant under-estimate)

# Who are the workers?



- A. Baio, November 2008. [The Faces of Mechanical Turk.](#)
- P. Ipeitorotis. March 2010. [The New Demographics of Mechanical Turk](#)
- J. Ross, et al. [Who are the Crowdworkers?...](#) CHI 2010.

# Road Map

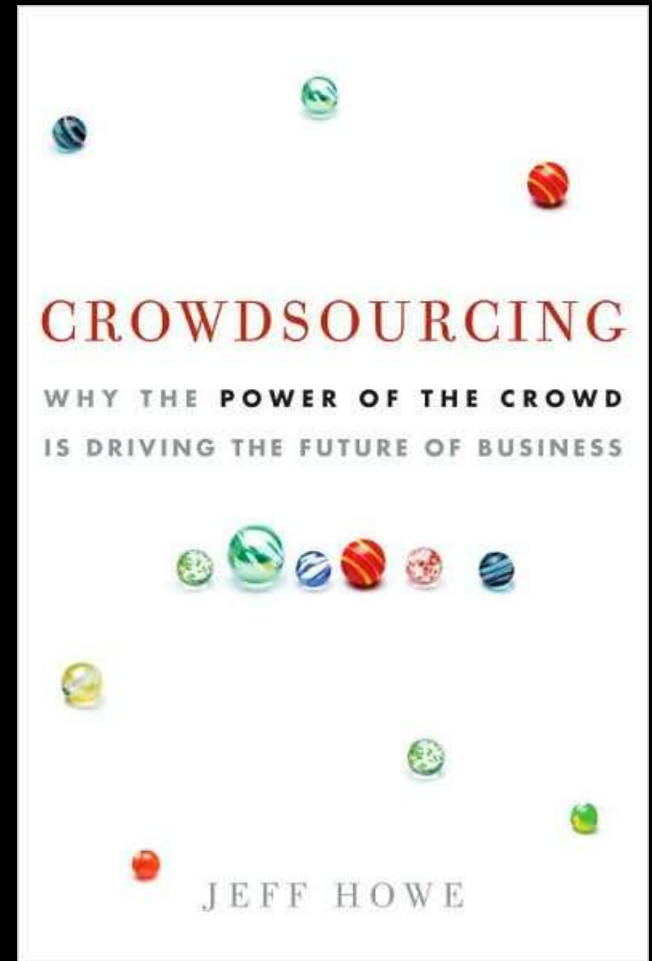


- Example and Introduction
- Crowdsourcing
- Human Computation
- Rethinking Application Design
- The Road Ahead
- Wrap-up

# Crowdsourcing

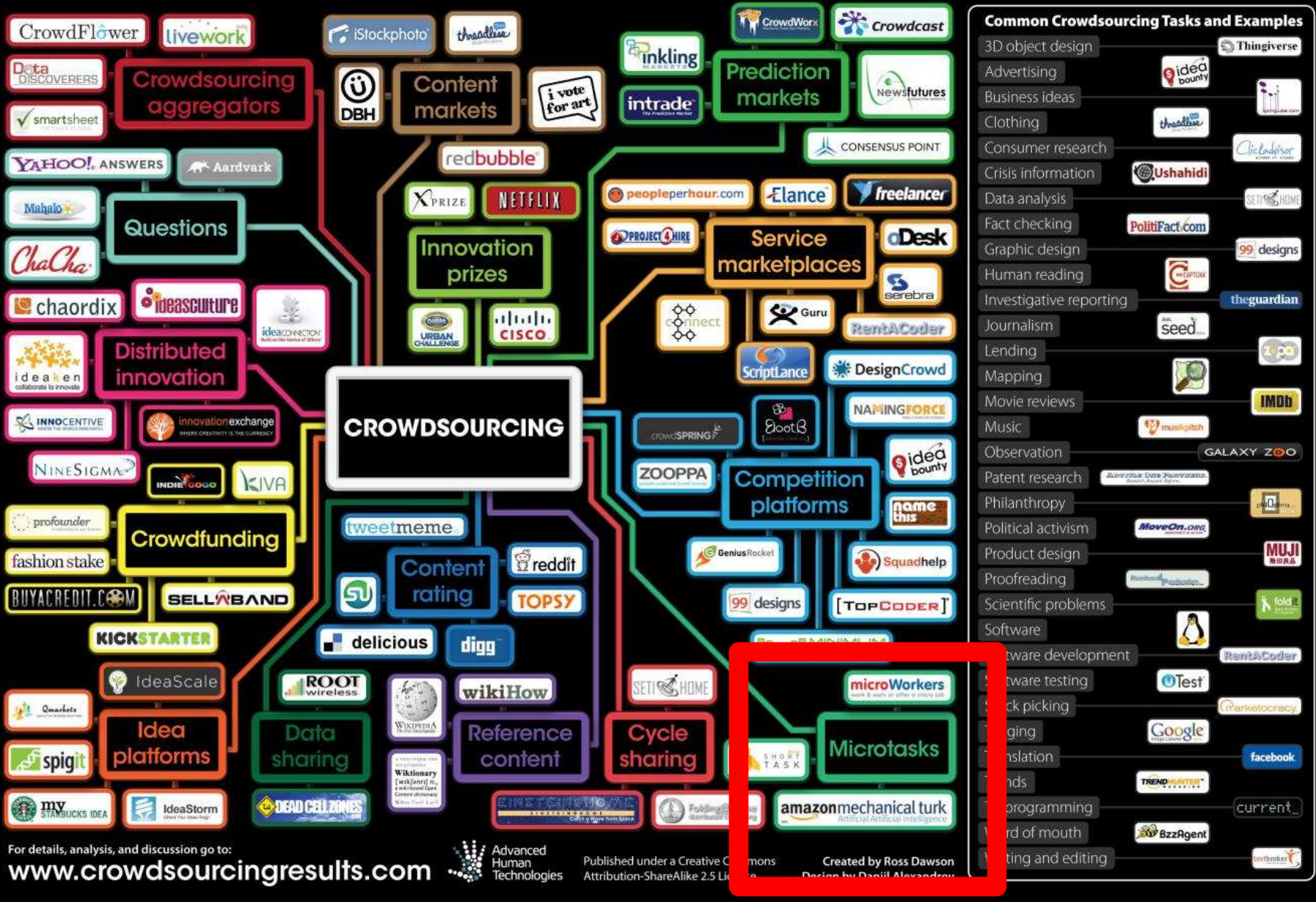
# Crowdsourcing

- Take a job traditionally performed by a known agent (often an employee)
- Outsource it to an undefined, generally large group of people via an open call
- New application of principles from open source movement





# CROWDSOURCING LANDSCAPE Beta v1



For details, analysis, and discussion go to:  
[www.crowdsourcingresults.com](http://www.crowdsourcingresults.com)



Published under a Creative Commons Attribution-ShareAlike 2.5 License  
 Created by Ross Dawson  
 Design by Danil Alexandrov

# Other Crowdsourcing Examples

**SAILOR MISSING SINCE 1/28/07**

Please contact the United States Coast Guard with any information.

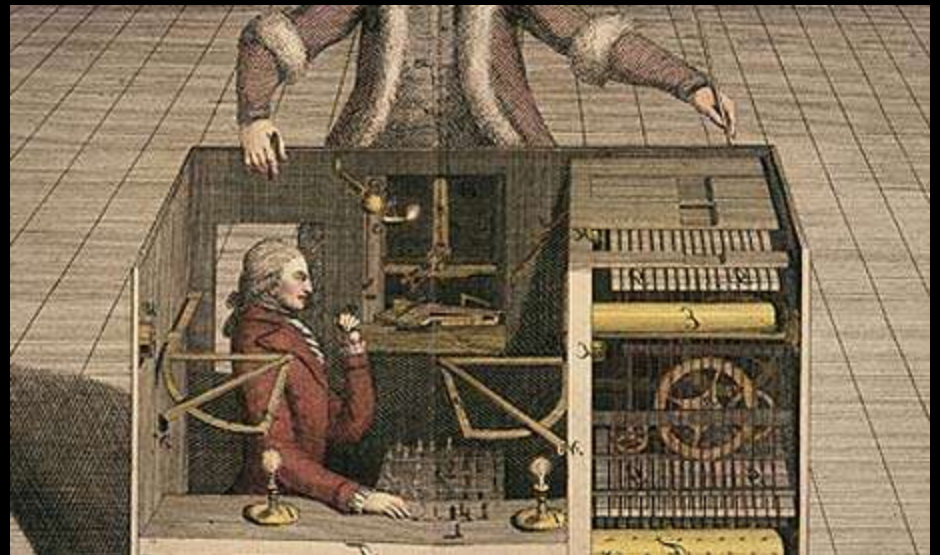
[Wired Article](#)   [NY Times Article](#)   [Ongoing Effort](#)   [I'd Like to Help!](#)   [Print a MISSING Poster](#)

...mination Done! We've examined more than 560,000 images from 3 satellites, covering nearly 3,500 square miles of ocean! We current



# Human Computation

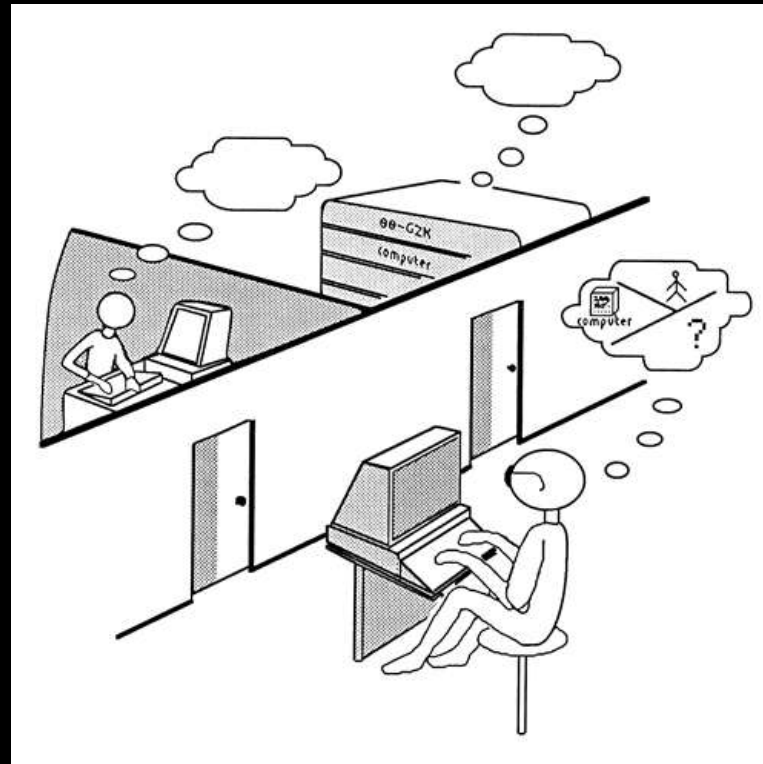
# The Mechanical Turk



The original, constructed and unveiled in 1770 by Wolfgang von Kempelen (1734–1804)



# The Turing Test (Alan Turing, 1950)





**Security Check**

Enter **both** words below, separated by a **space**.  
 Can't read this? Try another.  
 Try an audio captcha

Text in the box:

[← Back](#)

**PREMIER PHARMACY**  
 Contact **VIAGRA, CIALIS, LEVITRA** Online Price!

VIAGRA	30	\$134.95	CIALIS	30	\$169.95
VALIUM	30	\$85.45	SOMA	30	\$75.95
PROPECIA	30	\$64.95	AMBIEN	30	\$120.99

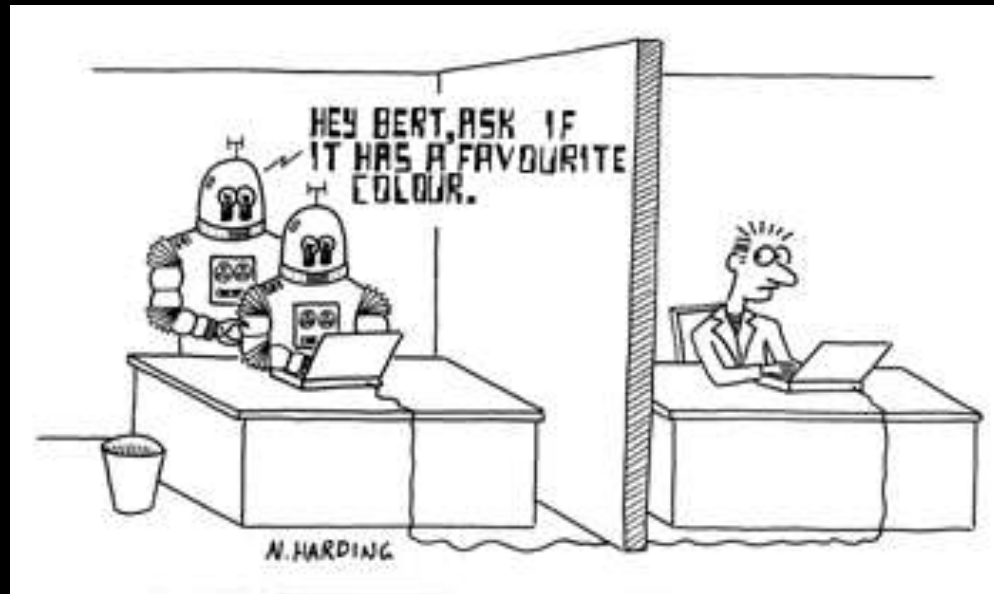
• XANAX 30 \$123.45    VIAGRA SOFT 50 \$250.99  
 ✓ New CIALIS SOFT 30 \$224.95

✓ Save up to 80% on your prescription Meds!

**5RX.ORG**

Do not click, type in your browser [www.5rx.org](http://www.5rx.org)

# The Turing Test (Alan Turing, 1950)



# What is a Computer?

com·put·er 



[kuh m-pyoo-ter]  [Show IPA](#)

–*noun*

1. Also called **processor**, an electronic device designed to accept data, perform prescribed mathematical and logical operations at high speed, and display the results of these operations. Compare analog computer, digital computer.

**Origin:**

1640–50; compute + -er<sup>1</sup>; compare Middle French *computeur*

–*Related forms*

com·put·er·like, *adjective*

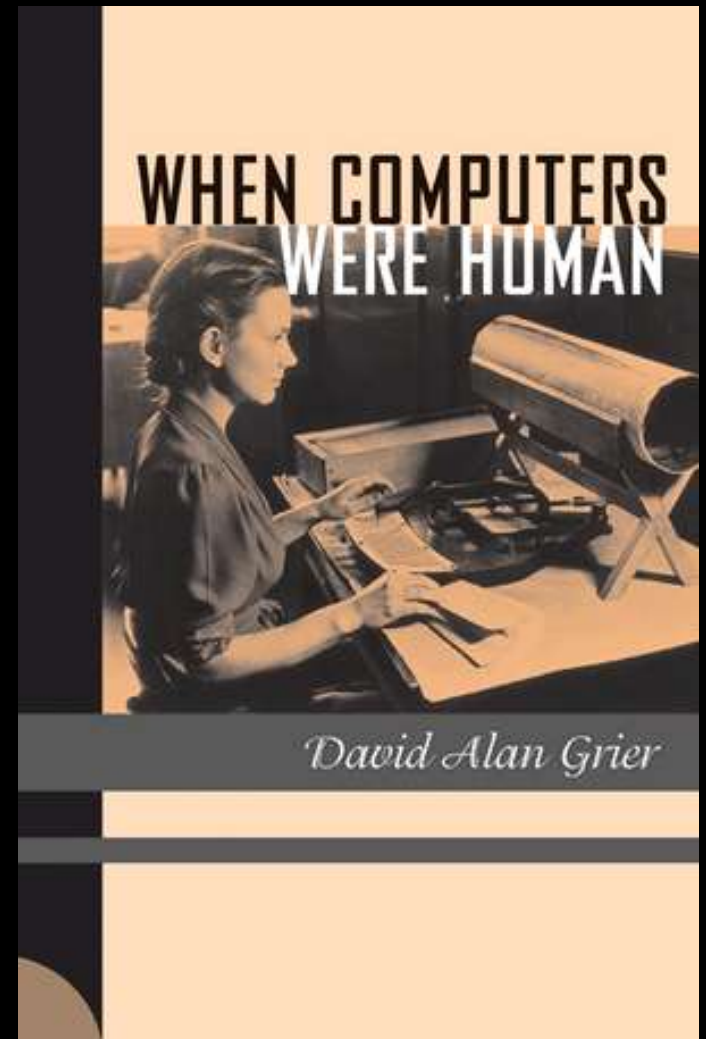
non·com·put·er, *adjective*

Dictionary.com Unabridged

Based on the Random House Dictionary, © Random House, Inc. 2011.

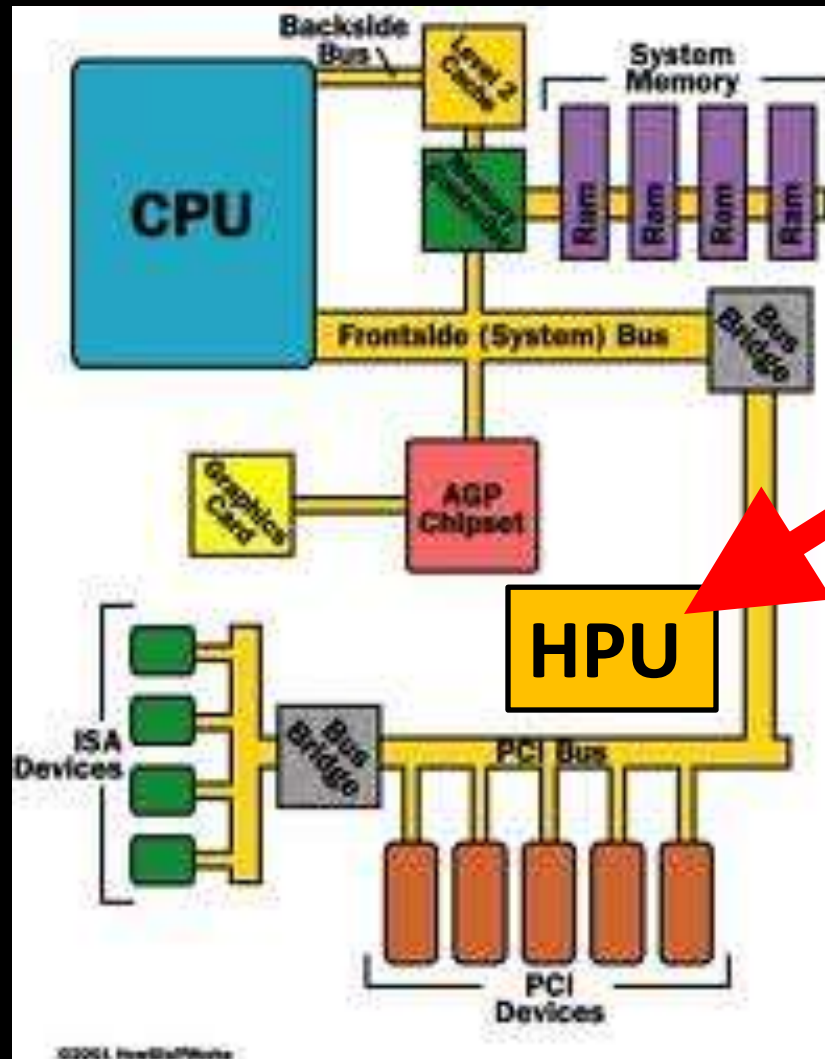


- What was old becomes new
- *“Crowdsourcing: A New Branch of Computer Science”*  
(March 29, 2011)



Princeton University Press, 2005

# Davis et al. (2010) *The HPU*.



# Human Computation

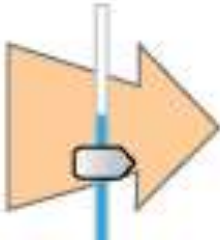
- People become ‘computists’ once more
  - Do tasks computers cannot (do well)
- 1. Detect robots (Captcha – “reverse Turing test”)
- 2. Micro-tasks and data labeling (at scale)
  - Game changer for improving practical AI: *starving for data*
- 3. Rethink what is possible in application design
  - Integrate CPU + HPU = new capabilities

# Blending Automation & Human Computation

# Soylent: A Word Processor with a Crowd Inside

- Bernstein et al., UIST 2010

**Shortn** action patterns  
olutions directly  
conceptual and  
vity. Authoring tools offer help with prag-  
thus present Soylent, a word processing  
e Find-Fix-Verify crowd programming pat-



This paper introduces architectural and  
plex endeavors that span many levels of  
other people. We thus present Soylent,  
ability, cost, wait time, and work time fo

**Crowdproof** they d  
use the software developed  
othing about programming r



let people be able to control  
'Be able to' is unnecessary.: let people  
allow people to control

**The Human Macro**

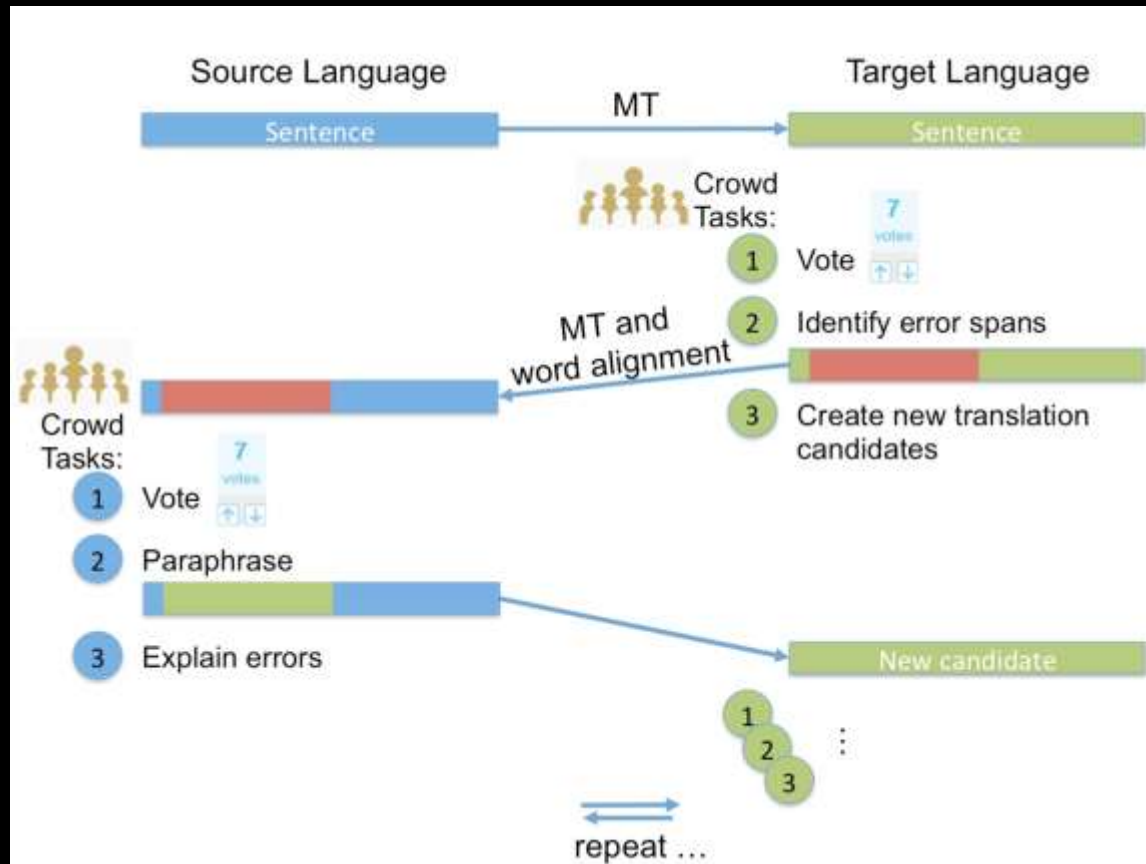
Write a request:  
Find Creative Commons figure for paragraph



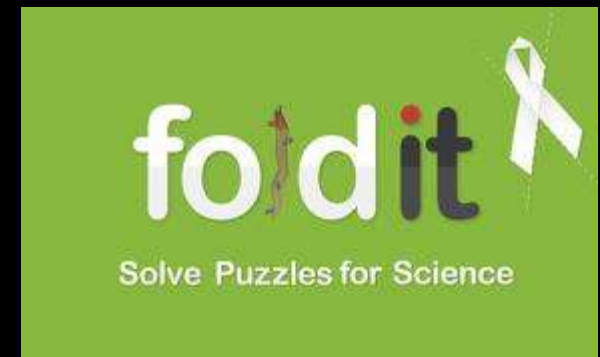
The paper introduces architectural and  
plex endeavors that span many levels of  
other people. We thus present Soylent,  
ability, cost, wait time, and work time fo

# Translation by monolingual speakers

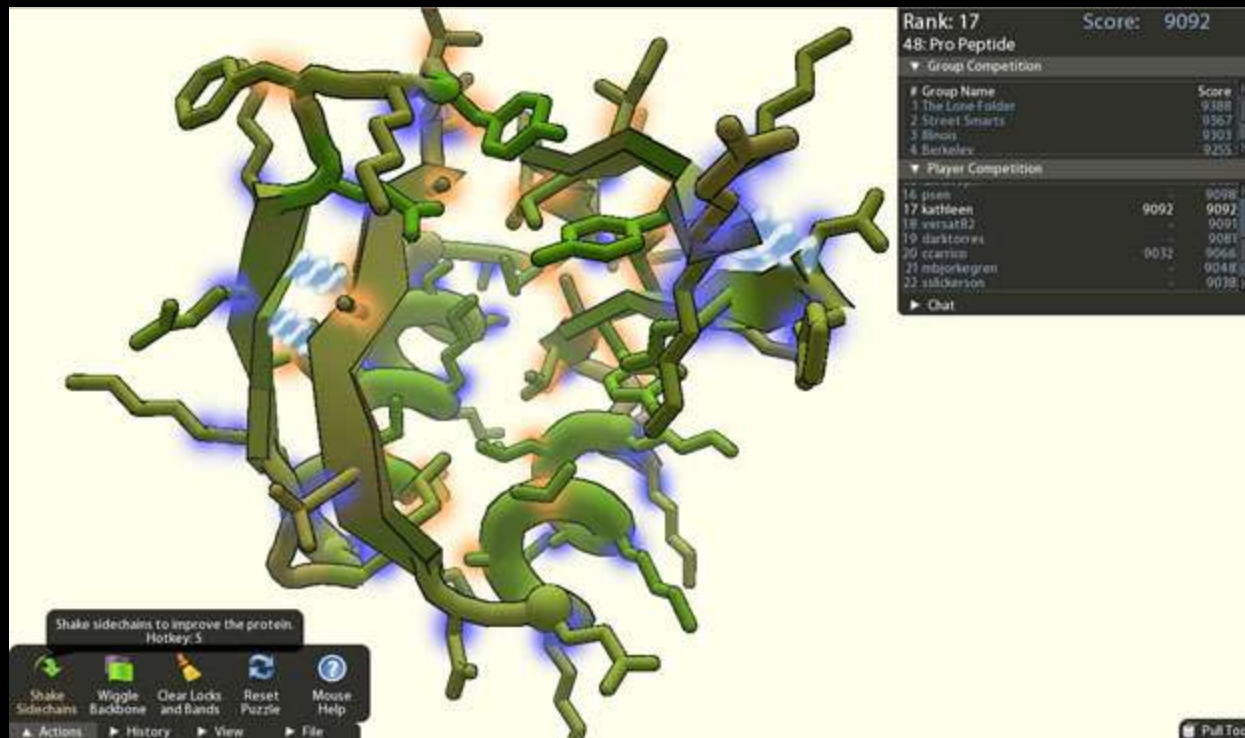
- C. Hu, CHI 2009



# fold.it



- S. Cooper et al. (2010).



Shake sidechains to improve the protein.  
Hotkey: S

Shake Sidechains Wiggle Backbone Clear Locks and Bands Reset Puzzle Mouse Help

Rank: 17 Score: 9092  
48: Pro Peptide

Group Competition	
#	Group Name
1	The Lone Folder
2	Street Smarts
3	Illmo
4	Berkalez

Player Competition		
#	Player Name	Score
16	pran	9088
17	kathleen	9092
18	versat82	9091
19	darktooves	9081
20	ccarion	9066
21	mbjorkgren	9048
22	siddarson	9038

► Chat

► Pull Tool

# Invisible By-product

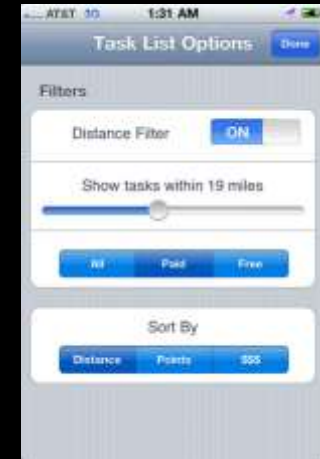


L. von Ahn et al. (2008). *recaptcha...* In *Science*.



# CrowdSearch and mCrowd

- T. Yan, MobiSys 2010





April 15, 2011

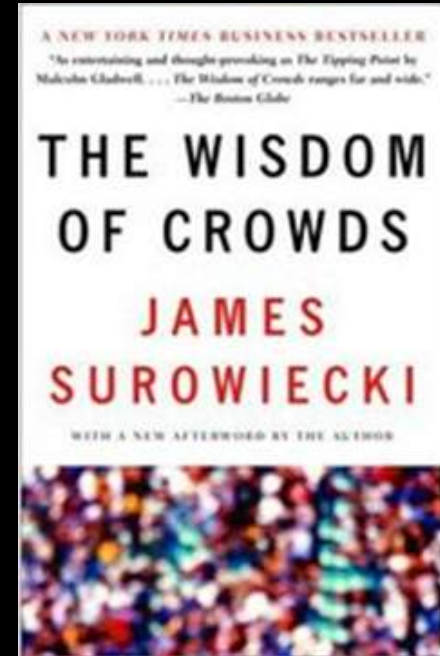
Matt Lease - ml@ischool.utexas.edu

26

# Wisdom of Crowds *Computing*

## Pre-conditions

- Diversity
- Independence
- Decentralization
- Aggregation



Input: large, diverse sample

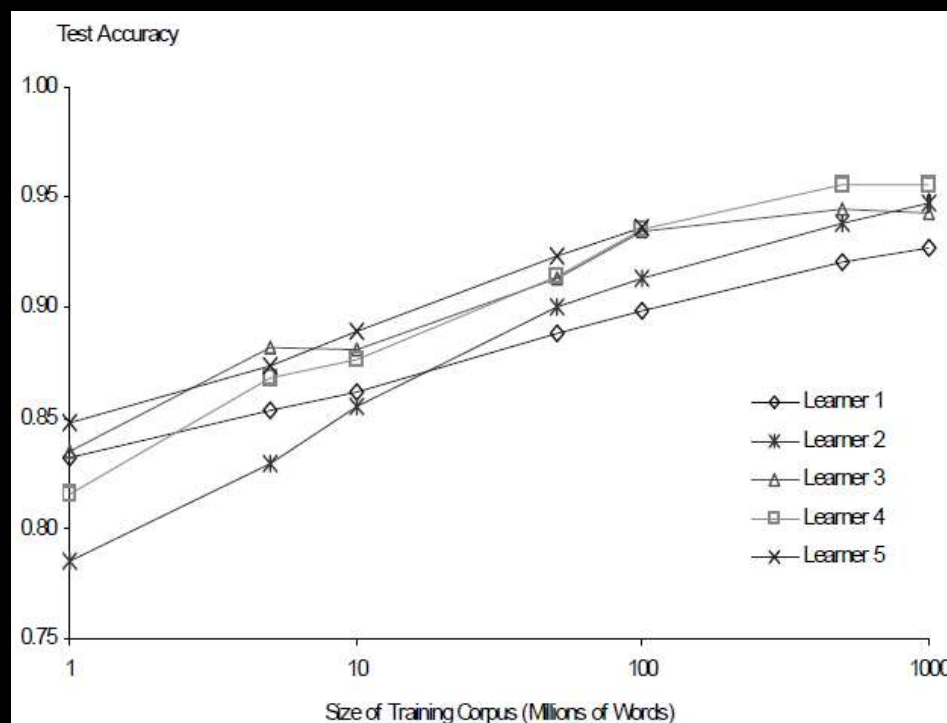
(increases likelihood of overall pool quality)

Output: consensus, selection, distribution

# Unreasonable Effectiveness of Data

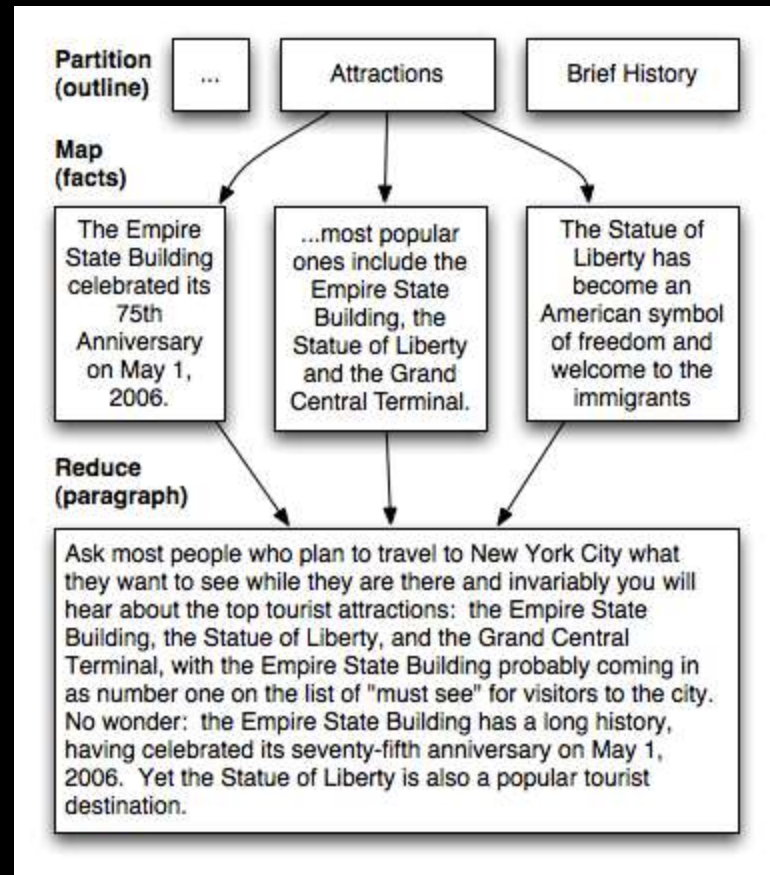
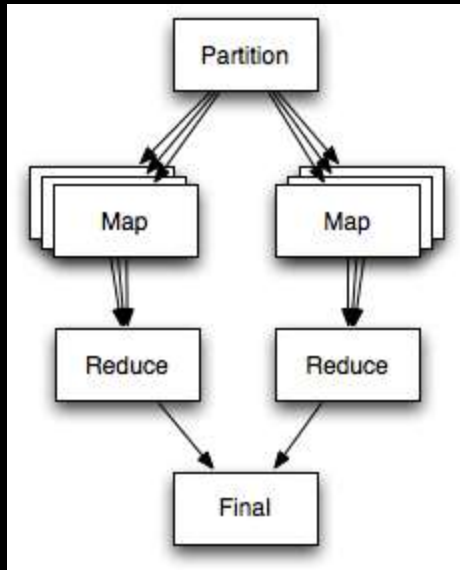
Massive free Web data changed how we train learning systems

- Banko and Brill (2001). Human Language Tech.
- Halevy et al. (2009). IEEE Intelligent Systems.



How might access to cheap & plentiful labeled data change the balance again?

# CrowdForge: MapReduce for Automation + Human Computation



# Wrap-up

# Conclusions

- Shift in practice that's here to stay
  - Fast, cheap, easy: collect data or do work
  - Emerging phenomenon to study and guide
- New capabilities in application design from automation + human computation
- Hot area, fast changing, many open problems

# Thank You!

- Students
  - Catherine Grady (iSchool)
  - Hyunjoon Jung (ECE)
  - Adriana Kovashka (CS)
  - Abhimanu Kumar (CS)
- Omar Alonso, Microsoft Bing
- Support
  - John P. Commons



[ir.ischool.utexas.edu/crowd](http://ir.ischool.utexas.edu/crowd)

[UT Mechanical Turk & Crowdsourcing Google Group](#)





# Resources & Upcoming Events

Special issue of Springer's Information Retrieval journal on Crowdsourcing (papers due May 6, 2011)

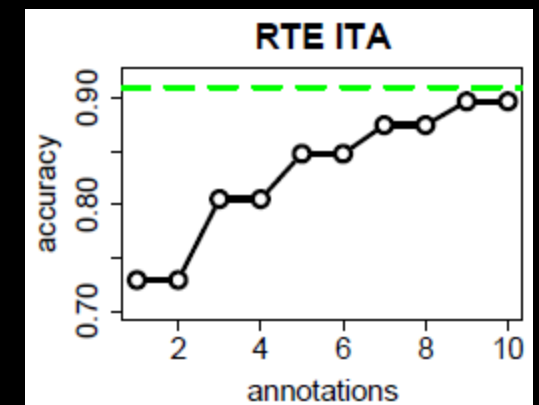
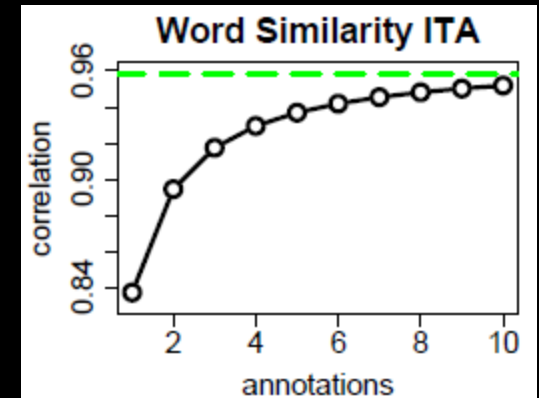
## Upcoming Conferences & Workshops

- 3<sup>rd</sup> HCOMP at AAI 2011
- CHI 2011 Workshop (proceedings online)
- SIGIR 2011 Workshop
- TREC 2011 Crowdsourcing Track
- CrowdConf 2011 (TBA)

# Data Collection

# Snow et al. (2008). EMNLP

- 5 Tasks
  - Affect recognition
  - Word similarity
  - Recognizing textual entailment
  - Event temporal ordering
  - Word sense disambiguation
- 22K labels for \$26
- high agreement between Turk annotations and expert “gold” labels



# Example – Dialect Identification

## Help Classify Arabic into Dialects!

This task is for Arabic speakers who understand the different local Arabic dialects (اللهجات المحلية، أو اللهجة)، and can distinguish them from *Fusha* Arabic (الصحى).

Below, you will see several Arabic sentences. For each sentence:

1. Tell us how much dialect (عربية) is in the sentence, and then
2. Tell us which Arabic dialect the writer intends.

This following map explains the dialects:



PLEASE READ the following. You MUST understand the classifications, otherwise your work might be rejected!!

- **Levantine** (لبناني) does NOT mean "Syrian" only. It includes Syrian, but ALSO: Jordanian is Levantine, Palestinian is Levantine, and Lebanese is Levantine. That's why all these countries are green in the map.
- **Maghrebi** (مغربي) does NOT mean "Moroccan" only. It includes Moroccan, but ALSO: Algerian is Maghrebi, Tunisian is Maghrebi, and Libyan is Maghrebi. That's why all these countries are purple in the map.
- The word "dialect" (لهجة) does NOT mean "spelling mistake" (خطأ إملائي). If the writer was trying to write in 100% صحى, classify it as No dialect, even if it has some spelling mistakes.

### Informed Consent Form

**Purpose of research study:** We are collecting human annotations to improve automatic translation of Arabic into other languages. These annotations might be class labels, judgments of output quality, or actual translations.

**Benefits:** Although it will not directly benefit you, this study may benefit society by improving how computers process human languages. This could lead to better translation software, improved web searching, or new user interfaces for computers and mobile devices.

**Risks:** There are no risks for participating in this study.

**Voluntary participation:** You may stop participating at any time without penalty by clicking on the "Return HIT" button, or closing your browser window.

**We may end your participation** if you do not have adequate knowledge of the language, or you are not following the instructions, or your answers significantly deviate from known translations.

**Confidentiality:** The only identifying information kept about you will be a WorkID serial number and your IP address. This information may be disclosed to other researchers.

**Questions/Concerns:** You may email questions to the principal investigator, [Chris Callison-Rosst](mailto:Chris.Callison-Rosst). If you feel you have been treated unfairly you may contact the Johns Hopkins University [Institutional Review Board](#).

Clicking on the "Accept HIT" button indicates that you understand the information in this consent form. You have not waived any legal rights you otherwise would have as a participant in a research study.

# Example – Spelling correction

## Evaluate a Spelling Correction for a Product Search Query

### Instructions

Imagine that a user is searching for products at an online shopping website. When the user searches for a term, the site suggests a spelling correction, such as "Did you mean: XYZ?". Evaluate whether this spelling correction is **GOOD** or **BAD**. If you aren't sure if the suggestion gives the proper spelling or are not familiar with the search terms, select **I DON'T KNOW**.

When evaluating corrections, ignore capitalization. All search terms and corrections are shown in lower case. A correction can be good even if a space is used instead of a hyphen. For example, "blu ray" and "blu-ray" are both good spelling corrections for "blue ray", even though the trademarked term is "Blu-ray".

Sample search results are provided for context. However, you should base your response on the accuracy of the spelling correction, not the relevance of the results.

**Note:** We pay bonuses for high-quality responses! You will earn a bonus if your answer is consistent with the majority of respondents. However, if you consistently disagree with the majority, you will be blocked from participating in our future experiments. (An answer is considered to be the majority response when it's selected by two-thirds or more of the workers who complete the HIT.)

[Instructions](#)

### Task

Please evaluate the following spelling correction, using the provided results for context:

User's search query: **enemax**



Photographic Print of Coloured X-ray of cancer of the colon from Science Photo Library (kitchen )  
productType: HOME\_FURNITURE\_AND\_DECOR  
productGroupID: gl\_kitchen  
Manufacturer: Science Photo Library  
superSaver: false  
numberReviews: 0  
averageRating: 0.0



Reports on Publications Issued and Registered in the Several Provinces of British India (Paperback )  
productType: ABIS\_BOOK  
productGroupID: gl\_book  
Author: Home Department, Government of India  
superSaver: true  
numberReviews: 0  
averageRating: 0.0  
fastTrack: true  
fastTrackEndDate:  
\$escapeUtils.unescapeHtml(\$highlighter.highlight(\$misspelling\_diff\_raw,\$engine.get('attributes').get(\$attrkey)))  
fastTrackGuaranteedDeliveryDate:  
\$escapeUtils.unescapeHtml(\$highlighter.highlight(\$misspelling\_diff\_raw,\$engine.get('attributes').get(\$attrkey)))  
listPrice: 18.99 GBP

Suggested correction: **enema**



Home **enema** kit: (Personal Care )  
productType: HEALTH\_PERSONAL\_CARE  
productGroupID: gl\_drugstore  
Manufacturer: Specialist Supplements Ltd.  
superSaver: false  
numberReviews: 1  
averageRating: 4.0



**Enema** Kit - 2 litre capacity for home and travel (Misc. )  
productType: BEAUTY  
productGroupID: gl\_beauty  
Manufacturer: Manifest Health Limited  
superSaver: false  
numberReviews: 0  
averageRating: 0.0

Is the correction of **enemax** to **enema** GOOD or BAD?

- GOOD: Yes, the suggested spelling correction corrects a misspelling.
- BAD: No, the suggested spelling correction is incorrect or unnecessary.
- I DON'T KNOW: Not sure if the suggested spelling correction gives the proper spelling, or not familiar with the search terms.

# Kovashka & Lease, CrowdConf'10



How similar is the artistic style in the paintings above?

- Very similar
- Somewhat similar
- Neither similar nor dissimilar
- Somewhat dissimilar
- Very dissimilar

# The Crowd



# Models & Incentives

- Pay (e.g. MTurk)
- Fun (or avoid boredom)
- Socialize
- Earn acclaim/prestige
- Altruism
- Learn something new (e.g. English)
- Invisible by-product (e.g. re-Captcha)
- Create self-serving resource (e.g. Wikipedia)

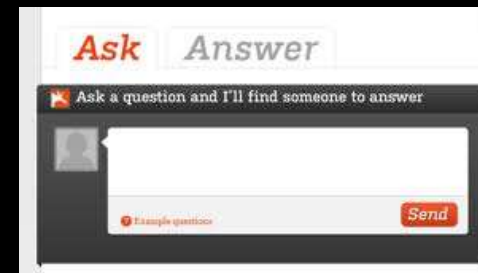
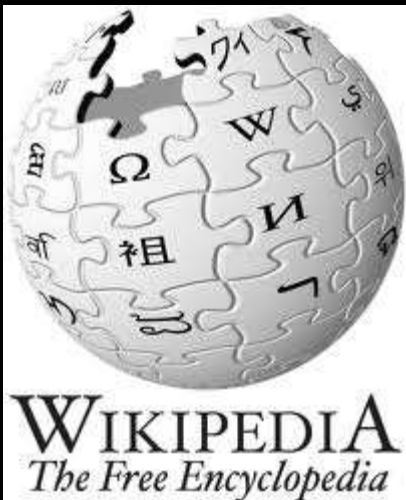


**MOTIVATION**  
SOMETIMES THERE JUST ISN'T ANY.

Multiple incentives are often offered in tandem

# Altruism

- Contribute knowledge
- Help others (who need knowledge)
- Help workers (e.g. *SamaSource*)
- Charity



# Games with a Purpose (L. von Ahn)

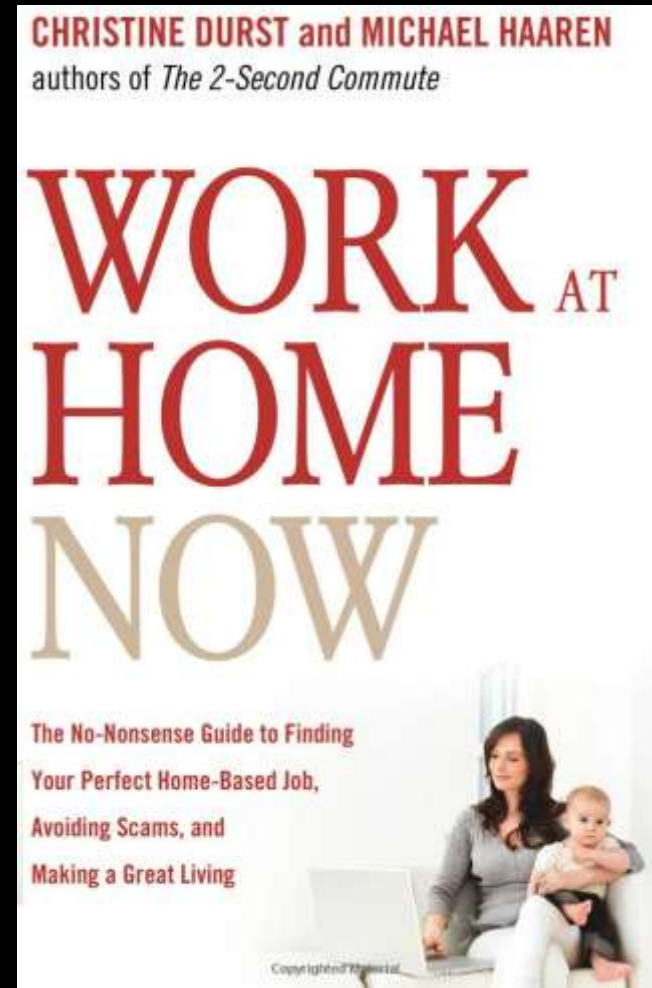
- Players have fun, creators get data as by-product



- distinct from *Serious Gaming / Edutainment*
  - Player learning / training / education is by-product

# Worker Demographics

- 2008-2009 studies found less global and diverse than previously thought
  - US
  - Female
  - Educated
  - Bored
  - Money is secondary



# 2010 shows increasing diversity

47% US, 34% India, 19% other (P. Ipeitorotis. March 2010)

