

Human-Powered Database Operations: Part 1



Dongwon Lee

Penn State University, USA

dongwon@psu.edu

Slide available @ <http://goo.gl/4pNUhB>

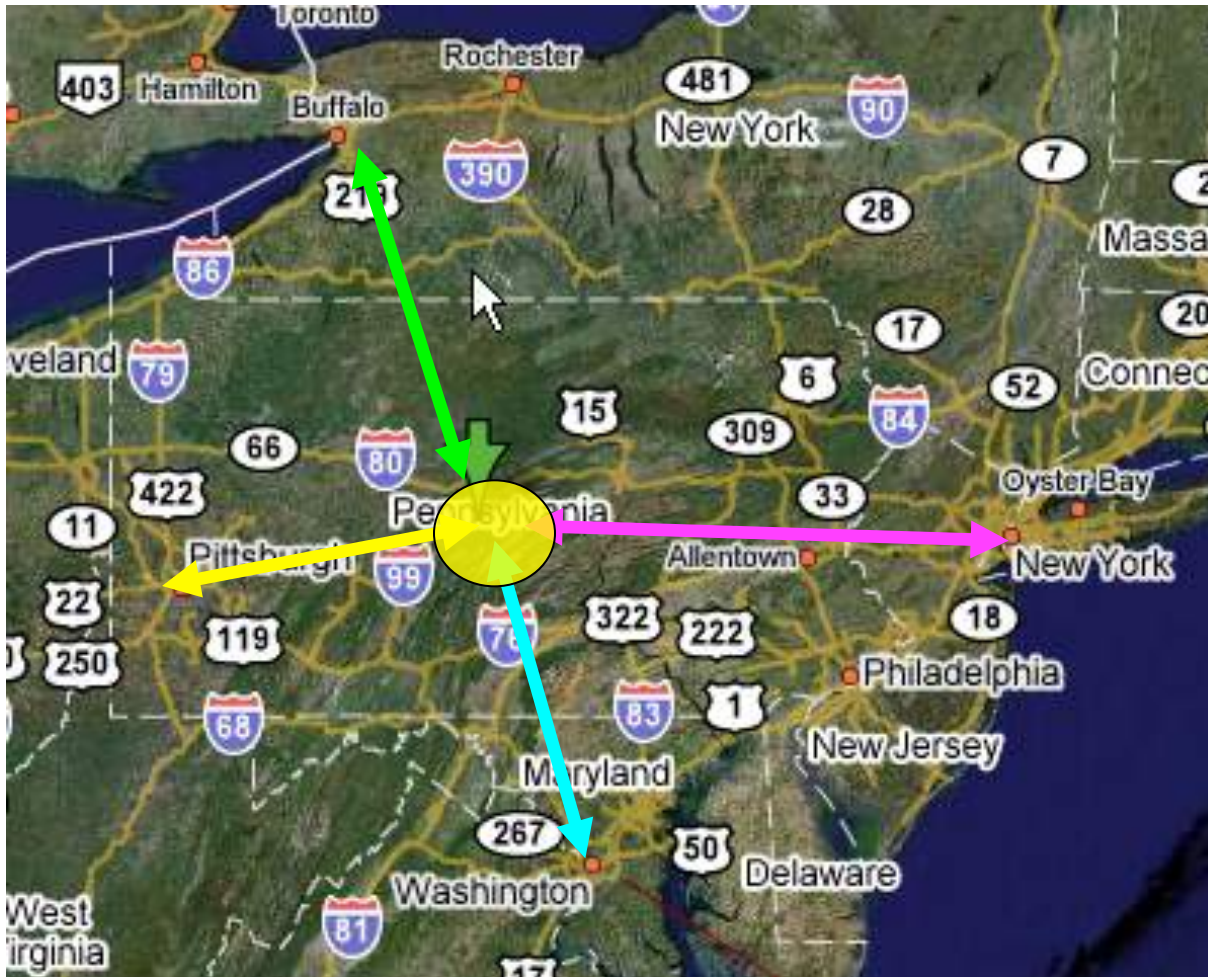
SBBD 2014 Tutorial



Where Am I From?



Penn State University



- State College, PA
 - Out of nowhere, but close to everywhere
- West: 2.5 hours to **Pittsburgh**
- East: 4 hours to **New York**
- South: 3 hours to **Washington DC**
- North: 3 hours to **Niagara Fall**

Penn State *i*-School

- College of Information Sciences and Technology (IST)
 - <http://ist.psu.edu/>
- 40+ tenure-track faculty on diverse areas
 - CompSci & EE
 - MIS & LIS
 - Design
 - Law
 - Psychology
 - Medical Infomatics



Other Tutorials on Crowdsourcing

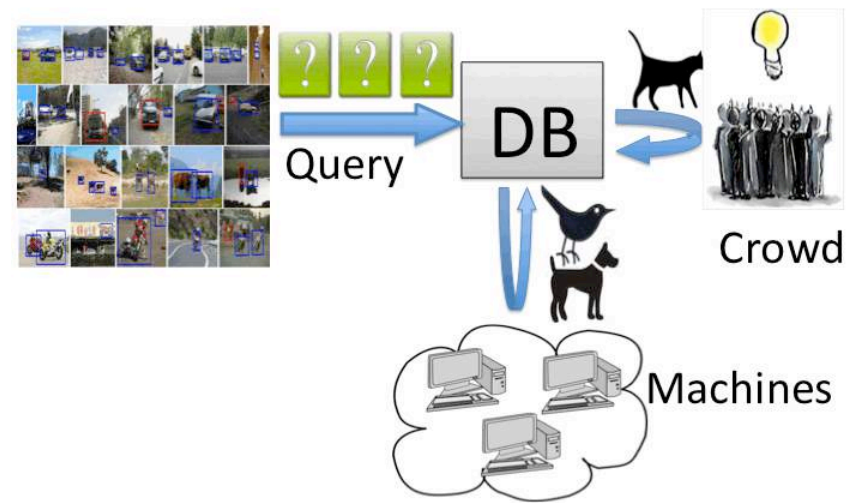
Year	Sub-field	Venue	Title
2013	Crowdsourcing	SDM	Crowdsourcing & Human Computation
2013	Crowdsourcing	HCOMP	Incentives in Human Computation
2012	HCI	AAAI	Crowdsourcing using MTurk for HCI Research
2012	Crowdsourcing	SBP	Crowdsourcing, Human Computation, and Collective Intelligence
2012	IR	SIGIR	Human Computation and Crowdsourcing
2012	DB	SIGMOD	Designing a Scalable Crowdsourcing Platform
2011	Crowdsourcing	AAAI	Human Computation: Core Research Questions and State of the Art
2011	IR	CLEF	Crowdsourcing for IR Experimentation and Evaluation
2011	ML	ICML	Collective Intelligence and Machine Learning
2011	Social Science	EC	Conducting Behavioral Research using AMT
2011	Multimedia	MM	Frontiers in Multimedia Search
2011	DB	VLDB	Crowdsourcing Application and Platforms
2011	Crowdsourcing	WWW	Managing Crowdsourced Human Computation
2010	Vision	CVPR	Mechanical Turk for Computer Vision
2008	IR	CIKM	Crowdsourcing for Relevance Evaluation

The Focus of This Tutorial

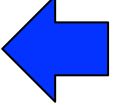
- Part 1 on **basics** of crowdsourcing
- Part 2 on **DB operations** that exploit crowdsourcing



When to Ask Whom

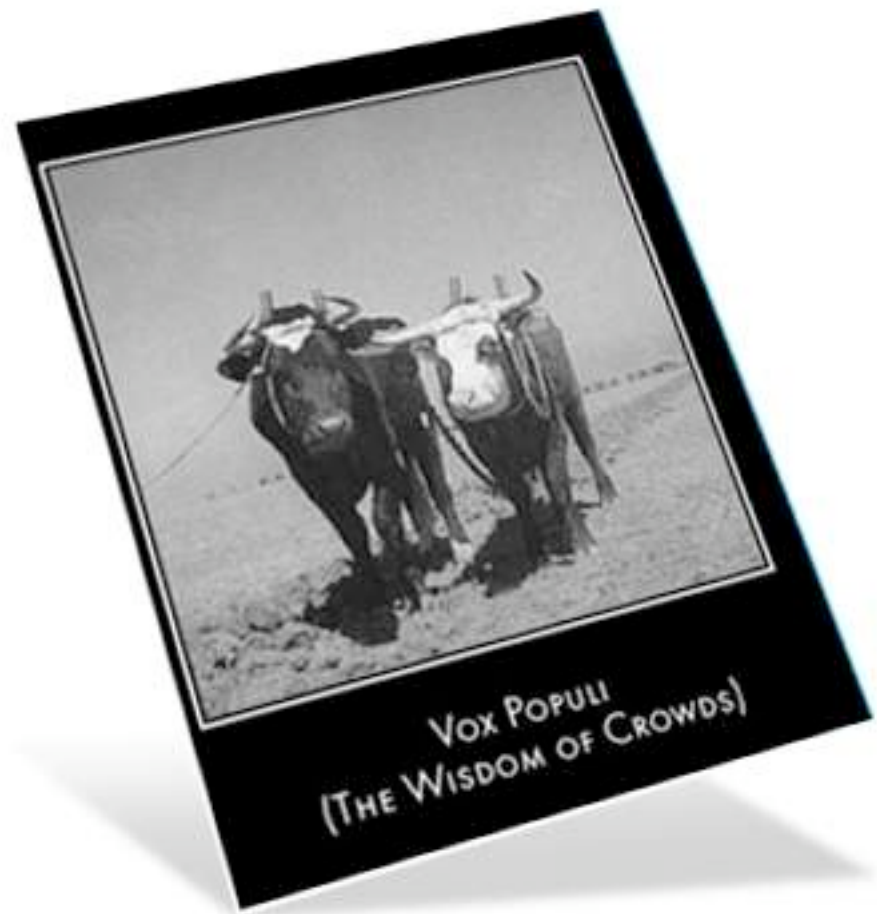
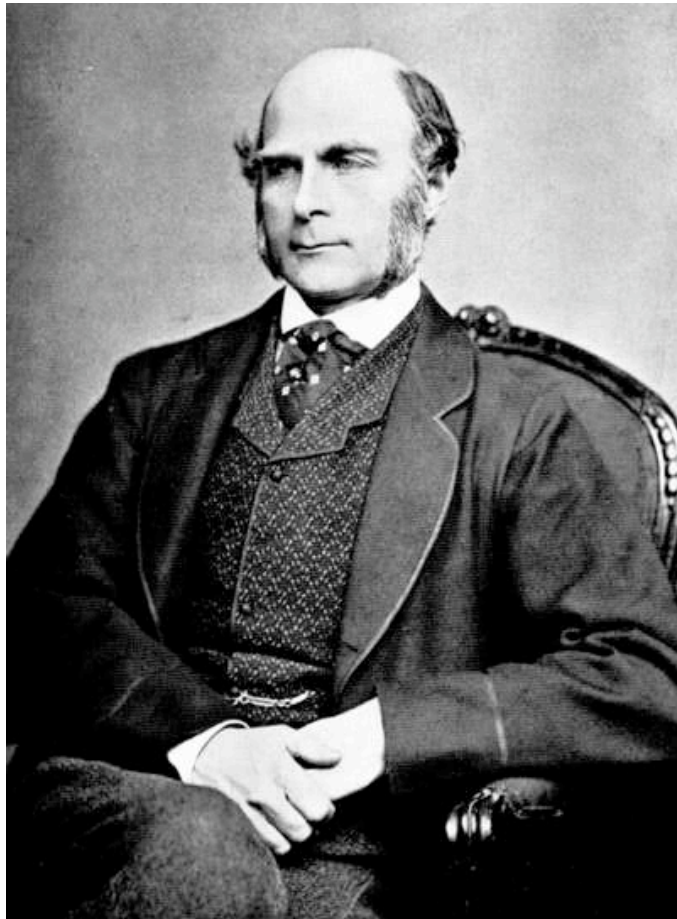


Part 1: Crowdsourcing Basics

- **Examples** 
- Definitions
- Marketplaces
- Computational Crowdsourcing
 - Preliminaries
 - Transcription
 - Sorting
- Demo

Eg, Francis Galton, 1906

**Weight-judging competition:
1,197 (mean of 787 crowds) vs. 1,198 pounds (actual measurement)**



Eg, StolenSidekick, 2006



9

- A woman lost a cellphone in a taxi
- A 16-year-old girl ended up having the phone
 - Refused to return the phone
- Evan Guttman, the woman's friend, sets up a blog site about the incident
 - <http://stolensidekick.blogspot.com/>
 - <http://www.evanwashere.com/StolenSidekick/>
 - Attracted a growing amount of attention → the story appeared in Digg main page → NY Times and CNN coverage → Crowds pressure on police ...
- NYPD arrested the girl and re-possessed the phone

Eg, Finding “Jim Gray”, 2007

COMMUNICATIONS OF THE ACM

HOME | CURRENT ISSUE | NEWS | BLOGS | OPINION | RESEARCH

Home / Magazine Archive / July 2011 (Vol. 54, No. 7) / Searching for Jim Gray: A Technical Overview / Full Text

CONTRIBUTED ARTICLES

Searching for Jim Gray: A Technical Overview

By Joseph M. Hellerstein, David L. Tennenhouse
Communications of the ACM, Vol. 54 No. 7, Pages 77-87
10.1145/1965724.1965744
[Comments](#)

VIEW AS:



SHARE:



- Loosely coupled teams of computer scientists developed software polytechniques with clean interfaces, decoupling data acquisition from analysis to enable use at a distance.

- The U.S. Coast Guard developed a system to aid search and rescue operations, an interesting potential research area for computer scientists.

- New open-source tools and techniques could help with group coordination, crowdsourced image acquisition, volume image processing, ocean drift modeling, and analysis of open-water satellite imagery.



Image A @ time t_1

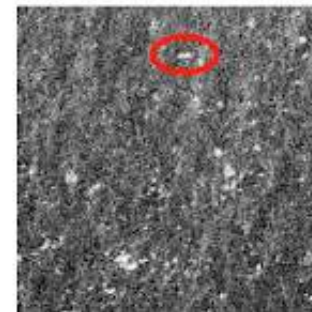


Image B @ time t_2

On Sunday, January 09, 2007, noted computer scientist Jim Gray

Eg, Threadless.com

- Sells t-shirts, **designed/voted** by crowds
- Artists whose designs are chosen get paid



The screenshot shows the Threadless.com website interface. At the top, there is a navigation bar with the Threadless logo, links for 'SHOP', 'PARTICIPATE', and 'BLOG', a search bar, and social media icons. The main banner features a woman wearing a green hoodie with a cartoon design of a monkey and a dinosaur. To her right, text reads: 'ALL THREADLESS HOODIES ARE \$29!' with a starburst indicating 'ENDS OCT 7 10am CT' and a 'SHOP NOW' button. Below the banner, a text block states: 'Threadless is a creative community that makes, supports, and buys great art.' At the bottom, there are four icons with corresponding text: 'Shop top sellers' (shopping bag icon), 'Submit your art' (pencil icon), 'Pick designs' (checkmark icon), and 'Meet artists' (ALUMNI icon).

Eg, KICKSTARTER

- Crowdfunding, started in 2009
- Project creators choose a deadline and a minimum funding goal
 - Creators only from US, UK, and Canada
- Donors pledge money to support projects, in exchange of non-monetary values
 - Eg, t-shirt, thank-u-note, dinner with creators
 - Donors can be from anywhere
- Eg, Pebble, smartwatch
 - 68K people pledged 10M



Eg, reCAPCHA

The Norwich line steamboat train, from New-London for Boston, this morning ran off the track seven miles north of New-London.

morning

morning overtakes

Type the two words:

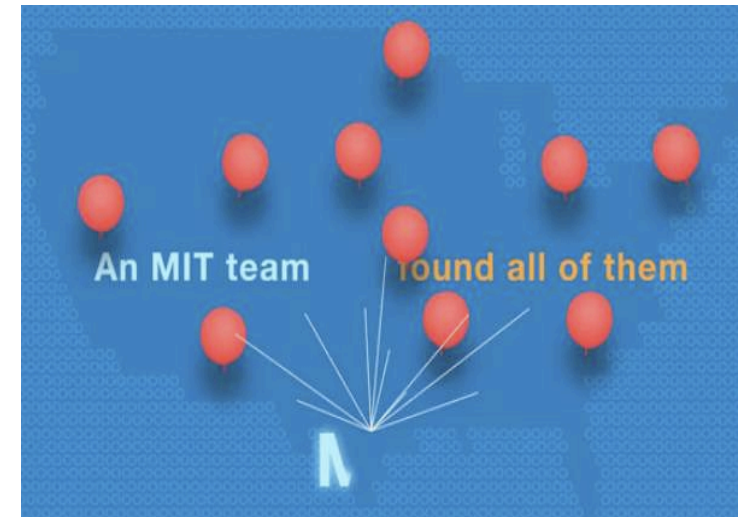
As of 2012

Captcha: 200M every day

ReCaptcha: 750M to date

Eg, DARPA Challenge, 2009

- To locate 10 red balloons in arbitrary locations of US
- Winner gets \$40K
- MIT team won the race with the strategy:
 - 2K per balloon to the first person, A, to send the correct coordinates
 - 1K to the person, B, who invited A
 - 0.5K to the person, C, who invited B, ...




Eg, Berkeley Mobile Millennium

mobile millennium

University of California, Berkeley


HOME
PROJECT
IN THE NEWS
GALLERY
ABOUT

Events

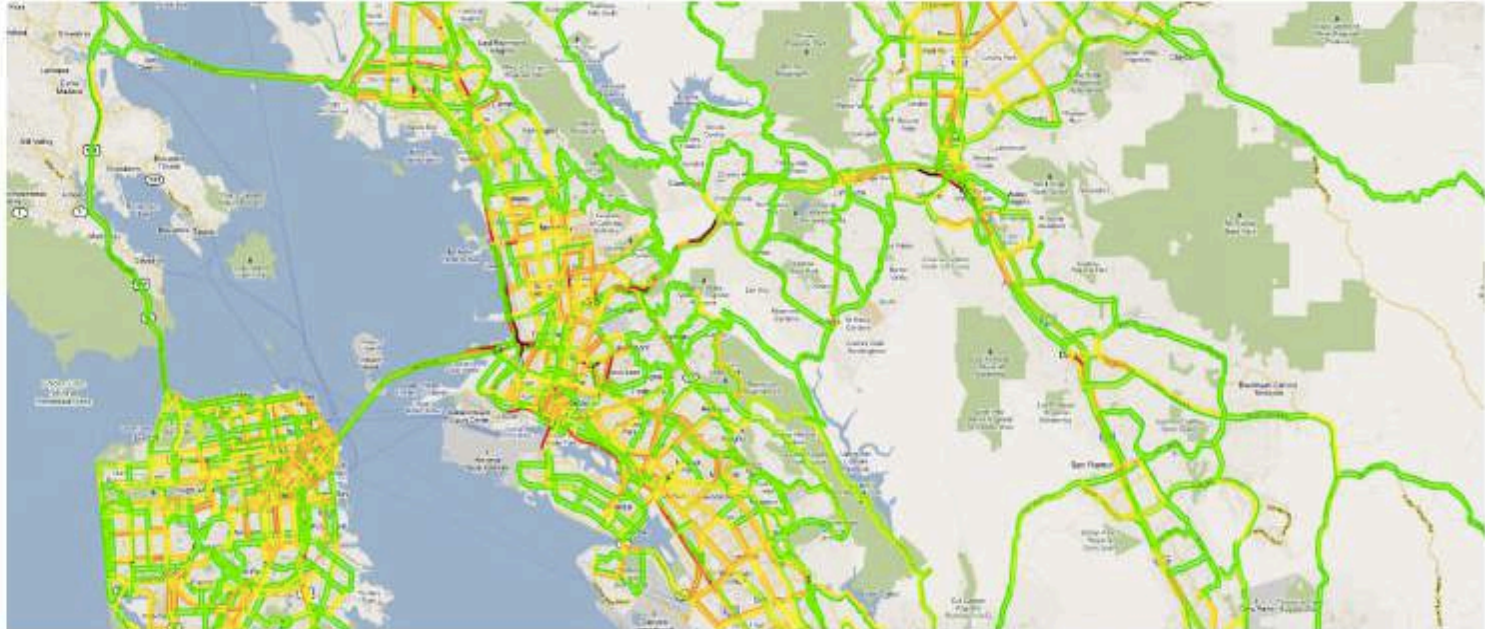


Mobile Millennium presented at the 2011 Transportation Research Board annual conference workshop on pervasive data, Washington DC, [see more](#)

Data Fusion



Snapshot of Mobile Millennium Traffic in San Francisco and the Bay Area



Project Overview

Mobile Millennium is a research project that includes a pilot traffic-monitoring system that uses the GPS in cellular phones to gather traffic information, process it, and distribute it back to the phones in real time.

Eg, Who Wants to be a Millionaire?



Asking the audience

usually works → Audience members have *diverse* knowledge that can be coordinated to provide a correct answer in sum

Eg, Who Wants to be a Millionaire?

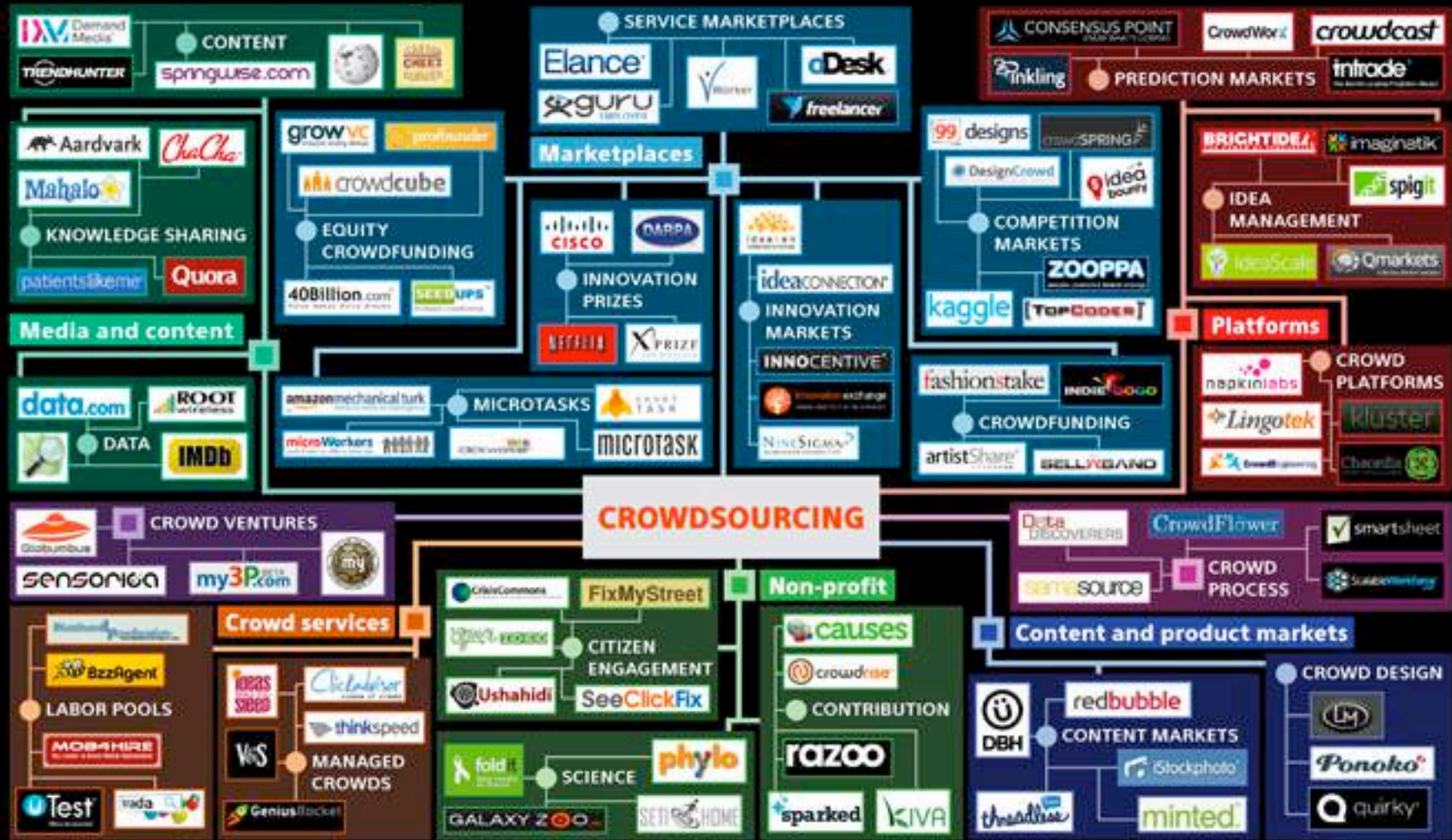
**Millionaire
Audience
BIGGEST
FAIL
EVER!!!!**

Eg, Game-With-A-Purpose: GWAP

- Term coined by Luis von Ahn @ CMU
- Eg,
 - ESP Game → Google Image Labeler: image recognition
 - Foldit: protein folding
 - Duolingo: language translation



Crowdsourcing landscape Beta v2



Excerpted from **Getting Results From Crowds** by Ross Dawson and Steve Bynghall

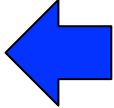
For definitions, analysis, free book chapters, and other crowdsourcing resources go to: www.resultsfromcrowds.com

Note: examples only, see website for full list of crowdsourcing services

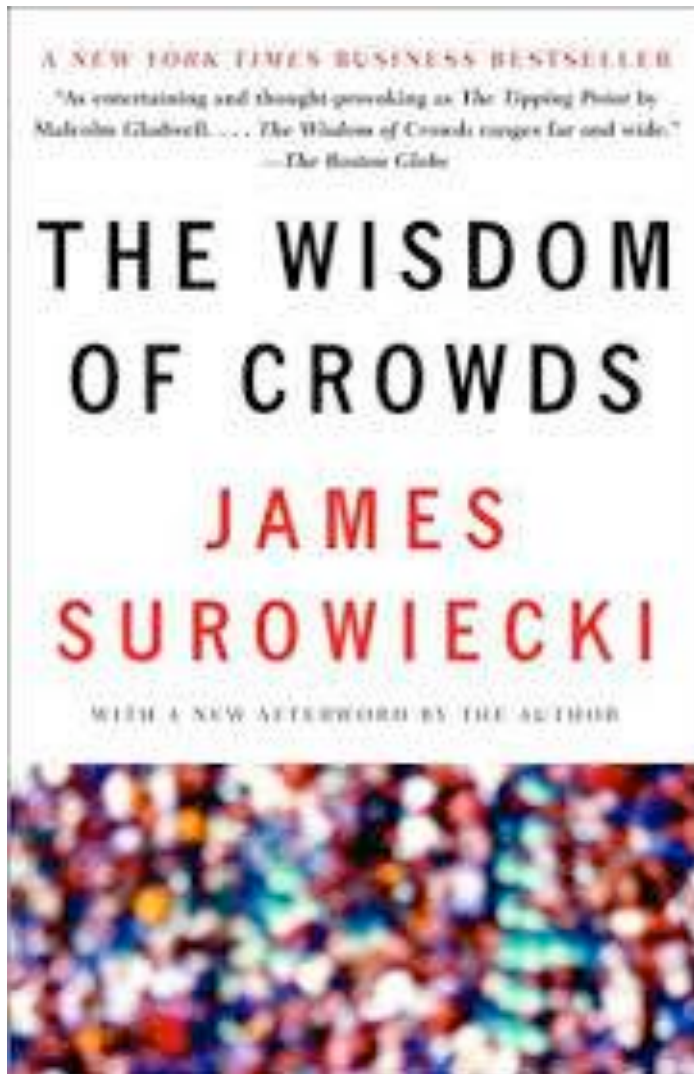


<http://www.resultsfromcrowds.com/features/crowdsourcing-landscape/>

Part 1: Crowdsourcing Basics

- Examples
- **Definitions** 
- Marketplaces
- Computational Crowdsourcing
 - Preliminaries
 - Transcription
 - Sorting
- Demo

James Surowiecki, 2004



“**Collective intelligence** can be brought to bear on a wide variety of problems, and complexity is no bar... conditions that are necessary for the crowd to be wise: *diversity*, *independence*, and ... *decentralization*”

Jeff Howe, WIRED, 2006



“**Crowdsourcing** represents the act of a company or institution 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. ... The crucial prerequisite is the use of the *open call* format and the *large* network of potential laborers...”

“Human Computation”, 2011

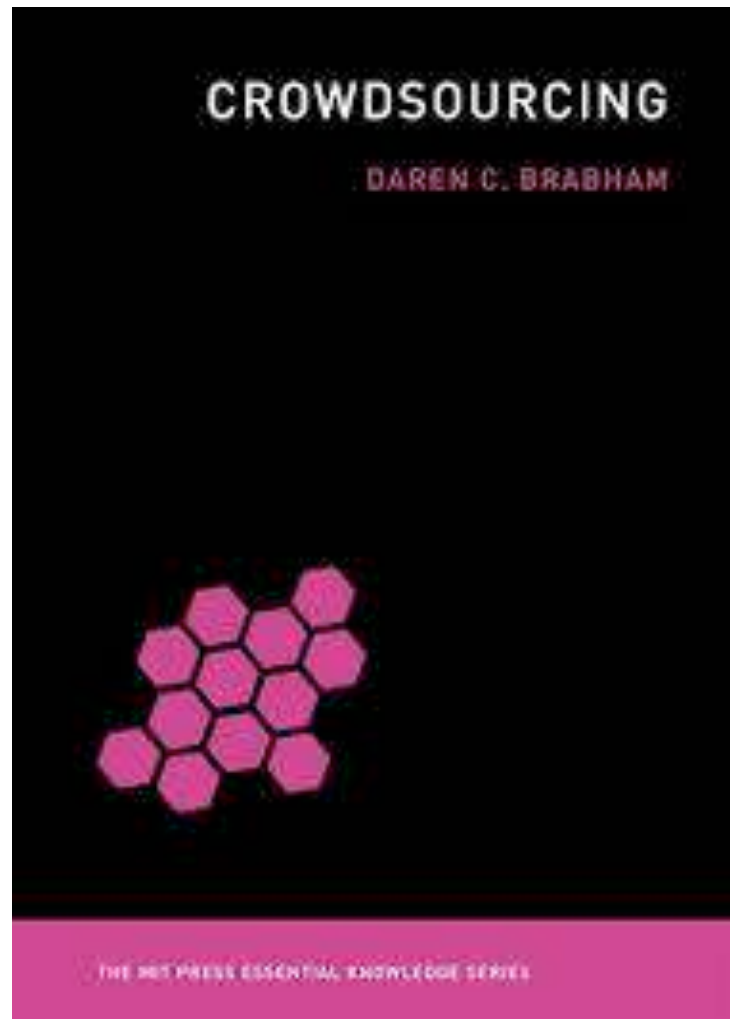


“**Human computation** is simply computation that is carried out by **humans**...

Crowdsourcing can be considered a method or a tool that human computation systems can use...”

By Edith Law & Luis von Ahn

Daren Brabhan, 2013



“**Crowdsourcing** as an *online, distributed* problem-solving and production model that leverages the collective intelligence of online communities to serve specific *organizational* goals... *top-down* and *bottom-up* ...”

What is Crowdsourcing?

- Many definitions
- A few characteristics
 - **Outsourced** to **human** workers
 - **Online** and **distributed**
 - Open call & right **incentive**
 - **Diversity** and **independence**
 - Top-down & bottom-up



What is **Computational Crowdsourcing**?

- Focus on computational aspect of crowdsourcing
 - Algorithmic aspect
 - Non-linear optimization problem
- Mainly use micro-tasks
- When to use Computational Crowdsourcing?
 1. Machine cannot do the task well
 2. Large crowds can probably do it well
 3. Task can be split to many micro-tasks

Part 1: Crowdsourcing Basics

- Examples
- Definitions
- **Marketplaces** ←
- Computational Crowdsourcing
 - Preliminaries
 - Transcription
 - Sorting
- Demo

Three Parties

- Requesters

- People submit some tasks
- Pay rewards to workers



Submit tasks



Collect answers

- Marketplaces

- Provide crowds with tasks



amazon mechanical turk™
Artificial Artificial Intelligence



Find tasks



Return answers

- Crowds

- Workers perform tasks



Notable Marketplaces

- Mechanical Turk



- CrowdFlower



- CloudCrowd



- Clickworker



- SamaSource



AMT: mturk.com

mechanical turk
Artificial Intelligence

Your Account

HITs

Qualifications

Introduction | Dashboard | Status | Account Settings

Mechanical Turk is a marketplace for work.

We give businesses and developers access to an on-demand, scalable workforce.

Workers

from thousands of tasks and work wh
200,645 HITs available. [View the](#)

Requesters

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. [Register Now](#)

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

AMT: Workers vs. Requesters

- Workers
 - Register w. credit account (only US workers can register as of 2013)
 - Bid to do tasks for earning money
- Requesters
 - First deposit money to account
 - Post tasks
 - Task can specify a qualification for workers
 - Gather results
 - Pay to workers if results are satisfactory

AMT: HIT

- Tasks
 - Called **HIT** (Human Intelligence Task)
 - **Micro**-task
- Eg
 - Data cleaning
 - Tagging / labeling
 - Sentiment analysis
 - Categorization
 - Surveying
 - Photo moderation
 - Transcription

Translate 3 lines from English to Russian (human translation needed).

Requester: Sergey Vasilyev **Reward:** \$0.05 per HIT **HITs Available:** 1 **Duration:** 15 minutes

Qualifications Required: HIT approval rate (%) is not less than 75

**Translate a text between the markers below from English to Russian.
Human translation only! Machine translations will be rejected.**

===== FROM HERE =====

Hello!
I am test text message to be translated from English to Russian.
If you ask me, I was born in a mind of a crazy web developer,
who tests the MTurk API to start a very promising service later.

===== TILL HERE =====

Any notes? Advices? Emotions? (Optional)

Translation task

Micro- vs. Macro-task: Eg, oDesk



Narrow results by:

CATEGORY

Software Development ▼

SUBCATEGORY

Game Development ▼

Workers



Rodrigo A.

Game Programmer (Desktop/Mobile/Web)

I'm a Computer Scientist who is always studying hard & developing softwares and games with a big passion. I've and apps using Java for the last 5 years and since ...

game-development

unity-3d

corona-sdk

unreal

\$31.11 HOURLY RATE | **4.8** ★★★★★ | **2 HOURS** | **9 BRAZIL**
9 PORTFOLIO ITEMS | **6** TESTS

ASSOCIATED WITH: [R2MGames](#)

Game Jobs

1,549 were found based on your criteria

Requesters

CATEGORIES

All Categories

Web Development (177)

Software Development (710)

Networking & Information Systems (8)

Writing & Translation (134)

Administrative Support (48)

Design & Multimedia (366)

Customer Service (10)

Sales & Marketing (76)

Business Services (20)

DATE POSTED

Any Timeframe

Last 24 Hours (50)

Last 3 Days (117)

Newest Jobs First ▼

1

Flash Game Interface Design

Hourly – 1 to 3 months – Less than 10 hrs/week – Posted 4 hours, 27 minutes ago

I'm looking for an artist experienced in designing flash game interfaces (Interface). I'm looking for someone who can provide a consistent style, have their own ideas, and be around in the coming months as needed. I need someone experienced in working with game developers in the past who have some understanding of the process of providing assets, either in psd, or directly in an .fla file, not some cluster of images in a folder. This could be a long-term project.

adobe-flash

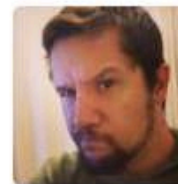
game-design

graphic-design

ui-design

Facebook App Development (Available NOW!!!)

Hourly – Less than 1 week – 30+ hrs/week – Posted 5 hours, 8 minutes ago



Gustavo V.

3D Generalist / Senior Game Programmer

I have a 15+ years of experience on programming using PHP. I have worked with Unity3D and other game engines on mobile devices and other platforms. I have also worked with ...

unity-3d

adobe-photoshop

blender3d

pixologic

AMT: HIT List

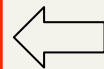
All HITS | **HITS Available To You** | HITS Assigned To YouFind **HITS** containingthat pay at least \$ **0.00** for which you are qualified require Master Qualification **GO**

All HITS

1-10 of 4372 Results

Sort by: **HITS Available (most first)** **GO!**[Show all details](#) | [Hide all details](#)1 2 3 4 5 > [Next](#) >> [Last](#)

Inv B 2	Requester: rohzit0d	HIT Expiration Date: Oct 25, 2013 (3 weeks 1 day)	Reward: \$0.00	View a HIT in this group
		Time Allotted: 48 minutes	HITS Available: 19606	
		The amount of time you have to complete the HIT, from the moment you accept it.		
Extract purchased items from a shopping receipt	Requester: Jon Breliq	HIT Expiration Date: Oct 10, 2013 (6 days 23 hours)	Reward: \$0.06	View a HIT in this group
		Time Allotted: 2 hours	HITS Available: 11518	
Você consegue encontrar o número de telefone ou endereço providenciados neste site?	Requester: CrowdFlower	HIT Expiration Date: Oct 8, 2013 (5 days 4 hours)	Reward: \$0.07	View a HIT in this group
		Time Allotted: 30 minutes	HITS Available: 7503	
Categorize: Businesses (US, Level III)	Requester: CrowdSource	HIT Expiration Date: Oct 3, 2014 (52 weeks)	Reward: \$0.12	View a HIT in this group
Can You Find the Provided Phone Number or Street Address on this Website?	Requester: CrowdFlower	HIT Expiration Date: Oct 9, 2013 (6 days 3 hours)	Reward:	
		Time Allotted: 30 minutes	HITS Available:	View a HIT in this group
Write Titles for Buying Guide	Requester: CrowdFlower	Description: <h3>Overview</h3>		
		Keywords: amazon , builders , delores , labs , crowd , flower , crowdfower , doloreslabs , deloreslabs , delores , address , business , verification , research , in		
	Qualifications Required:			
	Location is not VN			
	Location is not TR			
	Location is not RO			
	Location is not PK			
	Location is not PH			
	Location is not IN			
	Location is not ID			
	Location is not HK			
	HIT approval rate (%) is greater than 96			



Workers qualification

AMT: HIT Example

Can You Find the Provided Phone Number or Street Address on this Website?

Instructions ▾

Overview

In this task, you'll be provided a web page for a business, including its name, address, and phone number. Your goal is to answer a few questions about the business on the web page.

IMPORTANT: Sometimes the business will have multiple locations, and you will have to search the website for the specific business that we provide in order to verify the website.

Step by step instructions:

- Click the link to go to the provided site.
- First, please tell us whether or not the **name** of the business on the provided website is a **close** or **identical** match to the name of the business shown at the top of the page.
- Next, please tell us whether the provided business has
- Please be sure to click the appropriate option if the site

Wrinkles Day Spa

Phone: **+61893455333**
 Street: **Shop 5a Stirling Central Shopping Centre, 478 Wanneroo Rd**
 City: **Westminster**
 State: **WA**
 Postalcode (Zip): **6061**
 Country Code: **AU**

Click here to visit the website.

Is the name of the business on the web page similar or identical to 'Wrinkles Day Spa'?

- Yes: the name of the business is *similar* to *Wrinkles Day Spa*
 Yes: the name of the business is *nearly identical* to *Wrinkles Day Spa*
 No: the name is very different from *Wrinkles Day Spa*

? For the first option, the street **number** does not need to match, just the street, **Shop 5a Stirling Central Shopping Centre, 4**

AMT: HIT Example

Timer: 00:00:00 of 60 minutes

Want to work on this HIT?

Want to see other HITs?

Accept HIT

Skip HIT

Total Earned: Unavailable
Total HITs Submitted: 0

Store name, date, time, total, location on this receipt

Requester: Vishwanath Kumar

Reward: \$0.03 per HIT

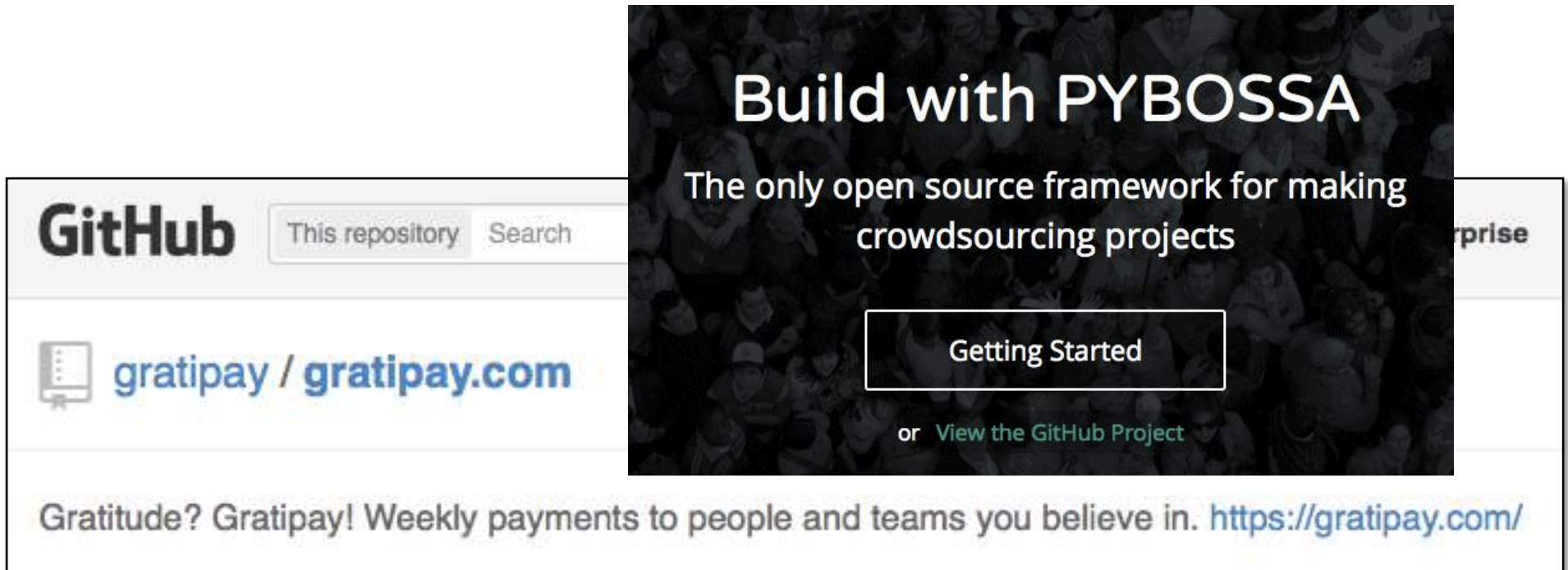
HITs Available: 71477

Duration: 60 minutes


Qualifications Required: Total approved HITs is greater than 1000



Open-Source Marketplace S/W



GitHub This repository Search

 [gratipay / gratipay.com](#)

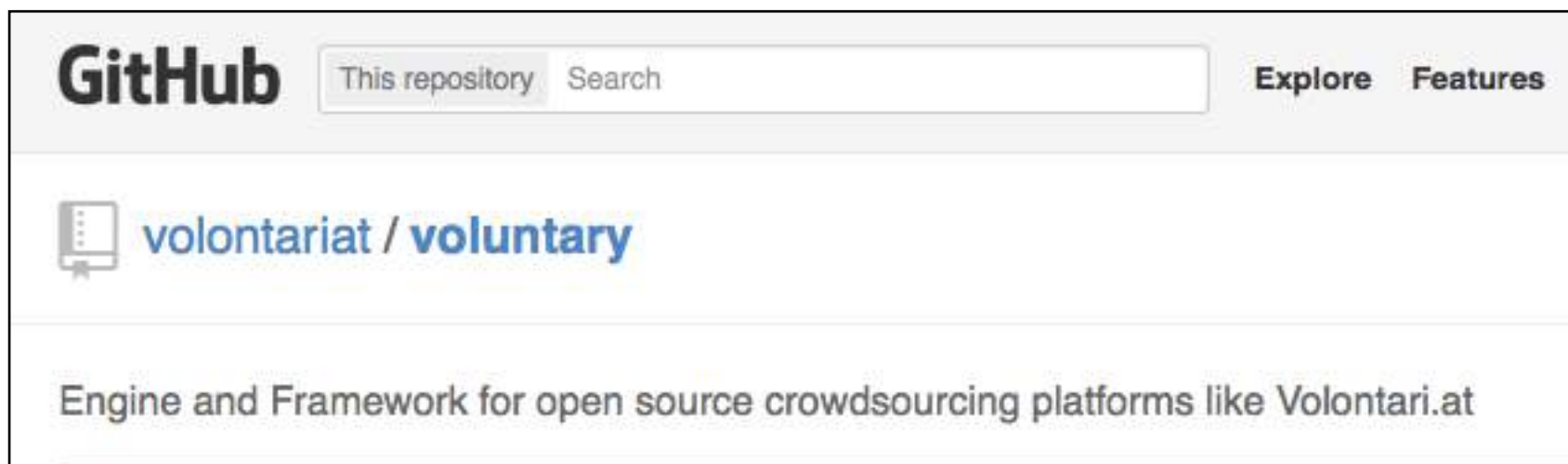
Gratitude? Gratipay! Weekly payments to people and teams you believe in. <https://gratipay.com/>

Build with PYBOSSA
The only open source framework for making crowdsourcing projects


Getting Started

or [View the GitHub Project](#)

prise



GitHub This repository Search Explore Features

 [volontariat / voluntary](#)

Engine and Framework for open source crowdsourcing platforms like Volontari.at

Part 1: Crowdsourcing Basics

- Examples
- Definitions
- Marketplaces
- **Computational Crowdsourcing**
 - **Preliminaries** ←
 - Transcription
 - Sorting
- Demo

Three Computational Factors

- **Latency (or execution time)**

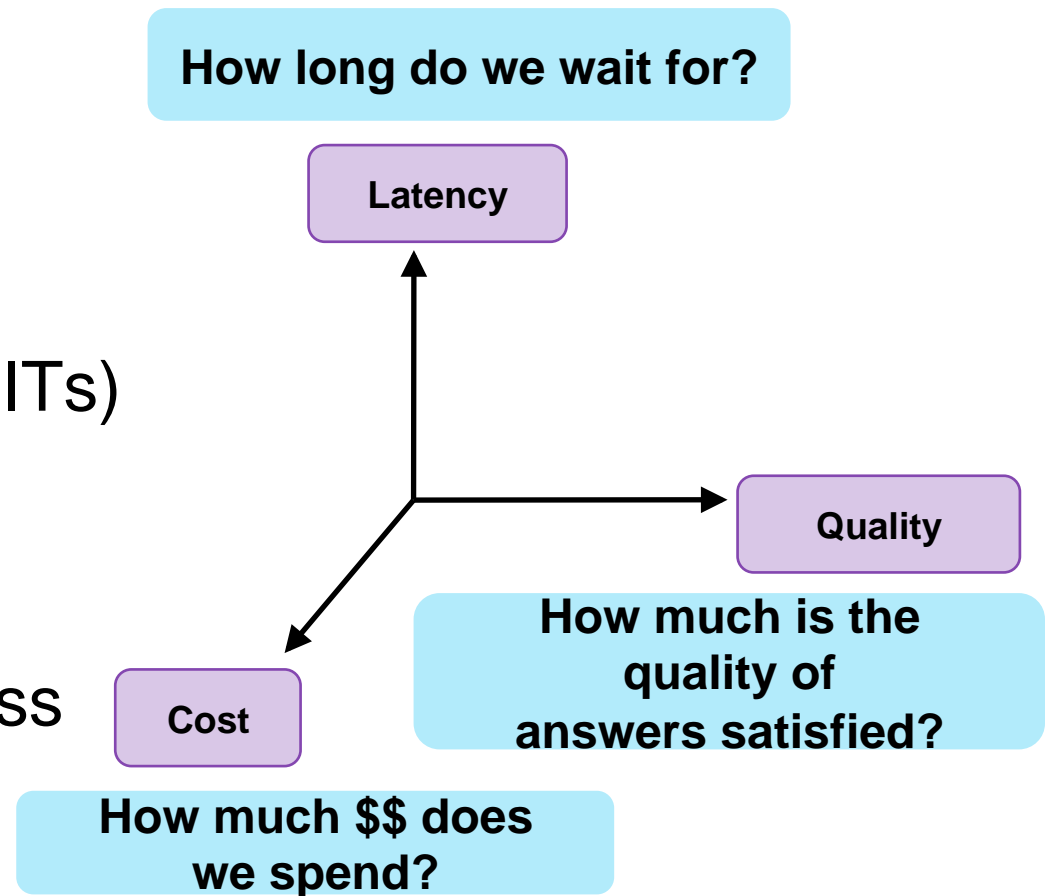
- Worker pool size
- Job attractiveness

- **Monetary cost**

- Cost per question
- # of questions (ie, HITs)
- # of workers

- **Quality of answers**

- Worker maliciousness
- Worker skills
- Task difficulty



#1: Latency

- Some crowdsourcing tasks finish faster than others
 - Eg, easier, or more rewarding tasks are popular
- Dependency among tasks

This is a password-protected HIT for a particular worker.		View a HIT in this group
Requester: Eric DeRosia	HIT Expiration Date: Oct 7, 2016 (104 weeks 3 days)	Reward: \$0.00
Time Allotted: 24 hours		HITs Available: 1
Faculty Development		View a HIT in this group
Requester: Kevin Dodds	HIT Expiration Date: Oct 28, 2014 (3 weeks)	Reward: \$0.20
Time Allotted: 15 minutes		HITs Available: 16
Rate an online article (required screening test)		View a HIT in this group
Requester: HubPages	HIT Expiration Date: Oct 7, 2014 (12 hours 10 minutes)	Reward: \$0.15
Time Allotted: 5 days		HITs Available: 15

#2: Cost

- Cost per question
- # of HITs

Remaining cost to pay:
 $\$0.03 \times 2075 = \62.25



<u>Image Keyword Verification</u>		View a HIT in this group	
Requester: Corbis Holdings, Inc	HIT Expiration Date: Oct 13, 2014 (6 days 19 hours)	Reward:	\$0.03
Time Allotted:	15 minutes	HITs Available:	2075
<u>Enter information about a forum discussion thread in which a vehicle is being built, rebuilt, or restored</u>		View a HIT in this group	
Requester: Jonathan R	HIT Expiration Date: Oct 14, 2014 (7 days 20 hours)	Reward:	\$0.20
Time Allotted:	30 minutes	HITs Available:	2000
<u>Transcribe up to 25 Seconds of Media to Text - Low Priority</u>		View a HIT in this group	
Requester: Crowdsurf Support	HIT Expiration Date: Oct 20, 2014 (2 weeks)	Reward:	\$0.08
Time Allotted:	15 minutes	HITs Available:	1836
<u>PADs US consumables 20140824-Thu Sep 11 16:15:50 PDT 2014</u>		View a HIT in this group	
Requester: Amazon Requester Inc.	HIT Expiration Date: Oct 11, 2014 (5 days)	Reward:	\$0.00
Time Allotted:	1 hour 46 minutes	HITs Available:	1605

#3: Quality of Answers

- Avoid spam workers
- Use workers with reputation

Store name, date, time, total, location on this receipt

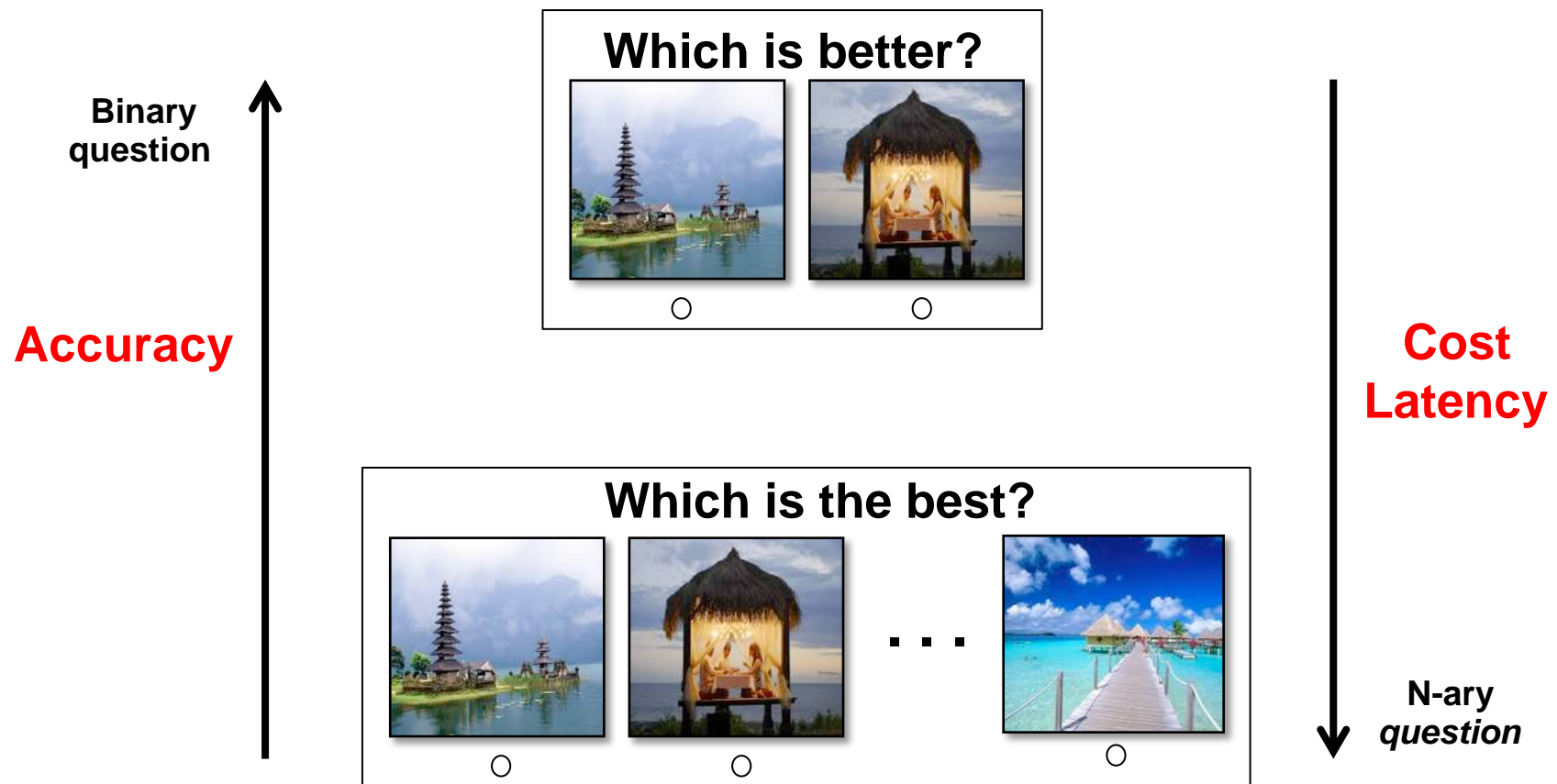
Requester: Vishwanath Kumar **Reward:** \$0.03 per HIT **HITs Available:** 71477 **Duration:** 60 minutes

Qualifications Required: Total approved HITs is greater than 1000

- Ask the **same** question to multiple workers to get consensus (eg, majority voting)
- Assign **more number** of (better-skilled) workers to **more difficult** questions

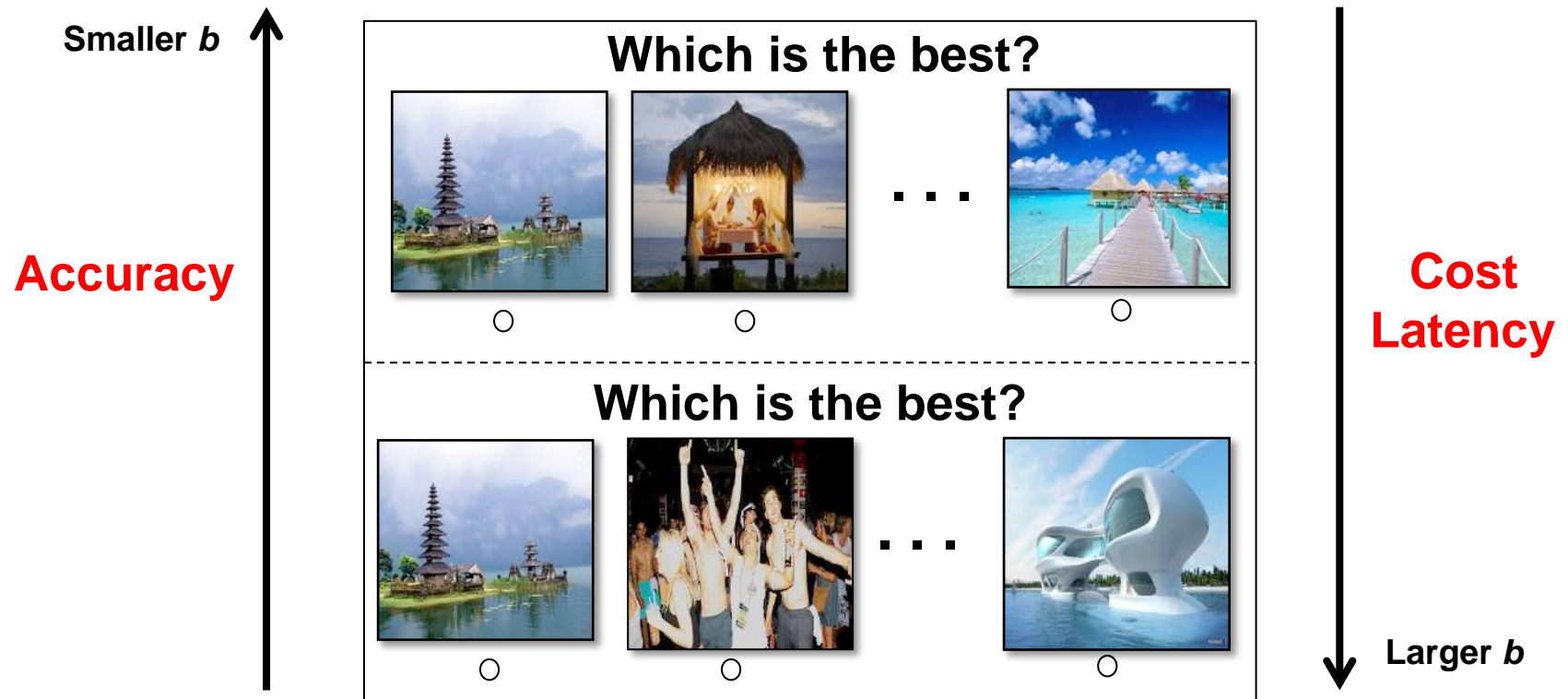
Size of Comparison

- Diverse forms of questions in a HIT
- Different sizes of comparisons in a question



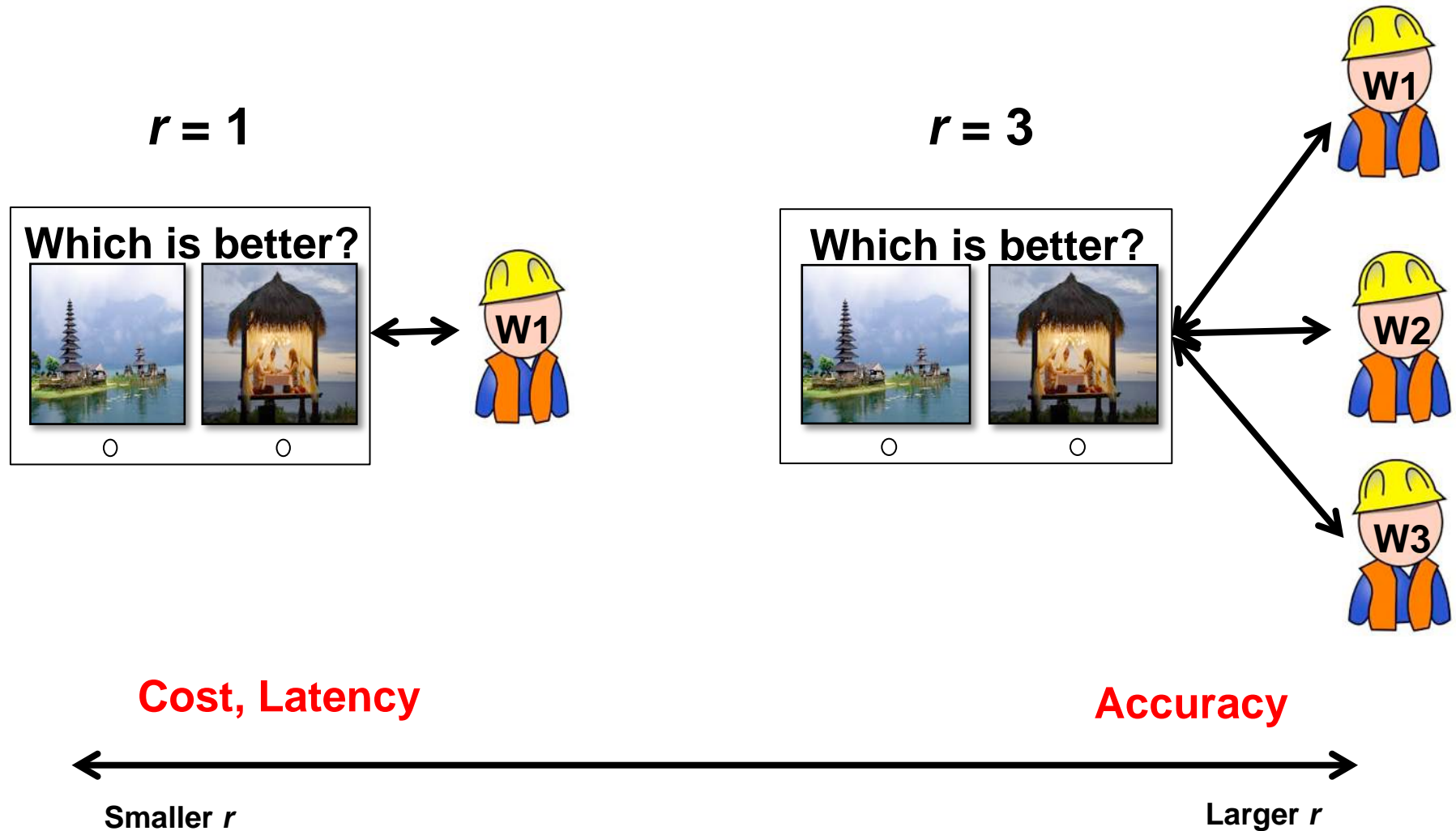
Size of Batch

- Repetitions of questions within a HIT
- Eg, two n -ary questions (batch factor $b=2$)



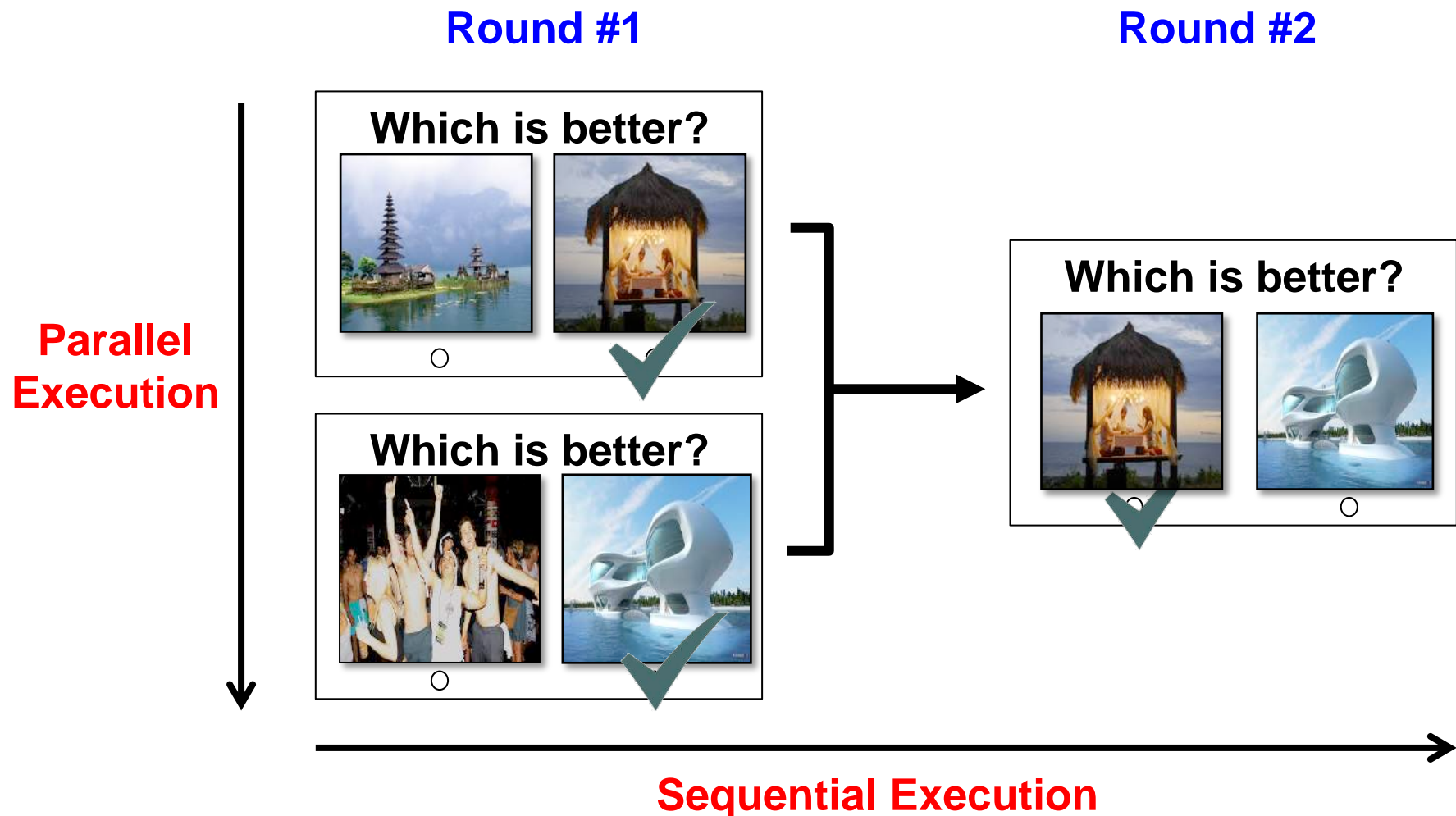
Response (r)

- # of human responses seeked for a HIT



Round (= Step)

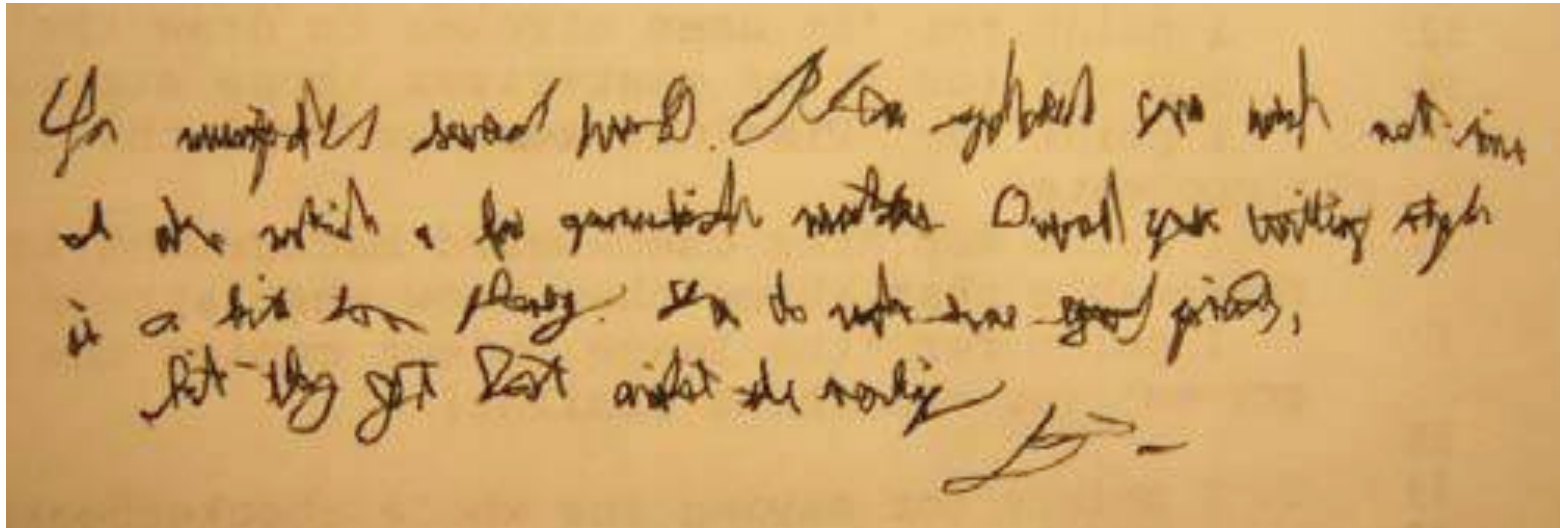
- Algorithms are executed in rounds
- # of rounds \approx **latency**



Part 1: Crowdsourcing Basics

- Examples
- Definitions
- Marketplaces
- **Computational Crowdsourcing**
 - Preliminaries
 - **Transcription** ←
 - Sorting
- Demo

Eg, Text Transcription [Miller-13]

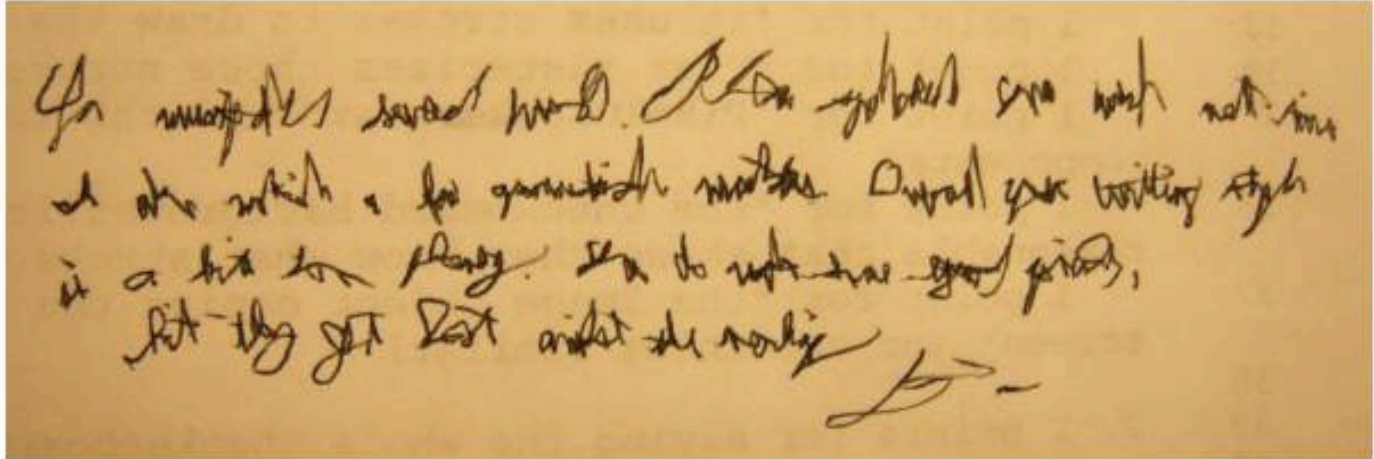


- **Problem:** one person cannot do a good transcription
- **Key idea:** iterative improvement by many workers

Greg Little *et al.* "Exploring iterative and parallel human computation processes." HCOMP 2010

Eg, Text Transcription [Miller-13]

Handwriting Recognition Task - Mozilla Firefox



- Please improve the transcription of this handwriting.
- People will vote whether to approve your changes.

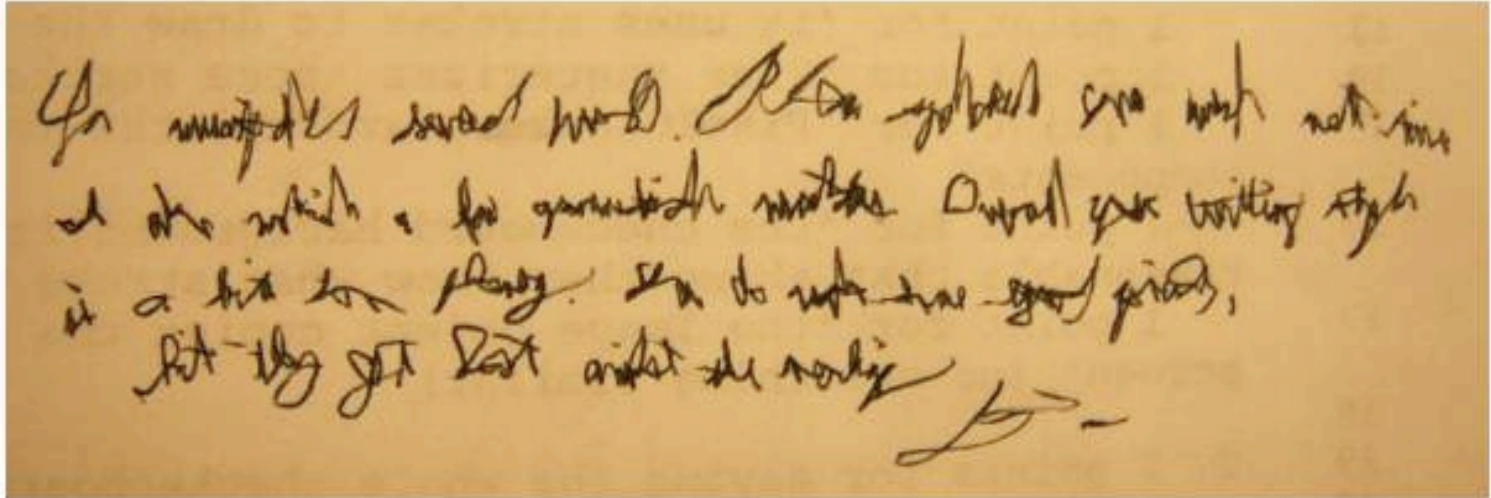
You (?) (?) (?) (work). (?) (?) (?) work (not) (time). I (?) (?) a few grammatical mistakes. Overall your writing style is a bit too (phoney). You do (?) have good (points), but they got lost amidst the (writing). (signature)

improvement \$0.05

Submit

Eg, Text Transcription [Miller-13]

MTurk Task - Mozilla Firefox



- Please select the better transcription for this handwriting.
- Differences are highlighted in yellow.

3 votes @ \$0.01

> You (misspelled) (several) (words) (work). (?) (?) (?) work next (time). I also notice a few grammatical mistakes. Overall your writing style is a bit too (phoney). You do (?) have good (points), but they got lost amidst the (writing). (signature)

> You (?) (?) (?) (work). (?) (?) (?) work (not) (time). I (?) (?) a few grammatical mistakes. Overall your writing style is a bit too (phoney). You do (?) have good (points), but they got lost amidst the (writing). (signature)

Eg, Text Transcription [Miller-13]

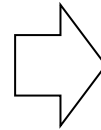
“You (misspelled) (several) (words). Please spellcheck your work next time. I also notice a few grammatical mistakes. Overall your writing style is a bit too **phoney**. You do make some good (points), but they **got** lost amidst the **(writing)**. **(signature)**”

According to our ground truth, the highlighted words should be “flowery”, “get”, “verbiage” and “B-” respectively.

After 9 iterations

Eg, Text Transcription [Miller-13]

I had intended to hit the nail, but I'm not a very good aim it seems and I ended up hitting my thumb. This is a common occurrence I know, but it doesn't make me feel any less ridiculous having done it myself. My new strategy will involve lightly tapping the nail while holding it until it is embedded into the wood enough that the wood itself is holding it straight, and then I'll remove my hand and pound carefully away. We'll see how this goes.



I had intended to hit the nail, but I'm not a very good aim it seems and I ended up hitting my thumb. This is a common occurrence I know, but it doesn't make me feel any less ridiculous having done it myself. My new strategy will involve lightly tapping the nail while holding it until it is embedded into the wood enough that the wood itself is holding it straight and then I'll remove my hand and pound carefully away. We'll see how this goes.

**After 8 iterations
with thousands of crowds**

Part 1: Crowdsourcing Basics

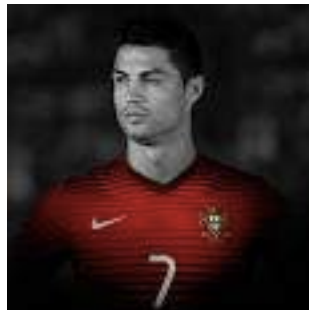
- Examples
- Definitions
- Marketplaces
- **Computational Crowdsourcing**
 - Preliminaries
 - Transcription
 - **Sorting** ←
- Demo

Human-Powered Sort

- Rank N items using crowdsourcing with respect to the constraint C
- Often C is subjective, fuzzy, ambiguous, and/or difficult-for-machines-to-compute
- Eg,
 - Which image is the most “representative” one of Brazil?
 - Which animal is the most “dangerous”?
 - Which actress is the most “beautiful”?

Human-Powered Sort

```
SELECT      *  
FROM        SoccerPlayers AS P  
WHERE       P.WorldCupYear = '2014'  
ORDER BY    CrowdOp('most-valuable')
```



...

Naïve Sort

- Eg, “Which of two players is better?”
- Naïve all pair-wise comparisons takes $\binom{N}{2}$ comparisons
 - Optimal # of comparison is $O(N \log N)$



...



...

...

...

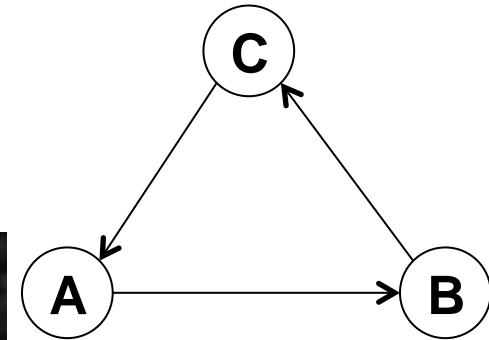


...



Naïve Sort

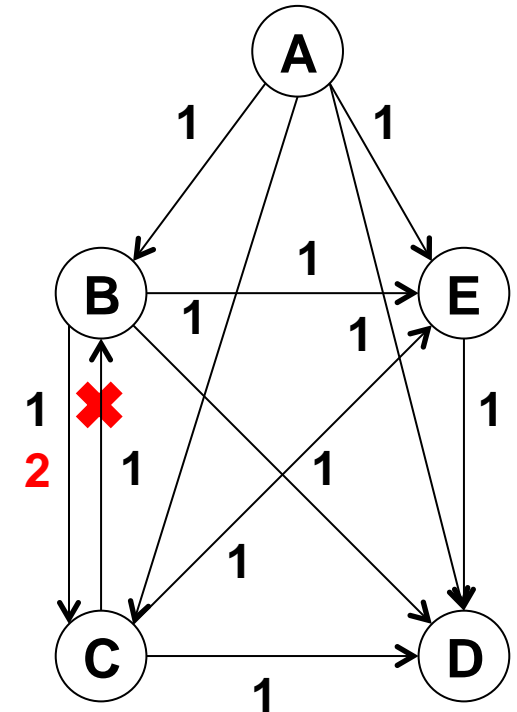
- Conflicting opinions may occur
 - Cycle: $A > B$, $B > C$, and $C > A$



- If no cycle occurs
 - Naïve all pair-wise comparisons takes $\binom{N}{2}$ comparisons
- If cycle exists
 - More comparisons are required

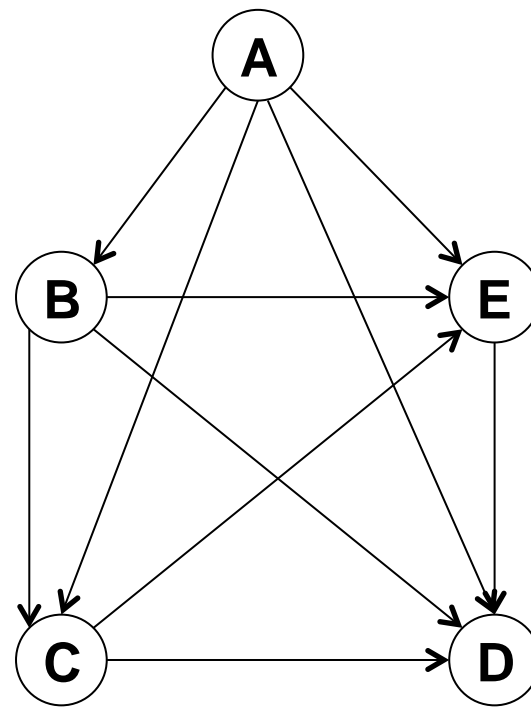
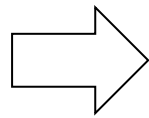
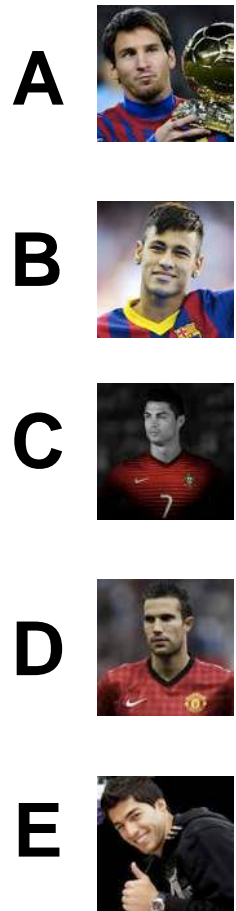
Sort [Marcus-VLDB11]

- $N=5$, $S=3$

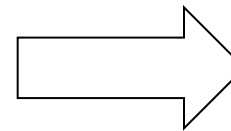


Sort [Marcus-VLDB11]

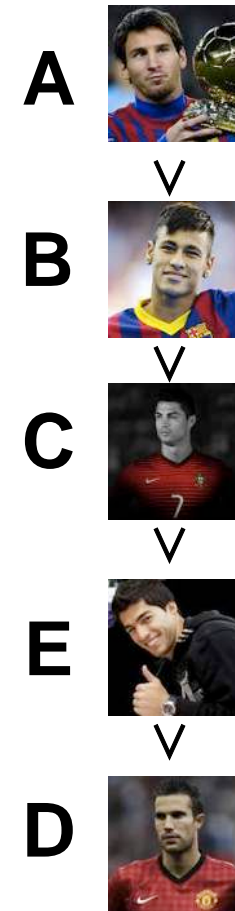
- $N=5$, $S=3$



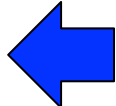
Topological
Sort



Sorted
Result



Part 1: Crowdsourcing Basics

- Examples
- Definitions
- Marketplaces
- Computational Crowdsourcing
 - Preliminaries
 - Transcription
 - Sorting
- **Demo** 

Demo: Human-Powered Sorting

- From your smartphone or laptop, access the following URL or QR code:

`http://goo.gl/3tw7b5`



Reference

- **[Brabham-13]** *Crowdsourcing*, Daren Brabham, 2013
- **[Franklin-SIGMOD11]** *CrowdDB: answering queries with crowdsourcing*, Michael J. Franklin et al, SIGMOD 2011
- **[Howe-08]** *Crowdsourcing*, Jeff Howe, 2008
- **[LawAhn-11]** *Human Computation*, Edith Law and Luis von Ahn, 2011
- **[Li-HotDB12]** *Crowdsourcing: Challenges and Opportunities*, Guoliang Li, HotDB 2012
- **[Marcus-VLDB11]** *Human-powered Sorts and Joins*, Adam Marcus et al., VLDB 2011
- **[Miller-13]** *Crowd Computing and Human Computation Algorithms*, Rob Miller, 2013
- **[Shirky-08]** *Here Comes Everybody*, Clay Shirky, 2008