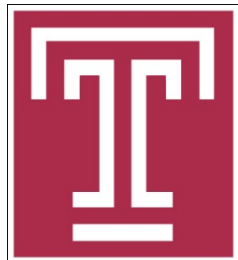


Algorithmic Crowdsourcing and Applications in the Big Data Era

Jie Wu

Dept. of Computer and Info. Sciences

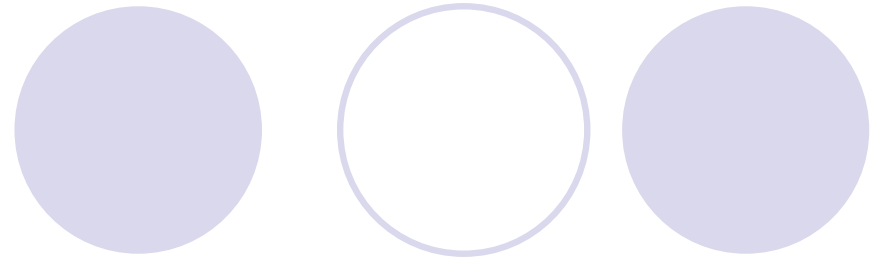
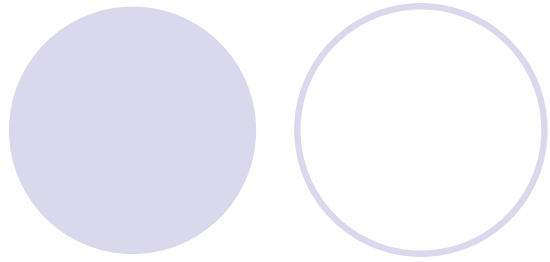
Temple University



Road Map

- Introduction
- Mechanical Turk
- Sample Applications
- Algorithmic Paradigms
- Challenges and Opportunities
- Conclusion





Big data era

What and why

Basic components

Motivation examples

INTRODUCTION

How Much Data?

- **Facebook:** 40 B photos; 30 B pieces of content shared every month
- **WeChat:** 846 M users and 20 B messages per day
- **Global Internet traffic:** quadrupled from 2010 to 2015, reaching 966 EB (10^{18}) per year

(The total amount of human knowledge since the dawn of man until 2003 equals 5 EB)



640K ought to be enough for anybody.

Big Data Era

- "In information technology, big data consists of datasets that grow so large that they become awkward to work with using on-hand database management tools."
- Computers are not efficient in processing or creating certain things: pattern recognition, complex communication, and ideation.



What is Crowdsourcing?

- Coordinating a **crowd** (a large group of people online) to do **microwork** (small jobs) that **solves problems** (that software or one user cannot easily do)
- Crowdsourcing: crowd + outsourcing (through Internet)

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



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



- Amazon's Mechanical Turk and CrowdFlower

The Benefits of Crowdsourcing

- Performance
 - Inexpensive and fast
 - The whole is greater than the sum of its parts
- Human Processing Unit (HPU)
 - More effective than CPU (for some apps)
 - Verification and validation: Image labeling
 - Interpretation and analysis: language translation
 - Surveys: Social network survey
- Widespread adoption in business (85% of the top global brands, according to eYeka)

Basic Components

- Requesters
 - People who submit jobs (microwork)
 - Human Intelligence Tasks (HITs)
- Workers
 - People who work on the submitted jobs
- Platform
 - Job management

Amazon **Mechanical Turk (MTurk)**: 18th century chess playing robot with a human inside



Requester

amazon
mechanical turk
beta

Jilin Univ.



Worker Pool

History



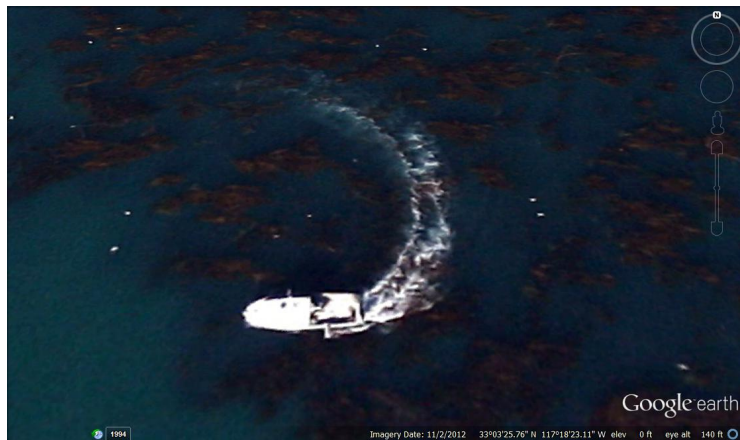
- **1714:** Longitude Prize: tried to find a way to measure a ship's longitudinal position
- **1884:** 800 volunteers catalogued words to create *Oxford English Dictionary*
- **2001:** Wikipedia: free-access, free content Internet encyclopedia
- **2005:** Threadless.com, with members creating own design
- **2006:** Howe and Robinson introduced the term in *Wired*
- **2008:** Brabham published first scholarly work using the term

Help Find Jim Gray

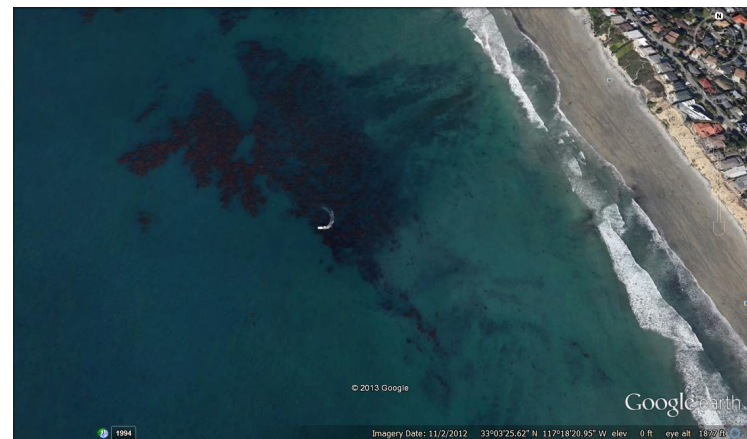


- Jim Gray, Turing Award winner, went missing with his sailboat outside San Francisco Bay in January 2007.

- Use satellite image to search for his sailboat.



Jilin Univ.



Malaysia Airlines Flight MH 370

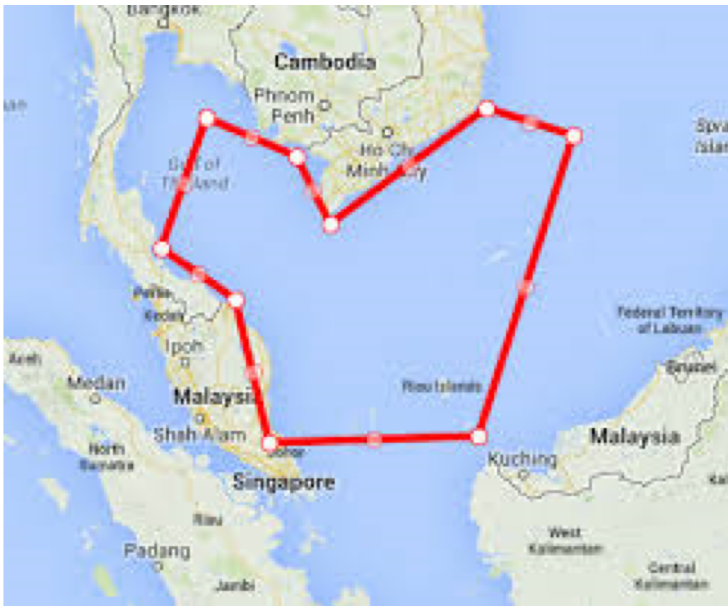


- DigitalGlobe

- Crowdsourcing volunteers combed through satellite photos for Malaysia Airlines jet

- March 11, 2014 (from CSU prof. email)

I just saw on our local Denver Fox news (KDVR.com) that a local company, DigitalGlobe, has reoriented their satellites to take high-res images in the area where the plane may have crashed. Crowdsourcing efforts are on to have people scan these images and find signs of debris. *I was reminded of Jie Wu's talk earlier this month.*



DARPA Network Challenges

WE HAVE A WINNER!

MIT RED BALLOON CHALLENGE TEAM

Read about the winner of the DARPA Network Challenge



- Problem (2009): \$40,000 challenge award for the first team to find 10 balloons.
- MIT team won under 9 hours.
- Winning strategy
 - \$2,000 per balloon to the first person to send the correct location
 - \$1,000 to the person who invited the winner
 - \$500 to whoever invited the inviter
 - ... (or to charity) ...

Tag Challenges



- Problem (March 31, 2012): Find five suspects in Washington, D.C., New York, London, Stockholm, and Bratislava.
- Winner from UCSD CrowdScanner: located 3 of the 5 suspects.
- Winning strategy: same as MIT. Also, recruiters of the first 2,000 eat gets \$1.

Washington DC



New York City

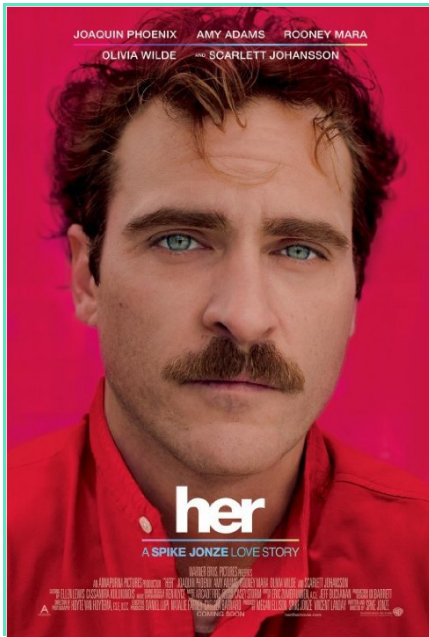


Bratislava



AI Could End Human Race (Stephen Hawking)

- Stephen Hawking
 - "Humans, who are limited by slow biological evolution, couldn't compete, and would be superseded."
- Recent movies
 - Her (2014) & Ex-Machina (2015)

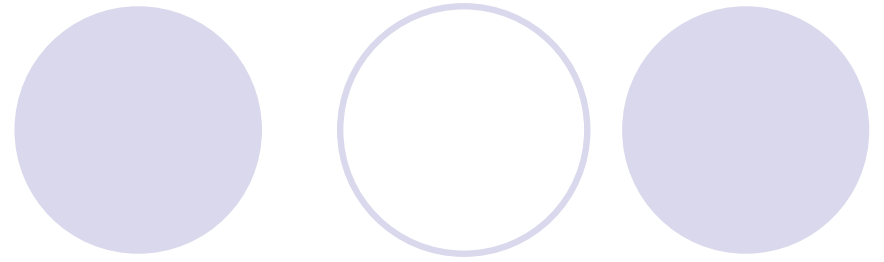
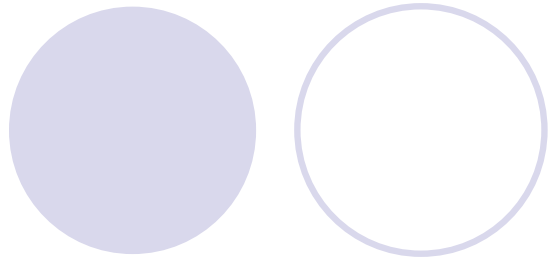


Smarter Than You Think



- Who is smarter
 - Human or computer?
 - AI will redefine
 - What it means to be human
- Our Machine Masters
NY Times, Oct. 31, 2014

- 1997 (Chess)
 - Kasparov vs. Deep Blue
- 1998
 - Kasparov vs. Topalov: 4:0
 - Kasparov + machine vs. Topalov + machine: 3:3
- 2005 (freestyle tournament)
 - Grand-master (>2,500)
 - Machine (Hydra)
 - Grand-master + machine
 - Amateurs (>1,500) + machine *
- 2016 (Go game)
 - AlphaGo vs. Lee Sedol: 4:1
 - AlphaGo vs. Jie Ke: 3:0 (May 2017)



Platform

HIT

Worker

MECHANICAL TURK

Mechanical Turk is a marketplace for work.

We give businesses and developers access to an on-demand, scalable workforce. Workers select from thousands of tasks and work whenever it's convenient.

293,089 HITs available. [View them now.](#)

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



or [learn more about being a Worker](#)

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



- As a worker, make an average of \$0.05 per task
- Paid directly to Amazon account
- 130 M tasks posted (2009-2014)
- As requester, set up simple tasks for workers to complete
- Quality control is possible through MTurk services

Worker: Contract for a HIT

All HITS

1-10 of 1982 Results

Sort by:

Show all details | Hide all details

1 2 3 4 5 > Next >> Last

Copy Text from Coupon Image	Not Qualified to work on this HIT (Why?) View a HIT in this group	
Requester: Coupon Vision	HIT Expiration Date: Jun 21, 2014 (51 weeks 2 days)	Reward: \$0.08
	Time Allotted: 10 minutes	HITS Available: 14
Proofread OCR Data	Take Qualification test (Why?) View a HIT in this group	
Requester: Brian Robertson	HIT Expiration Date: Jul 3, 2013 (6 days 23 hours)	Reward: \$0.30
	Time Allotted: 2 hours	HITS Available: 2
Get product codes and prices from receipt image (get bonuses for long receipts)	Request Qualification (Why?) View a HIT in this group	
Requester: Shopping	HIT Expiration Date: Jul 1, 2013 (4 days 23 hours)	Reward: \$0.03
	Time Allotted: 45 minutes	HITS Available: 2
Click and provide fast feedback B-US RHL-003	Not Qualified to work on this HIT (Why?) View a HIT in this group	
Requester: CrowdFlower	HIT Expiration Date: Jul 3, 2013 (6 days 23 hours)	Reward: \$0.01
	Time Allotted: 30 minutes	HITS Available: 219
Basic Caption Requirements	View a HIT in this group	
Requester: Redwood	HIT Expiration Date: Jun 26, 2014 (52 weeks)	Reward: \$0.02
	Time Allotted: 15 minutes	HITS Available: 11
Identify company/publication from a photo	Not Qualified to work on this HIT (Why?) View a HIT in this group	

- Select a HIT
 - By creation date, payment amount, time allotment

Worker: Reviewing a HIT

Timer: 00:00:00 of 10 minutes

Want to work on this HIT?

Accept HIT

Want to see other HITs?

Skip HIT

Total Earned: \$4.72
Total HITs Submitted: 7

Copy Text from Business Card

Requester: Oscar Smith

Qualifications Required: None

Reward: \$0.02 per HIT

HITs Available: 39

Duration: 10 minutes

Please Copy Text from Business Card:



Please **select/crop** company logo or image from the business card above. Click + Drag to select the company logo.

Your Current Quality Score is:

If you have a high enough score, you will be [?](#) **--** considered for promotion to a Trusted Worker.

Name [?](#)

Title Company

Email Website

Address: [?](#)

Address Line 1

[add line](#)

City State Zip Code

Phone: [click here if not a U.S. phone number](#) [?](#)

Work Ext.

Mobile

Fax

[add phone](#)

- Review the HIT before accepting
 - Shown full task, allotted time (10 minutes), reward amount (\$0.02)

Types of HITS



- Information Finding
 - Searching the web to answer a question
- Interpretation and Analysis
 - Interpreting web content
- Verification and Validation
 - Verifying and validating certain information
- Content Creation
 - Generating new content
- Content Access
 - Accessing web content
- Surveys
 - Taking appropriate action based on the survey result

Threadless

Order Status Help

LOG IN / JOIN US!



SHOP

PARTICIPATE

CREATE

Search



SHOP 50% OFF WALL ART

SHOP NOW

Sale prices as marked.
Promo ends 7/17 at 10am
CST. Excludes tapestries.

SHOP MEN'S BY POPULARITY / NEWEST

DEPARTMENT

[VIEW ALL](#)

- Men's
- Women's
- Kids
- Babies
- Toddlers
- Phone Cases
- Wall Art
- Accessories
- Home

STYLE

[VIEW ALL](#)

- Tees
- Triblend Tees
- V-Necks
- Longsleeves
- Hoodies
- Sweatshirts
- Tanks
- Sublimated Triblend Tees

SIZE

- X Small
- Small
- Medium
- Large
- X Large
- 2X Large
- 2X Large



24% OFF

Let's Summon Demons
Design by Steven Rhodes
~~\$25~~ \$19 and up



20% OFF

Can Knot
Design by Haasbroek
~~\$30~~ \$24



24% OFF

Many Lands Under One Sun
Design by Rick Crane
~~\$25~~ \$19



24% OFF

The Cat and the Koi
Design by roni saptoni
~~\$25~~ \$19



InnoCentive





INNOCENTIVE®

[Solvers](#) [Our Offering](#) [Challenge Center](#) [Resources](#) [About](#) [Blog](#) [My IC](#) | [Log Out](#)

[My Challenges to Solve](#) [My Interests](#) [My Account](#) [My Public Profile](#) [Solver Search](#)

My Challenges to Solve

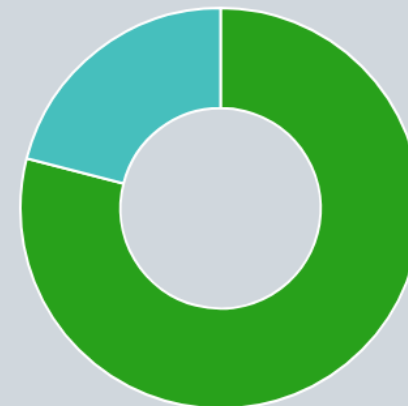
[Open \(0\)](#) [Recommended](#)

	Title	Posted	Deadline	Award	Solvers
	Culturing Canine Heartworm TAGS: Chemistry Life Sciences	Jul 12 2017	Sep 10 2017 23:59 EDT	\$20,000 USD	23
	Novel Approaches for Targeting an Orally-Administered Drug to the Liver TAGS: Chemistry Life Sciences	Jul 12 2017	Aug 11 2017 23:59 EDT	\$20,000 USD	68
	Seeking New Use Cases for Rollable Display Screens TAGS: Business/Entrepreneurship Engineering/Design	Jun 29 2017	Aug 28 2017 23:59 EDT	\$10,000 USD	357
	Seeking for a New Material to Increase Stiffness of Multilayered Cartonboard TAGS: Chemistry Engineering/Design Physical Sciences	Jun 22 2017	Aug 06 2017 23:59 EDT	\$20,000 USD	110

Profile completeness

Your profile is 79% complete

■ Completeness ■ Remaining



Featured Challenge



White Papers



Crowdsourcing Drug Development

This white paper looks at the role crowdsourcing can play throughout the drug development process, from discovery to early development and even late-stage manufacturing and marketing.

[View White Paper >](#)



Open Innovation for All: The General Fusion Experience

Fortune 500s with robust innovation programs have long benefited from crowdsourcing, but what do agile and innovative technology companies stand to gain? In this white paper we will look at the experiences of Brendan Cassidy, Open Innovation Manager at General Fusion, of using InnoCentive Challenges as a business tool.

[View White Paper >](#)



Maximizing ROI on Crowdsourcing and Internal Collaboration

There are four critical steps for optimizing your external and internal crowdsourcing initiatives and accelerating your internal collaboration. Find out how to tackle each of them in this white paper.

[View White Paper >](#)



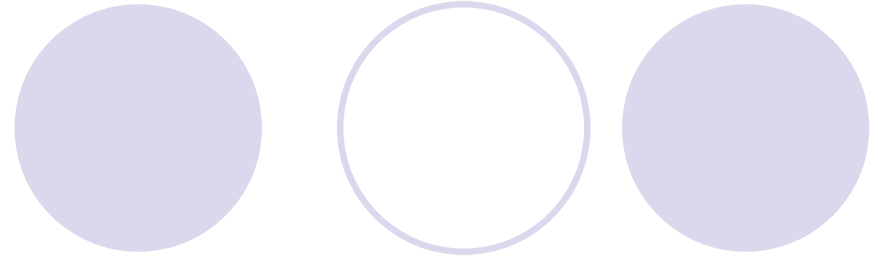
Innovation Metrics

Being able to accurately measure innovation across your organization is critical for optimizing your innovation programs and maximizing ROI. Through this white paper, we get to the core of what innovation really means and help you establish the innovation metrics that are right for your organization.

[View White Paper >](#)



Crowdfunder



Work mode 100% accuracy 1 task completed \$0.02 per task

Give up  0 Blog Help ▾ Nick ▾ 00:00

Imagine that you can talk to your phone's personal assistant while using the navigation app. You can ask the assistant to find places for you, get directions, and more. We programmed our computer to generate some navigation app requests for the assistant. Unfortunately, the computer's English is hard to understand. We need your help to improve it! Below is one of the computer-generated requests. Can you express the entire request in a natural English sentence?

Request:

find contact information and edit phone number

--> where contact information is for **Bob**

--> set phone number to **555-8374**

How would you say this in more natural English? Please provide 3 unique responses.

Response 1

Response 2

Response 3

Jilin Univ.

Crowdfunder



Quiz mode 7 of 10 to pass

Give up



Blog

Help ▾

Nick ▾

00:00

Business Name: ANGEL GARAGE DOOR REPAIR MTN

Street Address: 1174 CASTRO ST # 201

City: MOUNTAIN VIEW

State: CA

[Click here to Google the business](#)

Can you find the official website for this business?

- Yes
- No

i Make sure this is not a directory site.
Also make sure the business name matches!

Business Name: DAVIS HOMES INC

Street Address: 18800 CLARKE RD

City: PARKER

State: CO

[Click here to Google the business](#)

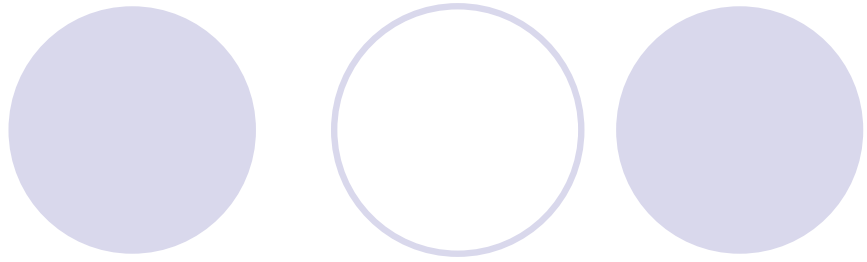
Can you find the official website for this business?

- Yes
- No

i Make sure this is not a directory site.
Also make sure the business name matches!

Jilin Univ.

iStock Photo



Photos Illustrations Video Audio

iStock
by Getty Images

Your Board Pricing Account

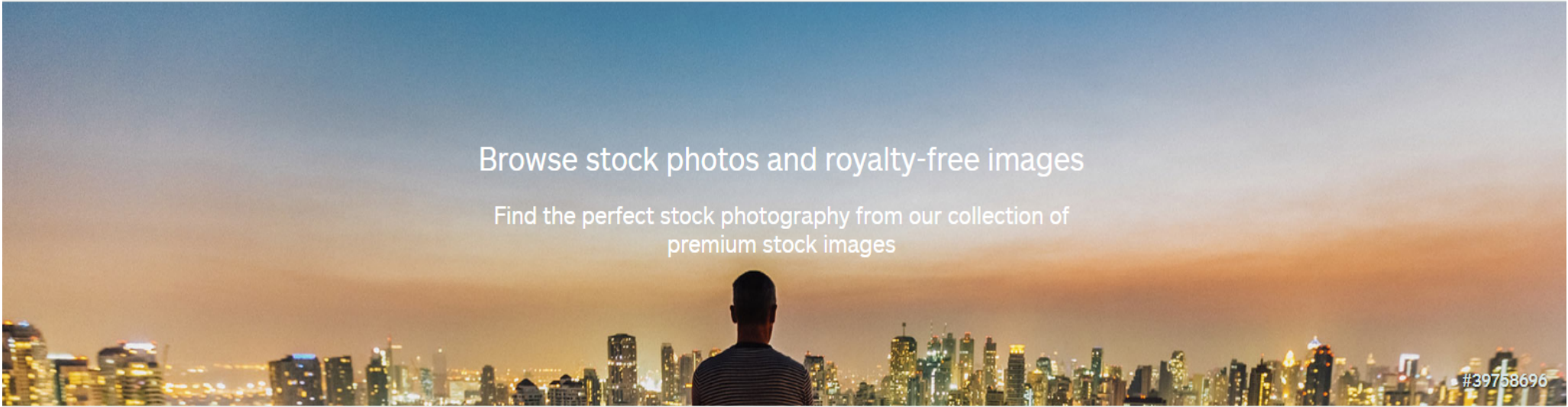
Find the perfect stock photos...



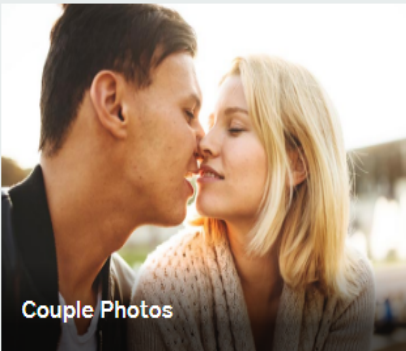
Photos

Browse stock photos and royalty-free images

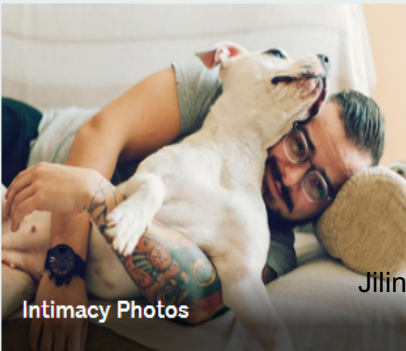
Find the perfect stock photography from our collection of premium stock images



#39758696



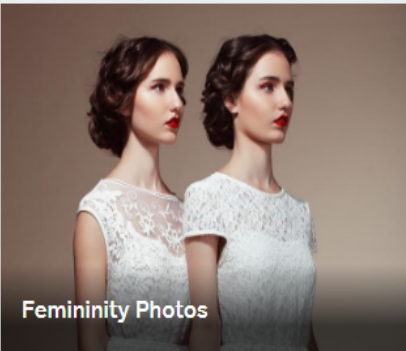
Couple Photos



Intimacy Photos



Woman Photos



Femininity Photos

TopCoder



COMPETE

LEARN

COMMUNITY

nick_boyd7  

Challenges [SRMs](#)



























Design Development Data Science

 Filters

Open for registration

Sort by: Prize high to low

 	Action Vector - First 100 Setup challenge Ends Jul 31 Action Vector – SMP Analytics Database	\$ 10,000 Purse	Submission <div><div style="width: 100%;"></div></div> 16d 23:19h to go	 187  100	16d 23:19h to register
 	Financial Machine Learning PoC Ends Jul 17 IBM Watson	\$ 2,850 Purse	Submission <div><div style="width: 100%;"></div></div> 2d 19:19h to go	 54  0	2d 19:18h to register
 	QED Learning Library Mobile Application Design Challenge Ends Jul 21	\$ 2,750 Purse	Checkpoint <div><div style="width: 100%;"></div></div> 2d 16:17h to go	 62  0	2d 16:18h to register
 	Hercules C++ Fog App - Multi-location redirect support Ends Jul 16 C C++	\$ 2,250 Purse	Submission <div><div style="width: 100%;"></div></div> 2d 7:21h to go	 33  0	2d 7:20h to register
 	XPrize Visioneers Native iOS App API Integration Part II Challenge Ends Jul 19 Swift iOS	\$ 2,100 Purse	Submission <div><div style="width: 100%;"></div></div> 4d 16:31h to go	 8  0	2d 16:31h to register
 	GE - Future of Air & Surface System Management Software Wireframe Challenge Ends Jul 24	\$ 2,000 Purse	Checkpoint <div><div style="width: 100%;"></div></div> 4d 18:19h to go	 27  0	4d 18:13h to register

All Challenges	38
Open for registration	22
Ongoing challenges	16
Open for review	0
Upcoming challenges	0
Past challenges	0

[Get the RSS feed](#)

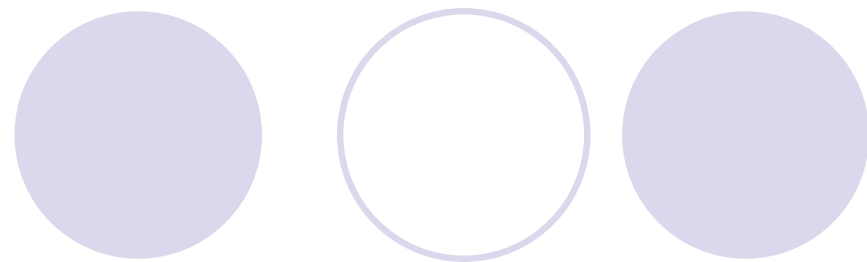
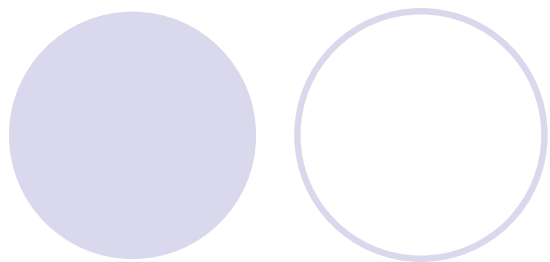
[About](#) • [Contact](#) • [Help](#) • [Privacy](#) • [Terms](#)

Topcoder © 2017

 Support

Major Types of Crowdsourcing

- **Virtual labor markets (VLM)** - Platforms where users can complete work for monetary compensation (e.g., Amazon's Mturk)
- **Tournament Crowdsourcing (TC)** - Also known as idea competitions, where only the best solution is compensated (e.g., Crowdfunder and TopCoder)
- **Open collaboration (OC)** - Typically do not offer monetary compensation, people are often prompted through social media with the opportunity to help out; like an open call to anyone (e.g., Wikipedia)

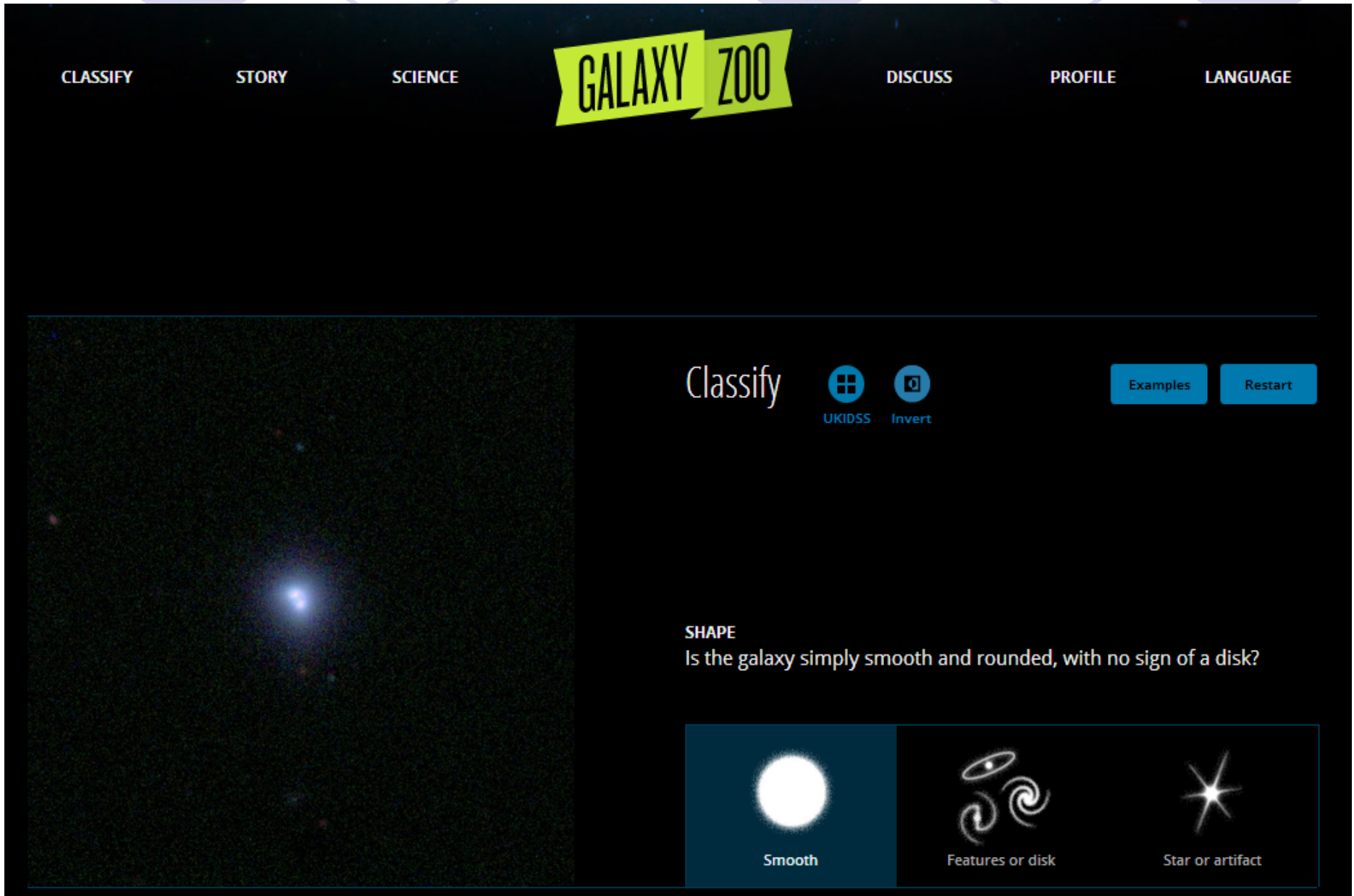


Galaxy Zoo

Fine-grained Recognition



APPS: IMAGE PROCESSING

Galaxy Zoo: Zooniverse






The screenshot displays the Galaxy Zoo Zooniverse interface. At the top, a navigation bar includes links for CLASSIFY, STORY, SCIENCE, DISCUSS, PROFILE, and LANGUAGE. The 'GALAXY ZOO' logo is prominently displayed in the center of the navigation bar. Below the navigation bar, the main content area is divided into two sections. On the left, a large image shows a galaxy with a bright central core and a diffuse, rounded structure. On the right, the 'Classify' section contains several interactive elements: a 'UKIDSS' icon, an 'Invert' icon, and two buttons labeled 'Examples' and 'Restart'. Below these icons, the text 'SHAPE' is followed by the question 'Is the galaxy simply smooth and rounded, with no sign of a disk?'. At the bottom of the interface, three options are presented as buttons: 'Smooth' (with a white circle icon), 'Features or disk' (with a spiral galaxy icon), and 'Star or artifact' (with a star icon).

CLASSIFY STORY SCIENCE **GALAXY ZOO** DISCUSS PROFILE LANGUAGE

Classify   [Examples](#) [Restart](#)

UKIDSS Invert

SHAPE
Is the galaxy simply smooth and rounded, with no sign of a disk?

Smooth Features or disk Star or artifact

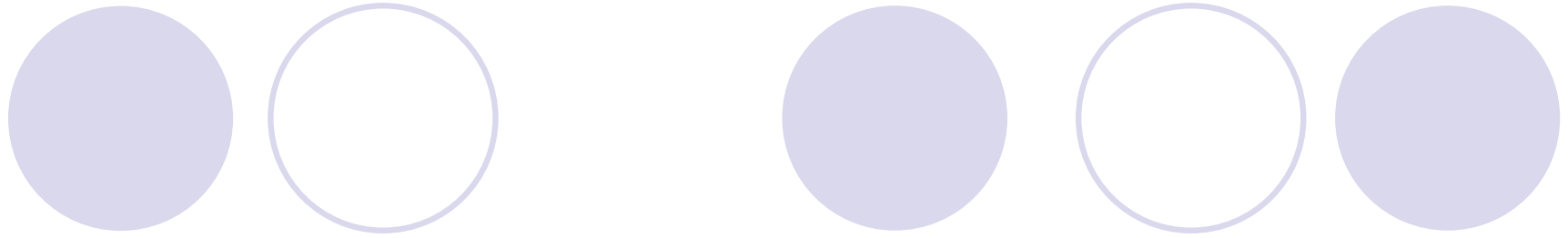
Galaxy Zoo: Zooniverse

The screenshot displays the Galaxy Zoo Zooniverse interface. At the top, a navigation bar includes links for CLASSIFY, STORY, SCIENCE, DISCUSS, PROFILE, and LANGUAGE. The central logo features the text "GALAXY ZOO" in a stylized font. The main content area is divided into two sections. On the left is a large image of a galaxy. On the right, the "Classify" section contains two icons: "UKIDSS" (a 3x3 grid) and "Invert" (a camera icon). Below these are two buttons: "Examples" and "Restart". The classification question is "ROUND How rounded is it?". At the bottom, three options are presented with corresponding galaxy images: "Completely round" (a circular galaxy), "In between" (an oval galaxy), and "Cigar shaped" (a thin, elongated galaxy).

Fine-Grained Recognition: Tohme



K. Hara et al, "Tohme: Detecting Curb Ramps in Google Street View Using Crowdsourcing, Computer Vision, and Machine Learning," UIST 2014



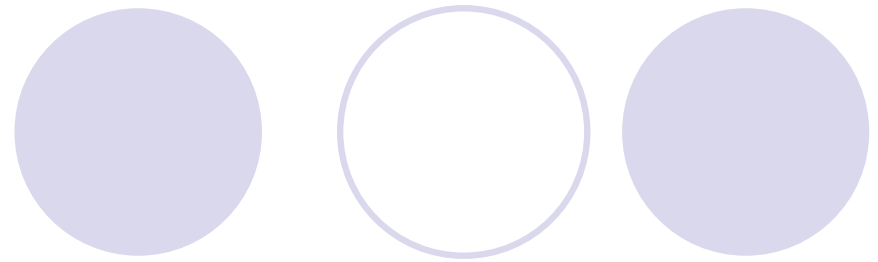
GWAP.com

reCAPTCHA

Crowdvoting (Crowdfunding and Crowdsearching)

APPS: COMMON SENSE KNOWLEDGE

GWAP.com: CMU



ESP Game

- Labeling images

ESP Game
Concentrate...

How to Play

- 1 You and a partner see the same image.
- 2 Each of you must guess what words your partner is typing.

Got it, Let's Play!

View Top Scores

Tag a Tune

- Labeling tunes

Tag a Tune
Hear Here

How to Play

- 1 You and a partner hear a tune and must describe it.
- 2 Based on the descriptions, you have to figure out if you're both listening to the same tune!
- 3 There are several other bonus rounds which are self descriptive. Enjoy!

Got it, Let's Play!

View Top Scores

reCAPTCHA: CMU



- WHAT IS reCAPTCHA
- GET reCAPTCHA
- PROTECT YOUR EMAIL
- MY ACCOUNT
- RESOURCES: DOCS & PLUGINS

reCAPTCHA IS A FREE ANTI-BOT SERVICE THAT HELPS DIGITIZE BOOKS.

steamboat train, from New
this **morning** ran off the track
New-London. Four cars plunged



→ [LEARN HOW reCAPTCHA WORKS](#)

[USE reCAPTCHA ON YOUR SITE](#)

- STRONG SECURITY**
- ACCESSIBLE TO BLIND USERS**
- 30+ MILLION SERVED DAILY**

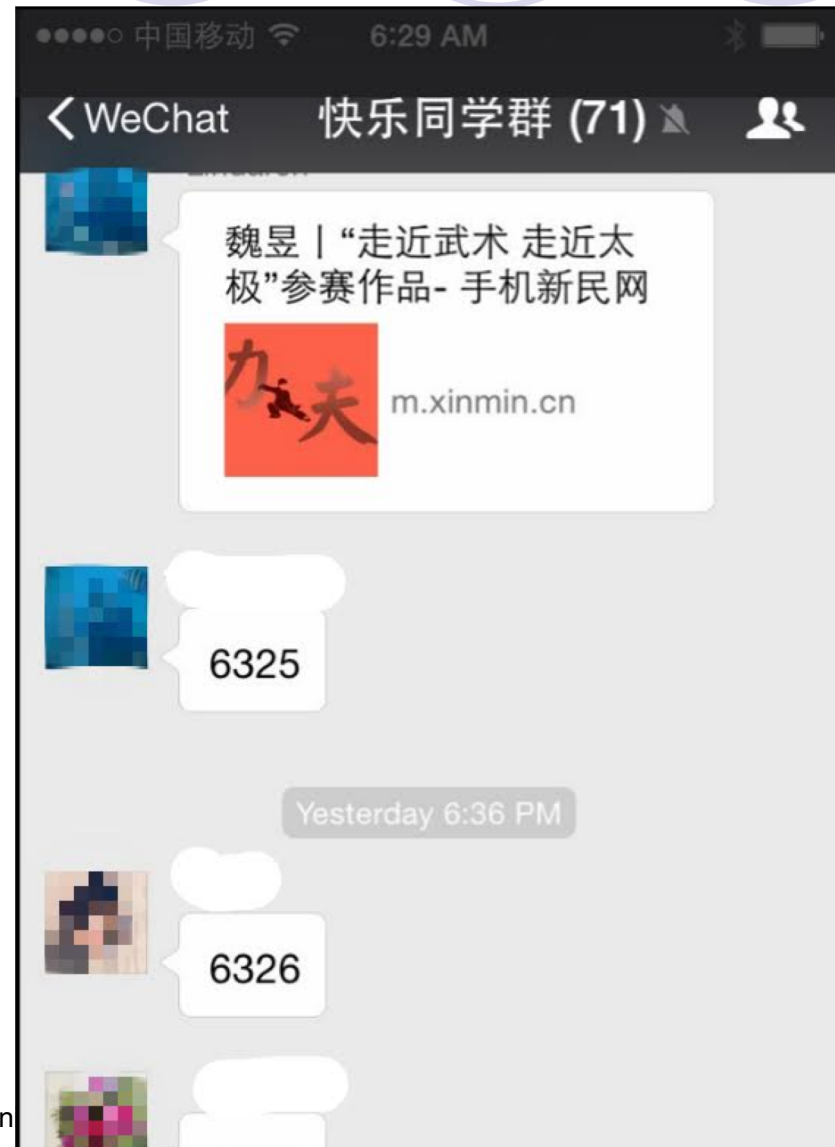
NEW See how accurate reCAPTCHA is at digitizing content!

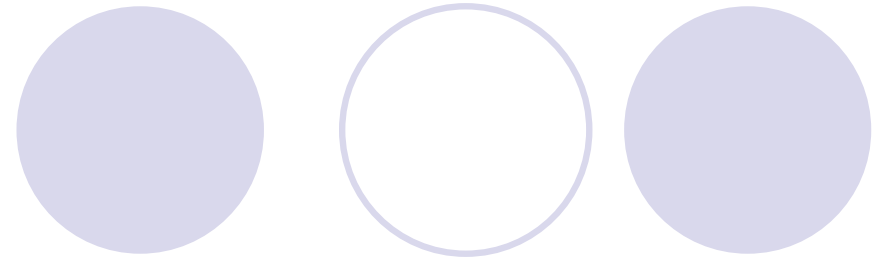
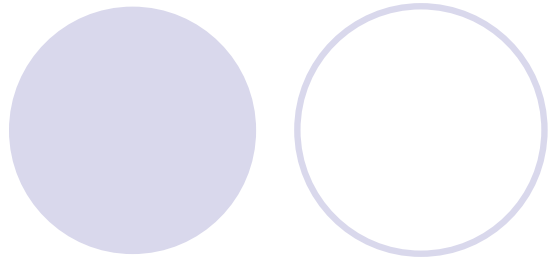
Jilin Univ.

[Blog](#) | [About Us](#) | [Contact](#) | [FAQs](#) | [Terms](#) | [Privacy](#)
© 2009, all rights reserved.

Crowdvoting in Social Networks

- Online voting
- Example
 - WeChat group
 - Social recognition as incentive
- Others
 - Crowdfunding
 - Crowdsearching





Software Engineering

Online feedback

Waze

Crowdsensing (Smart city)

APPS: LARGE PROJECTS

Software Engineering

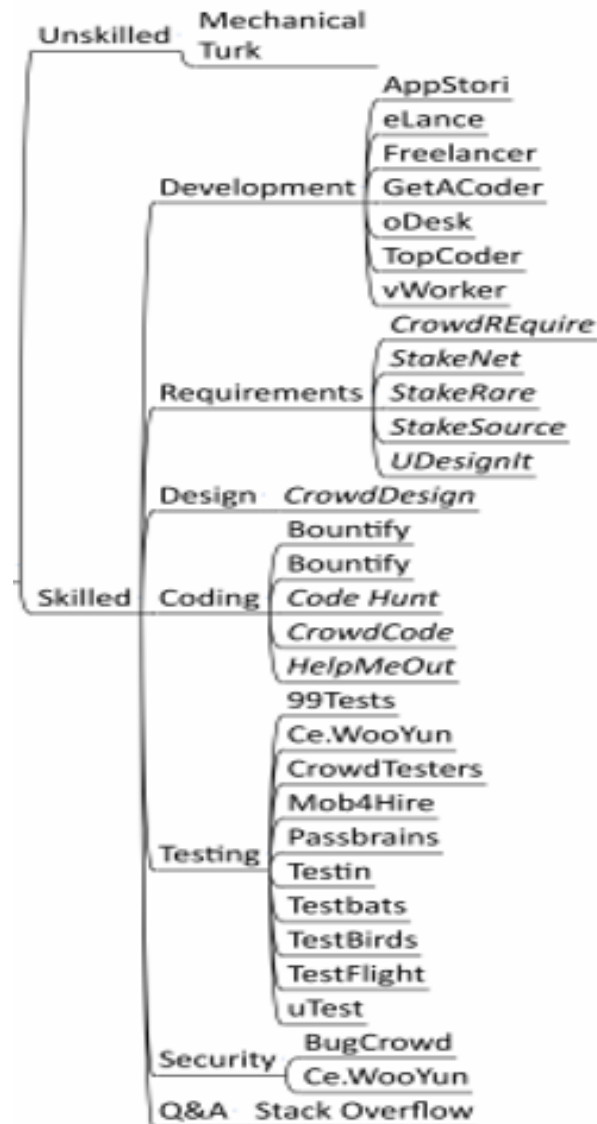
- Issues/concerns

- Planning, scheduling, coordination, motivation, intellectual property

- Some solutions

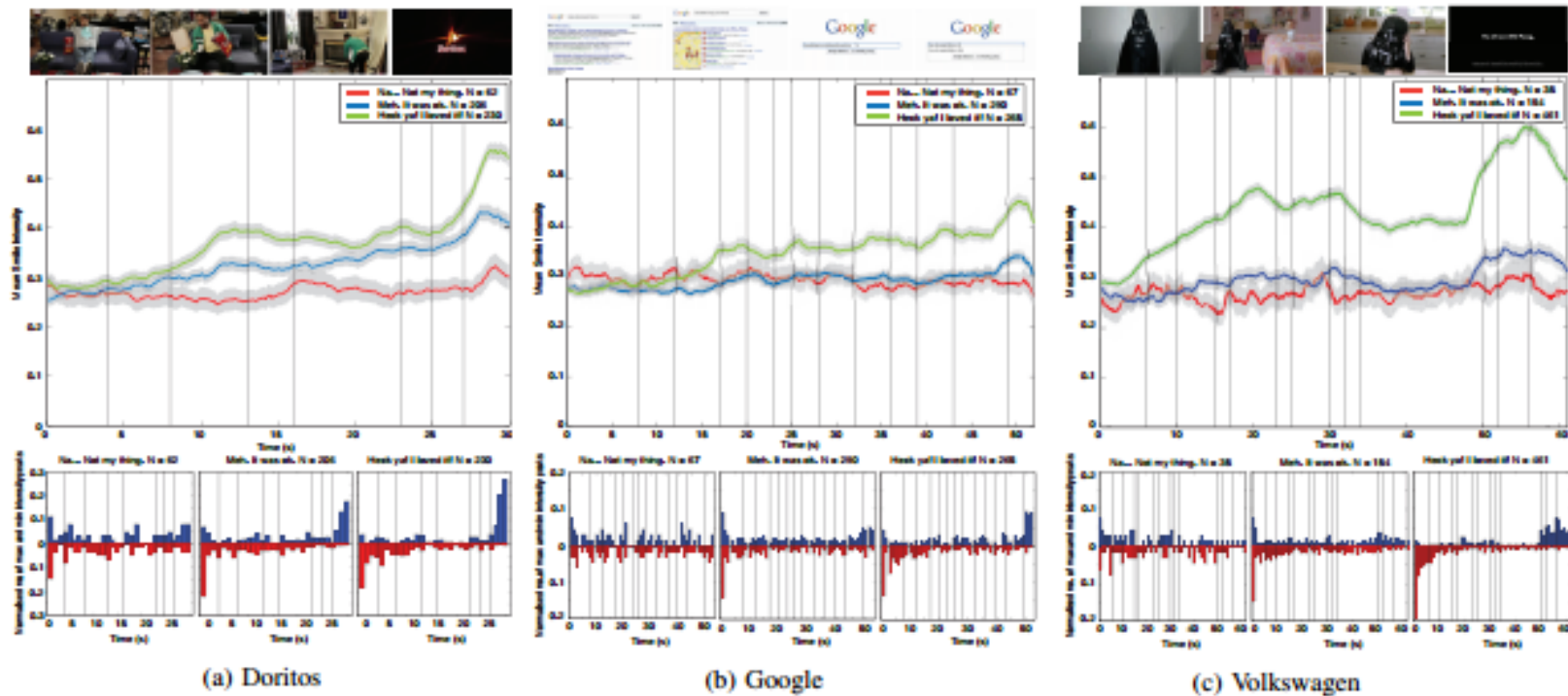
- Mturk and TopCoder use monetary rewards

Ke Mao et al., "A survey of the use of of crowdsourcing in software engineering," Journal of Systems and Software, 2017



Online Feedback

- Crowdsourcing facial responses to online videos



D. McDuff et al., "Crowdsourcing facial responses to online videos." *Proc. of Int'l Conf. on Affective Computing and Intelligent Interaction (ACII)*, 2015.

Waze as Mobile Crowdsourcing

- Waze is the world's largest community-based traffic and navigation app.

SEND YOUR ETA & REAL-TIME DRIVE TO UPDATE THOSE YOU'RE MEETING

REPORT REAL-TIME TRAFFIC & ROAD CONDITIONS TO IMPROVE DRIVING FOR ALL

GET ALERTED BEFORE YOU APPROACH POLICE, ACCIDENTS, HAZARDS & MORE

SEE FRIENDS HEADING TO THE SAME DESTINATION & WHEN THEY'LL ARRIVE

Jilin Univ.

The image consists of four vertical panels, each showing a smartphone screen with a different Waze feature. The first panel (red background) shows the 'Send ETA' feature, with a circular callout highlighting the 'SEND YOUR ETA' button. The second panel (blue background) shows the 'Report' menu with various icons for reporting traffic conditions like 'Traffic Jam', 'Police', 'Accident', 'Hazard', 'Gas Prices', 'Map Chat', 'Map Issue', and 'Place'. The third panel (light blue background) shows a 'Police' alert on the map. The fourth panel (orange background) shows a map with a purple route and a circular callout highlighting the 'See Friends' feature, which shows a list of friends heading to the same destination.

Smartphone-based Crowdsensing

● Smart city

○ Public information reposting

B. Guo et al, "FlierMeet: A Mobile Crowdsensing System for Corss-Space Public Info. Reposting, Tagging, and Sharing," IEEE TMC, Oct. 2015

○ Public transportation

Z. He et al, "High Quality Participant Recruitment in Vehicle-based Crowdsourcing using Predicatable Mobility," INFOCOM 2015

○ Indoor map construction

R. Gao et al, "Multi-story Indoor Floor Plan Reconstruction via Mobile Crowdsensing," IEEE TMC April 2016

○ Street view service

Y. Wu et al, "Photo Crowdsourcing for Area Coverage in Resource Constrained Environments," INFOCOM 2017

● Some challenges

○ Data redundancy

C. Meng et al, "Tackling the Redundancy and Sparsity in Crowd Sensing Applications," SenSys 2016

○ Trustfulness & Game

G. Wang et al, "Defending Against Sybil Devices in Crowdsourced Mapping Services," MobiSys 2016

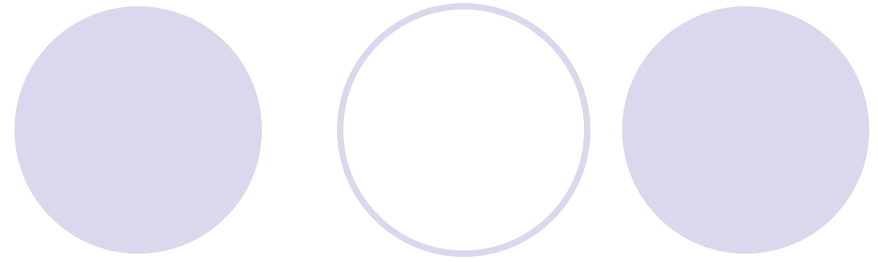
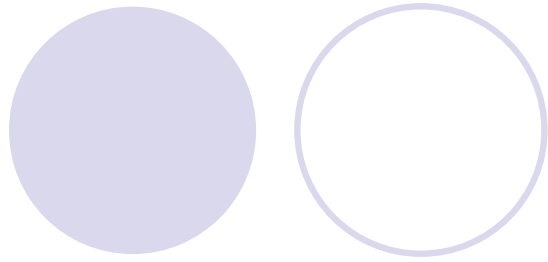
T. Luo et al, "Crowdsourcing with Tullock Contests: a New Perspective," INFOCOM 2015

○ Online assignment

Y. Tong et al, "Online Mobile Micro-Task Allocation in Spatial Crowdsourcing," ICDE 2016

○ Incentive

H. Jin et al, "CENTURION: Incentivizing Multi-Requester Mobile Crowd Sensing," INFOCOM 2017



Sequential

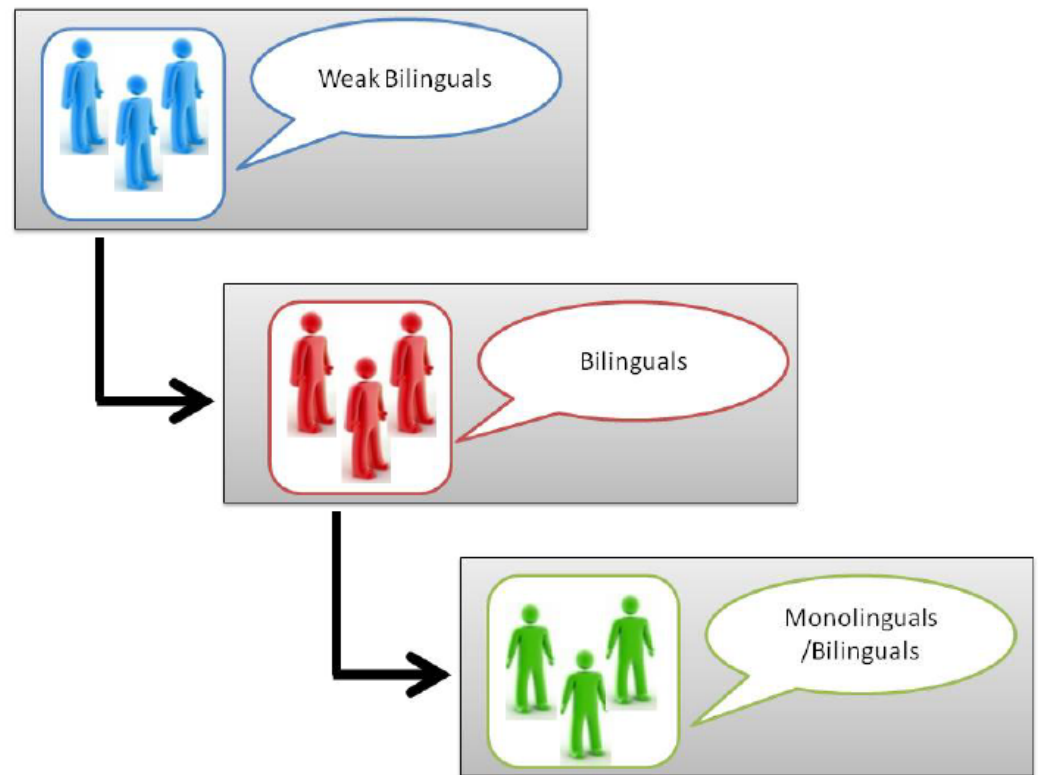
Iterative and Parallel

Divide-and-Conquer and Aggregate

PARADIGMS

Sequential: Collaborative Workflow

- Lexical translation
(weak bilinguals or machine)
- Assistive translation
(strong bilinguals)
- Refined sentences
(monolinguals)



V. Ambati et al, "Collaborative Workflow for Crowdsourcing Translation," CSCW 2012

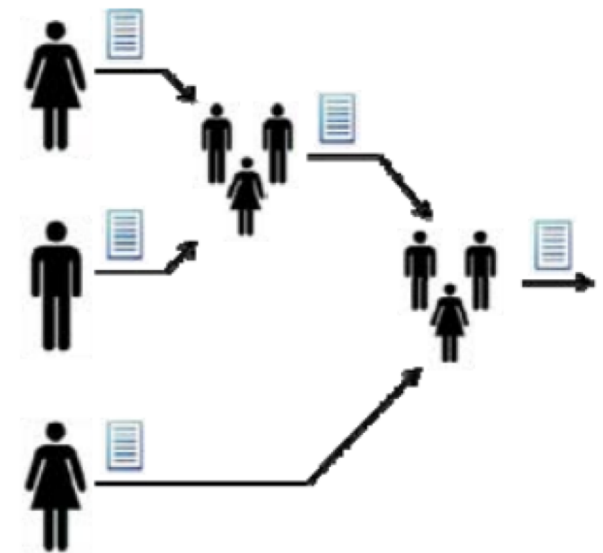
Iterative and Parallel

- Iterative improve and vote

The screenshot shows two browser windows side-by-side. The left window, titled 'Handwriting Recognition Task - Mozilla Firefox', displays a handwritten document and a list of instructions: 'Please improve the transcription of this handwriting.' and 'People will vote whether to approve your changes.' Below the instructions, a feedback message reads: '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)'. A 'Submit' button is at the bottom right.

The right window, titled 'M Turk Task - Mozilla Firefox', displays the same handwritten document and instructions: 'Please select the better transcription for this handwriting.' and 'Differences are highlighted in yellow.' Below, two options are presented for selection. The first option has several words highlighted in yellow: 'You (misspelled) (several) (words) (work). (?) (?) (?) work (not) (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)'. The second option has different words highlighted: '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)'. A 'Submit' button is at the bottom right.

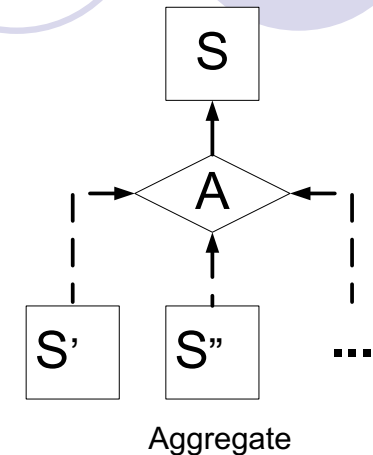
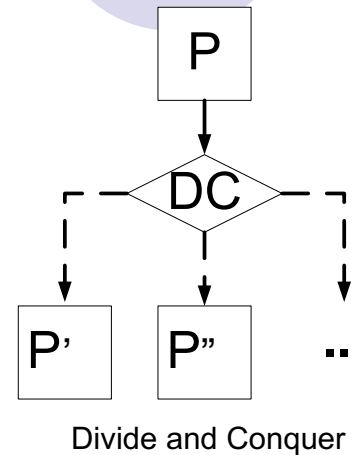
Below the screenshots, a diagram illustrates the iterative process. It shows a sequence of human icons. The first icon is labeled 'improvement \$0.05'. An arrow points to a group of three icons, with a label '3 votes @ \$0.01' pointing to them. This process repeats, showing a person submitting work, receiving votes, and then submitting an improved version.



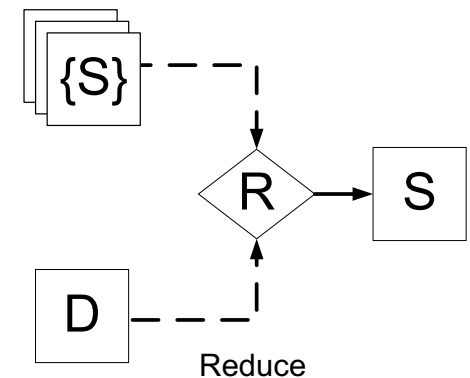
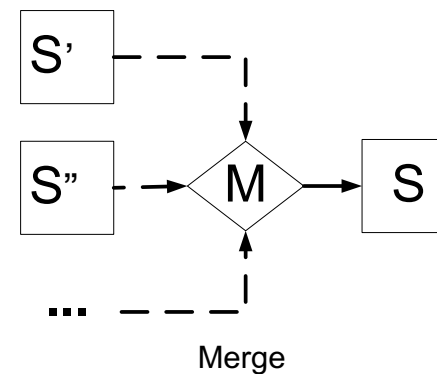
G. Little et al, "Exploring Iterative and Parallel Human Computation Processes," HCOMP 2010

Divide-and-Conquer and Aggregate

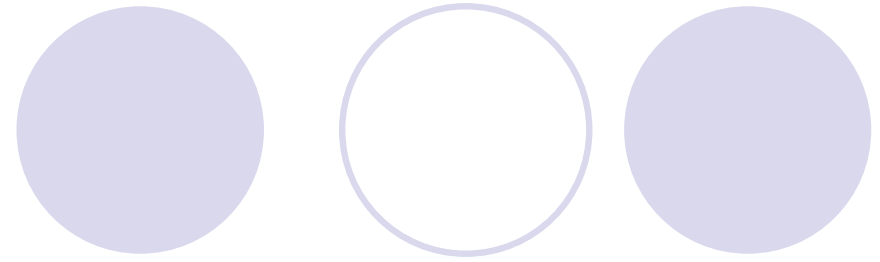
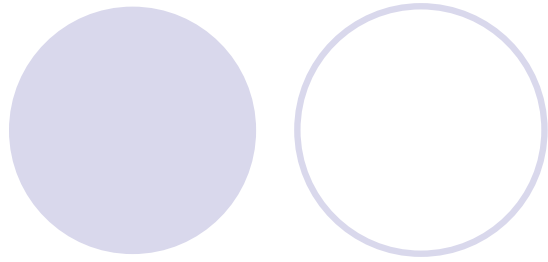
- Divide-and-Conquer and Aggregate
 - Decompose a problem statement and aggregate the results



- Two special aggregates
 - Merge
 - Reduce



P. Minder et al, "Crowdlang - First Steps Towards Programmable Human Computers for General Computation," AAI 2011.



Challenges

Opportunities

CHALLENGES AND OPPORTUNITIES

Challenges

Each set has $S/2$ items

r workers r

Each set has $S/10$ items

r r r r r r r r r r

- Trade-offs: time, cost, and quality

- Max algorithm with human error (with a probability)
- Maximize quality (via redundancy) subject to cost and time

P. Venetis et al, "Max Algorithms in Crowdsourcing Environments," WWW 2012

R. Kawajiri et al. "Steered Crowdsensing: Incentive Design Towards Quality-oriented Place-centric Crowdsensing", UBICOMP 2014

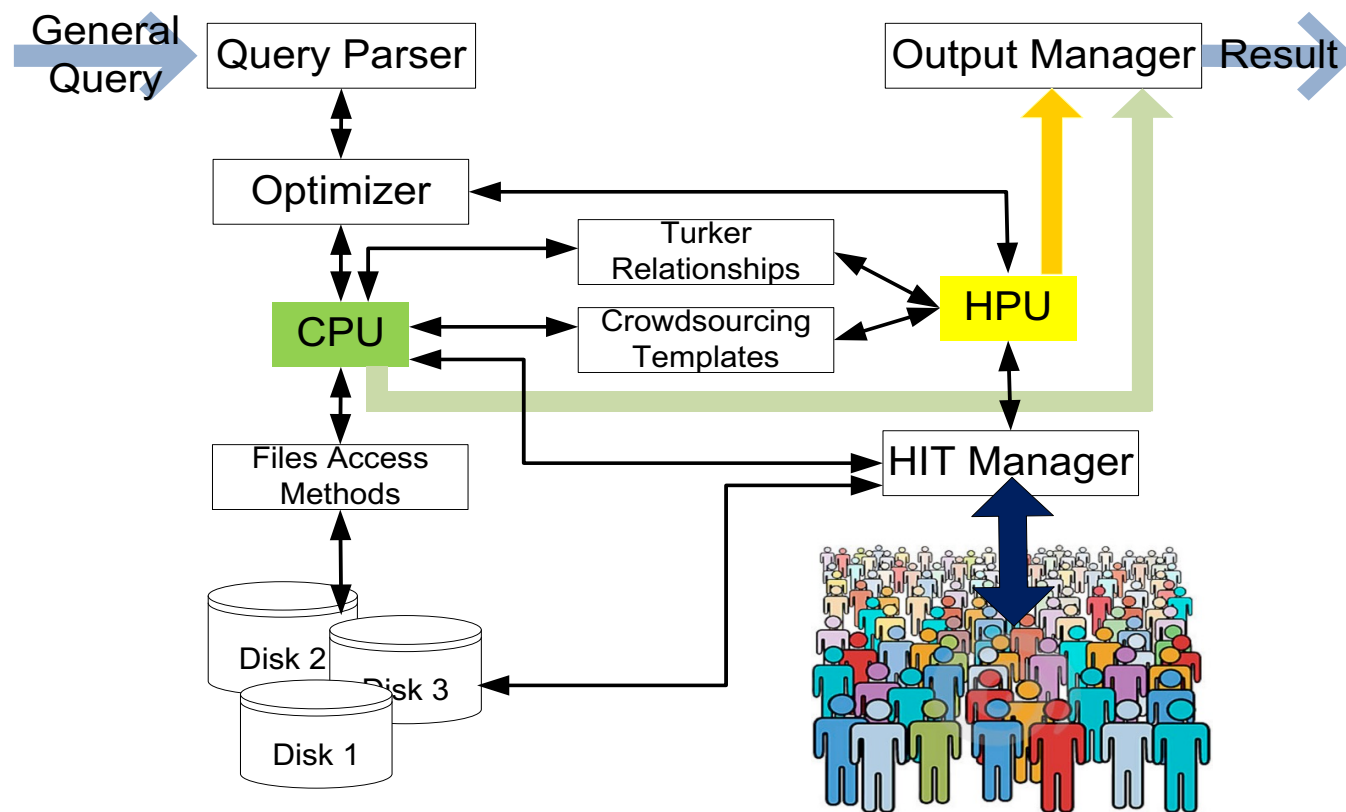
- Incentives: money, social recognition, and self esteem

- Platform-centric: a Stackelberg game
- User-centric: auction-based incentive mechanism

D. Yang et al, "Crowdsourcing to Smartphones: Incentive Mechanism Design for Mobile Phone Sensing," MobiCom 2012.

Challenges: HPU + CPU

- CrowdDB:

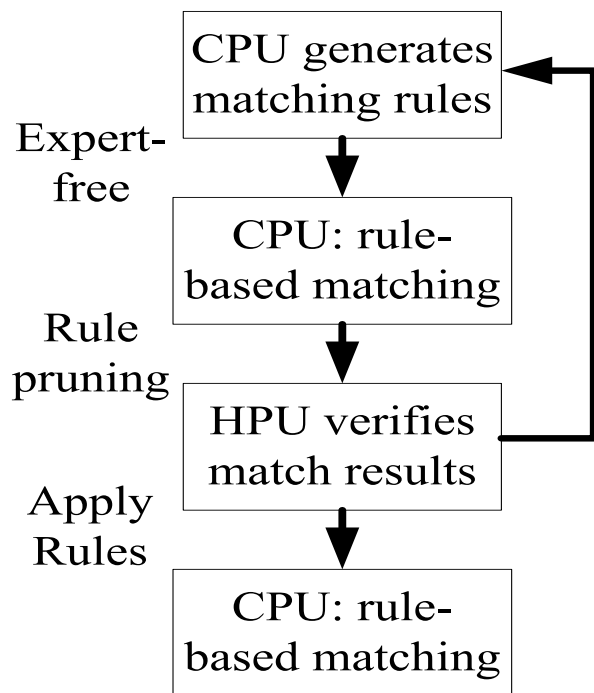


M. Franklin et al, "CrowdDB: Answering Queries with Crowdsourcing,"
SIGMOD 2011

CPU-assisted HPU

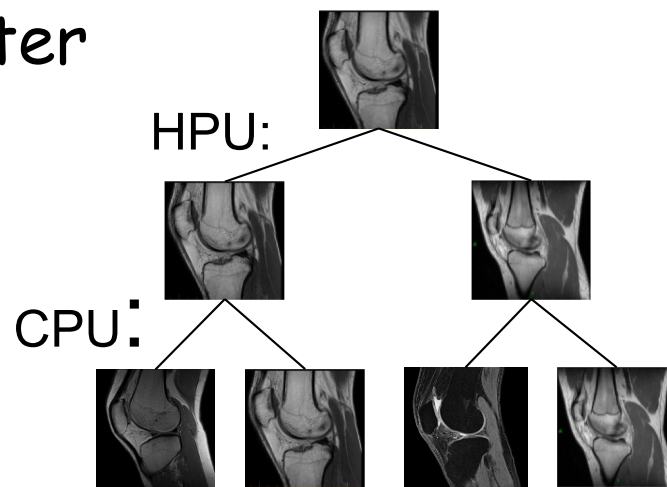
- Entity Matching

e.g. JHU matches John Hopkins Univ.

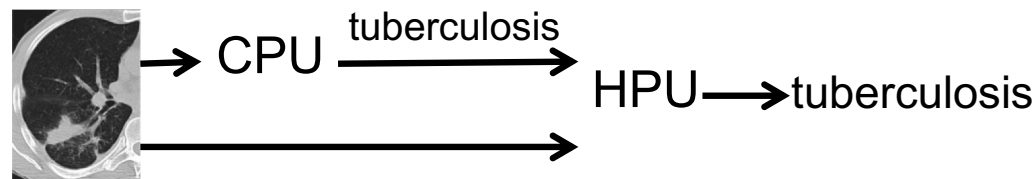


- Radiology

- Filter



- Diagnose



C. Gokhale et al, "Corleone: Hands-off Crowdsourcing for Entity Matching," SIGMOD 2014 Jilin Univ.

Opportunities



- Beyond simple workflows
 - Graph search
 - Graph match
- Beyond simple worker selection
 - Dynamic procurement
- Beyond independent workers
 - Social networks

Beyond Simple Workflows

Blend of bottom-up and open process with top-down organization goals

- **Graph search** (for macrowork)
 - Human-assisted graph search
 - Best sequence of questions with simple Y/N answers
 - A. Parameswaran et al, "Human-Assisted Graph Search: It's Okay to Ask Questions," VLDB 2010
- **Graph match**
 - People graph (who knows and/or communicates with whom)
 - Puzzle graph (ideas are compatible and can merge)
 - Natural dynamic for people to merge their compatible ideas
 - C. Brummitt et al, "Jigsaw Percolation: What Social Networks Can Collaboratively Solve a Puzzle," 2012

Beyond Simple Worker Selection

Dynamic Procurement (multi-armed bandit)

- A gambler facing a row of slot machines
- Which one to play, how many times, and in which order
- Each machine having a random reward from a fixed distribution
- Objective: maximizing the sum of rewards earned through a sequence of lever pulls



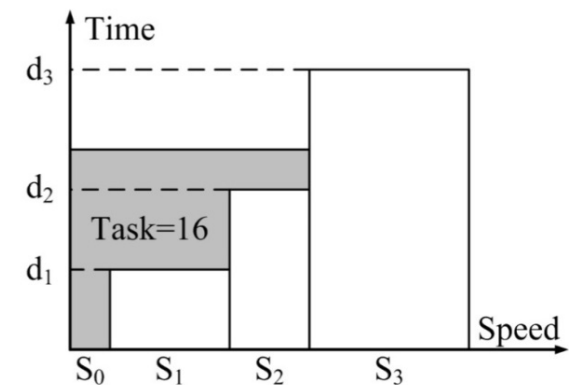
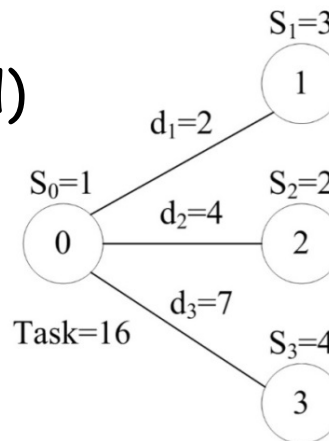
A. Badanidiyuru et al,
"Bandits with Knapsacks:
Dynamic Procurement for
Crowdsourcing," 2013

Beyond Independent Workers

Social network of workers

Recruitment of workers through **social ties**

- Friends help friends
 - Response delay
 - Computation (by a friend)
 - Reply delay
- Water-filling
 - d_i : response + reply



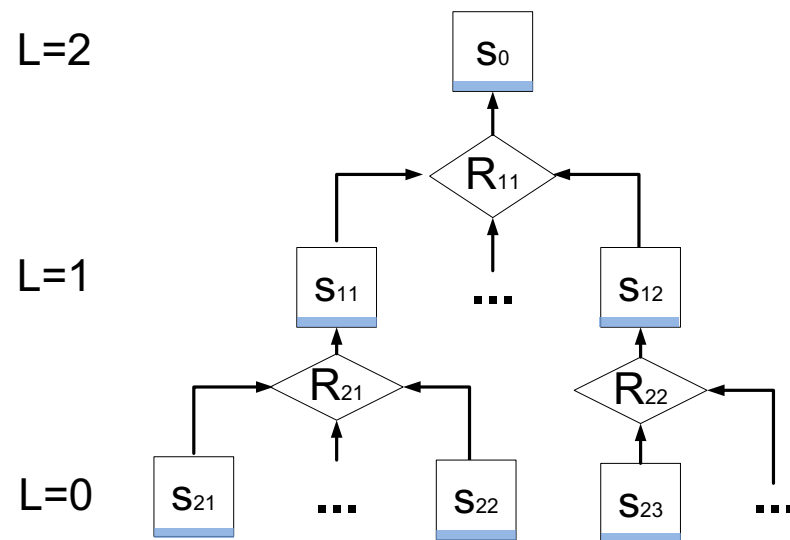
M. Xiao, J. Wu, L. Huang, Y. Wang, and C. Liu, "Multi-task Assignment for Crowdsensing in Mobile Social Networks," INFOCOM 2015

Survey: recursive (gather through reduce)

- Initial label is $L = "2"$ (subtract L by 1 when forwarding this request to QQ friends)

- When $L = 0$, return the total number of QQ friends

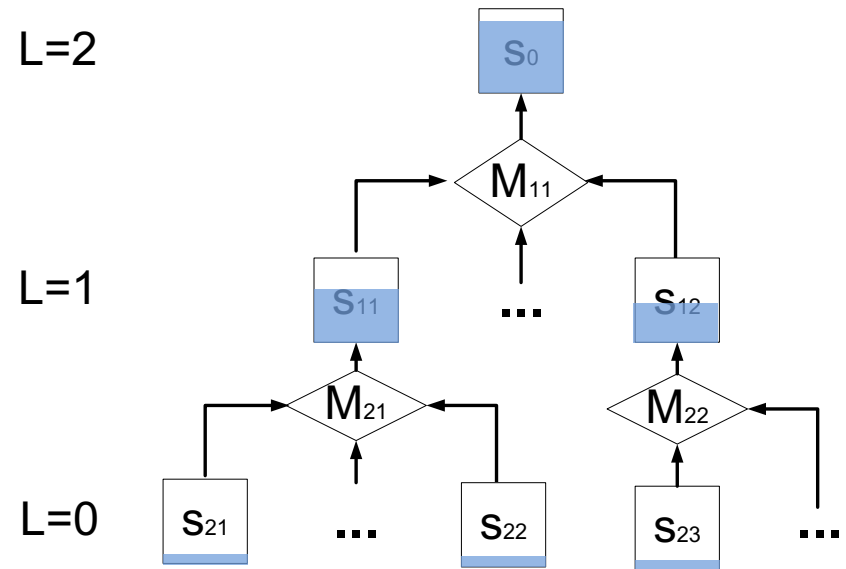
- When $L > 0$, do the following:
 - Forward this request to all QQ friends
 - After receiving the first 10 replies, compute the average number of friends, and send them back to me

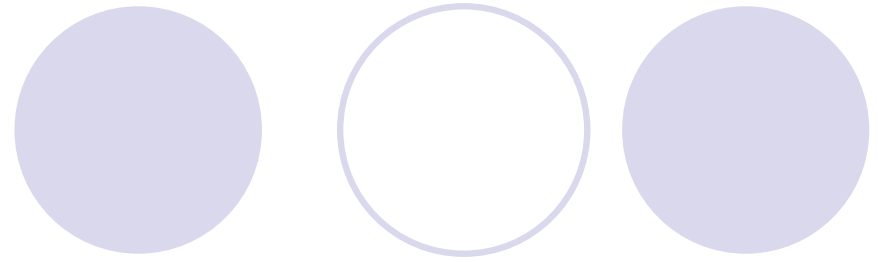
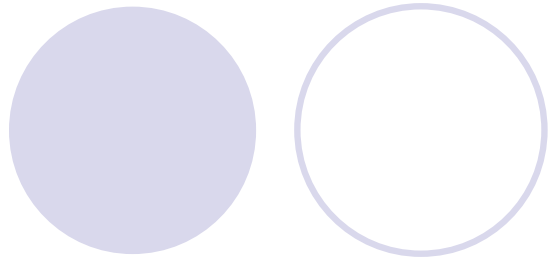


QQ: 200 million simultaneous online users

Survey: recursive (gather through merge)

- Initial label is $L = "2"$ (subtract L by 1 when forwarding this request to QQ friends)
- When $L = 0$, return the following:
 - Basic information (B)
 - Number of friends (N)
 - Timestamps (T)
- When $L > 0$, do the following:
 - Forward this request to all QQ friends
 - **Pack** the first 10 replies, together with your own information (B, N, T), and send them back to me





Summary

Acknowledgements

CONCLUSION



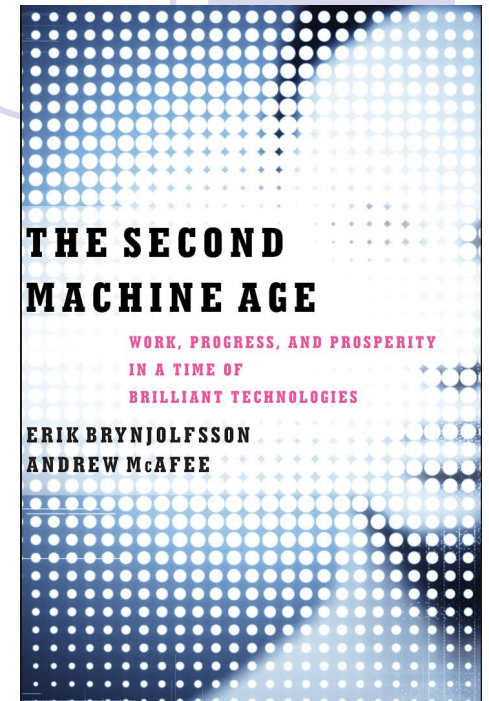
Summary

- HPU as a new paradigm to complement the traditional CPU-based computing for big data
- Many un(der)explored algorithmic problems
 - Social connections and proper training of workers
 - Workflow design
 - Cost-time-quality-uncertainty trade-offs
 - Incentive, gamification, and satisfaction mechanisms
 - Mobile crowdsourcing: energy consumption, communication cost, truthfulness, and privacy

Summary

Collective intelligence = networked brain +
massively connected & intelligent machine

- Finding ways to race *with* the machine rather than racing *against* it
- MTurk as “*artificial artificial intelligence*” as a way for people to race with machine



Acknowledgements

- Wei Chang
Saint Joseph's University
- Grace Ju
Carnegie Mellon University
- Wenjun Jiang
Hunan University
- Ying Dai
Google

