CROWDSOURCING FOR THE SEMANTIC WEB

FULL-DAY TUTORIAL ESWC2013

MONTPELLIER, FRANCE

TUTORIAL OVERVIEW

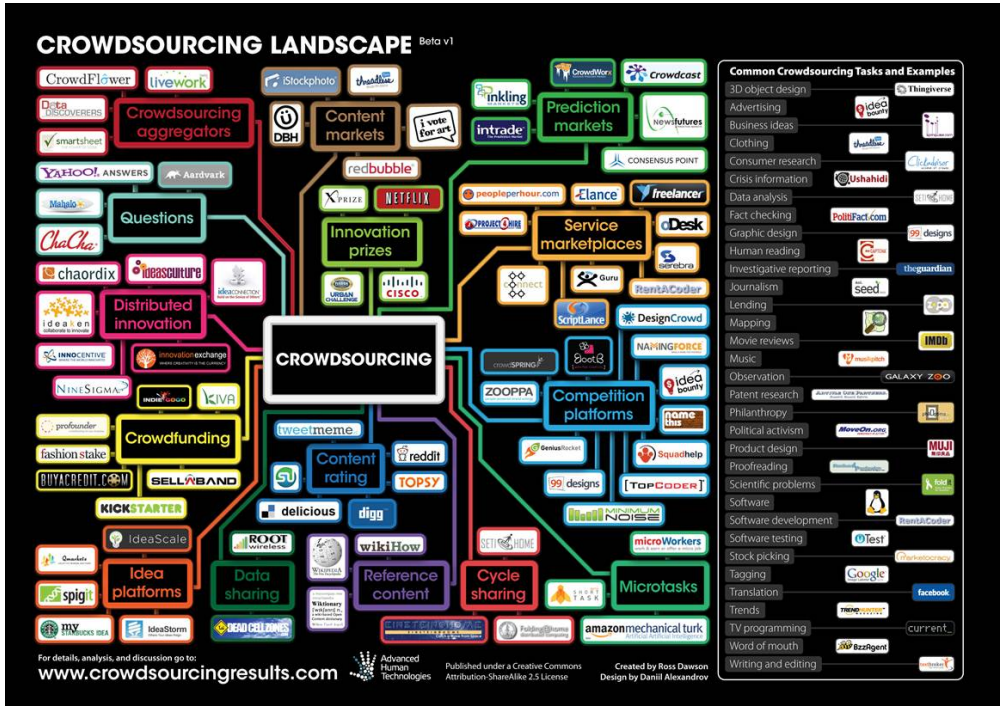
- **Semantic technologies are mainly about automation, but many Semantic Web tasks rely on human input**
  - Modeling a domain
  - Understanding text and media content (in all their forms and languages)
  - Integrating data sources originating from different contexts
- **Crowdsourcing offers an approach to solve Semantic Web tasks using human and computational intelligence**

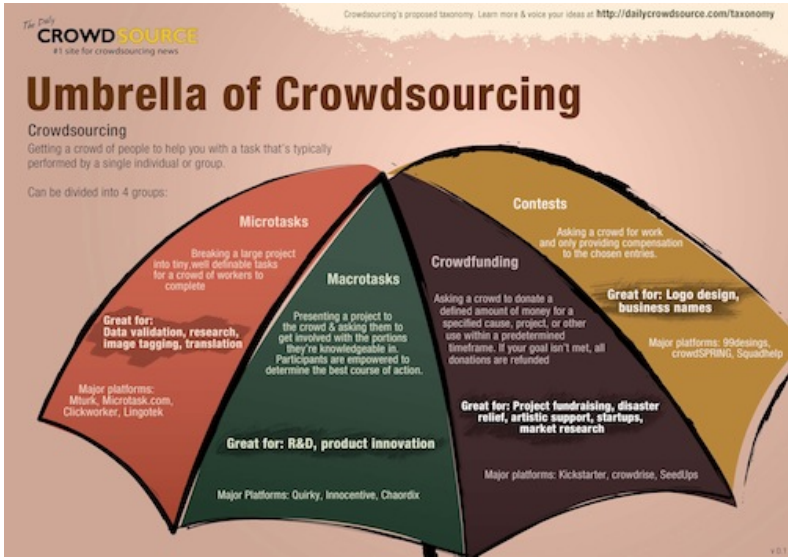
CROWDSOURCING: PROBLEM SOLVING VIA OPEN CALLS

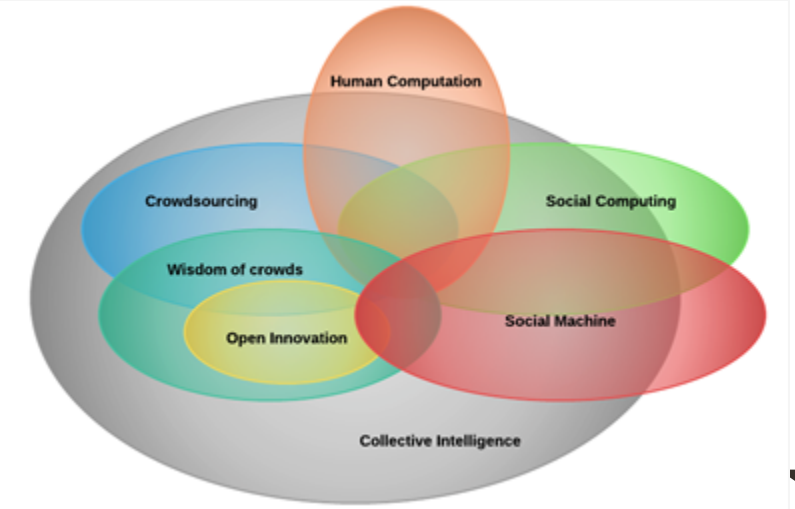
"Simply defined, crowdsourcing represents the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call. This can take the form of peer-production (when the job is performed collaboratively), but is also often undertaken by sole individuals. The crucial prerequisite is the use of the open call format and the large network of potential laborers."

[Howe, 2006]

CROWDSOURCING COMES IN DIFFERENT FORMS AND FLAVORS

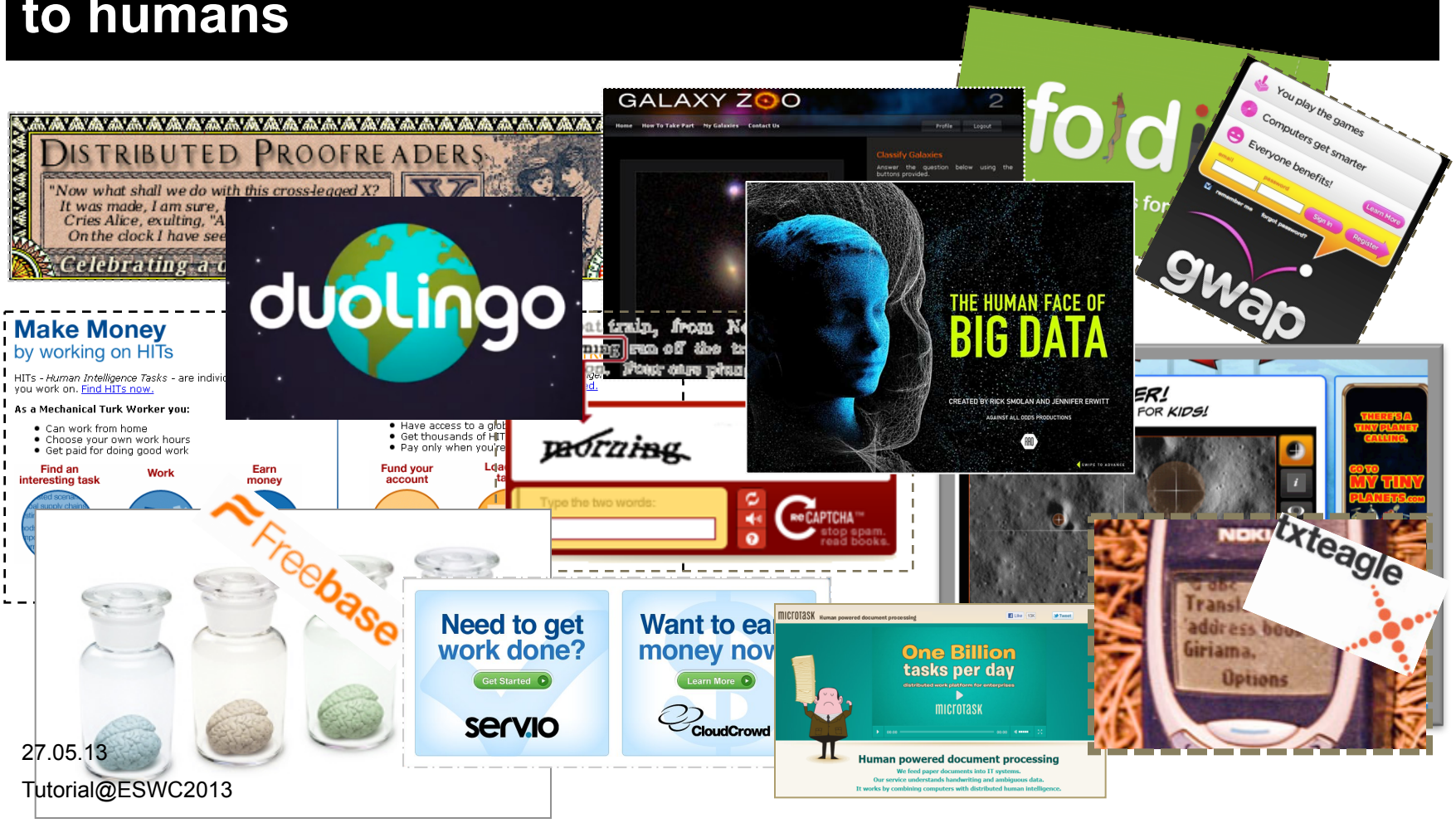






IN THIS TUTORIAL: CROWDSOURCING AS HUMAN COMPUTATION

Outsourcing tasks that machines find difficult to solve to humans

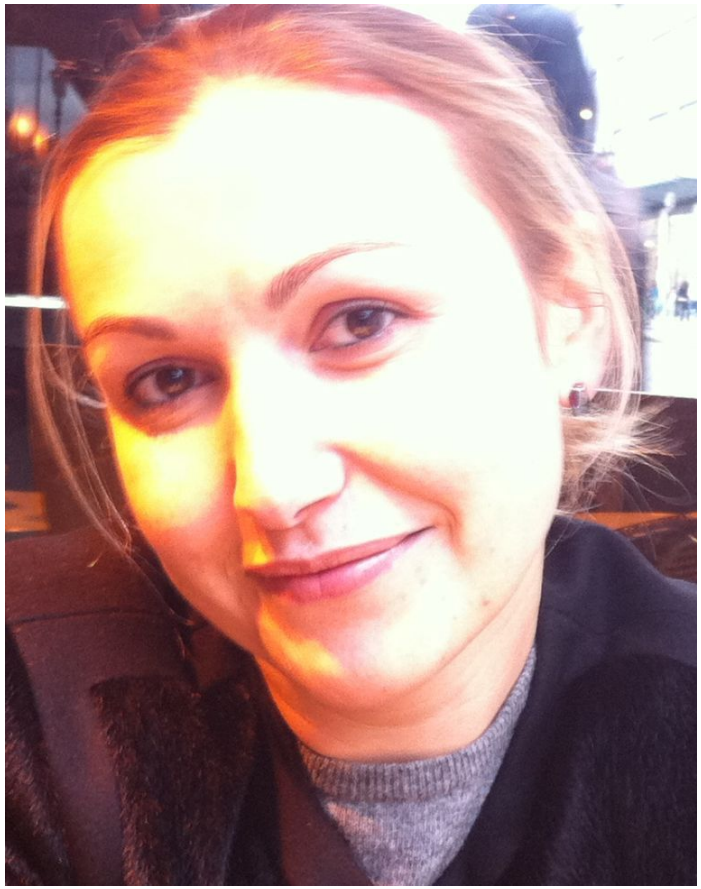


AGENDA FOR TODAY

Human computation fundamentals	09:30 – 10:30
Coffee break	10:30 – 11:00
Games with a purpose	11:00 – 11:30
Microtasks: management and automation	11:30 – 12:30
Lunch break	12:30 – 14:00
Microtasks: quality control	14:00 – 15:00
Human computation and the Semantic Web	15:00 – 15:30
Coffee break	15:30 – 16:00
Hands-out: Amazon Mechanical Turk	16:00 – 17:00
Wrap-up	17:00 – 17:30

PRESENTERS

- **Elena Simperl**
  - PhD in Computer Science FU Berlin, Germany
  - Worked for FU Berlin, Germany; STI Innsbruck, Austria; KIT, Germany
  - Senior lecturer WAIS, University of Southampton, UK
  - Coordinator of **Insemtives** project and tutorials at ESWC/ISWC
- **Research interests**
  - Social computing
  - Semantic technologies



e.simperl@soton.ac.uk

PRESENTERS

- **Gianluca Demartini**
  - MSc University of Udine, Italy
  - PhD University of Hannover, Germany on Entity Retrieval
  - Worked for UC Berkeley (on crowdsourcing);Yahoo! Research, Spain; L3S Research Center , Germany
  - PostDoc eXascale Infolab, University Fribourg, Switzerland
  - Lecturer for Social Computing in Fribourg
  - Tutorial on Entity Search at ECIR 2012
- **Research interests**
  - Information retrieval
  - Social and Semantic Web

demartini@exascale.info

PRESENTERS

- **Maribel Acosta**
  - MSC in Computer Science University Simon Bolivar, Venezuela
  - Worked for University Simon Bolivar, Venezuela
  - PhD student, KIT, Germany
- **Research interests**
  - Data base management
  - Linked Data query processing
  - Social Web

maribel.acosta@kit.edu

ALL MATERIALS AVAILABLE AT HTTPS://SITES.GOOGLE.COM/SITE/CROWDSOURCINGTUTORIAL/

NEXT EVENTS

TUTORIAL ON MICROTASKS TO SOLVE SEMANTIC WEB PROBLEMS

WORKSHOP ON CROWDSOURCING AND THE SEMANTIC WEB (PAPERS DUE JULY 12)

AIFB KIT Karlsruhe Institute of Technology

UNIVERSITAS FRIBURGENSIS

University of Southampton

eXascale Infolab

HUMAN COMPUTATION FUNDAMENTALS

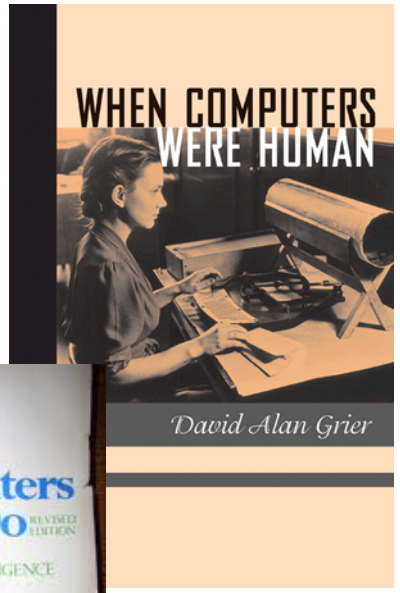
ELENA SIMPERL

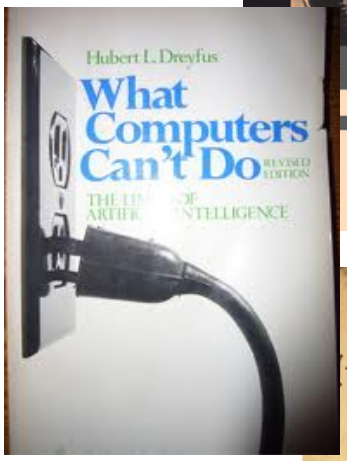
UNIVERSITY OF SOUTHAMPTON

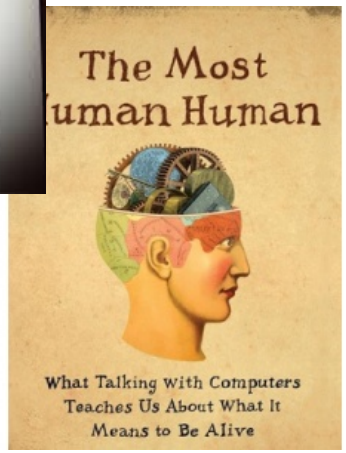
HUMAN COMPUTATION REVISITED

- **Outsourcing tasks that machines find difficult to solve to humans**
  - Difficult not the same as impossible
  - Accuracy, efficiency, cost

- **Historically humans were the first computers**
  - 17th century: Halley‘s commet
  - 19th century: computing factories
  - 20th century: professionalization of human computation
  - *Characteristics: division of labor, redundancy, multiple methods to find or check the correctness of a solution*







DIMENSIONS OF HUMAN COMPUTATION

DIMENSIONS OF HUMAN COMPUTATION

See also [Quinn & Bederson, 2012]

WHAT IS OUTSOURCED

- Tasks based on human skills not easily replicable by machines (visual recognition, language understanding, knowledge acquisition, basic human communication etc)

WHO IS THE CROWD

- Open call
- Call may target specific skills and expertise
- Requester typically knows less about the workers than in other work environments

HOW IS THE TASK OUTSOURCED

- Explicit vs. implicit participation
- Tasks broken down into smaller units undertaken in parallel by different people
- Coordination required to handle cases with more complex workflows
- Partial or independent answers consolidated and aggregated into complete solution

EXAMPLE: CITIZEN SCIENCE VIA HUMAN COMPUTATION

WHAT IS OUTSOURCED

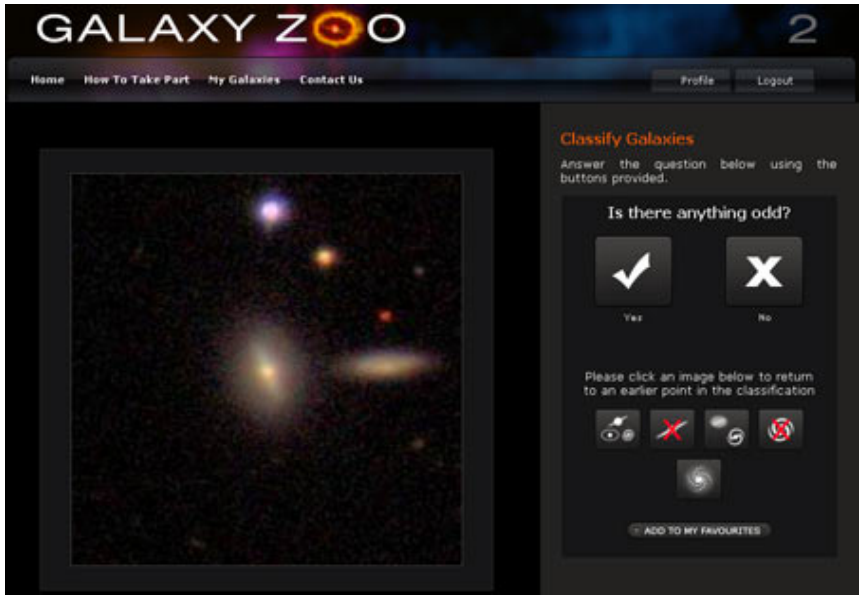
- Object recognition, labeling, categorization in media content

WHO IS THE CROWD

- Anyone

HOW IS THE TASK OUTSOURCED

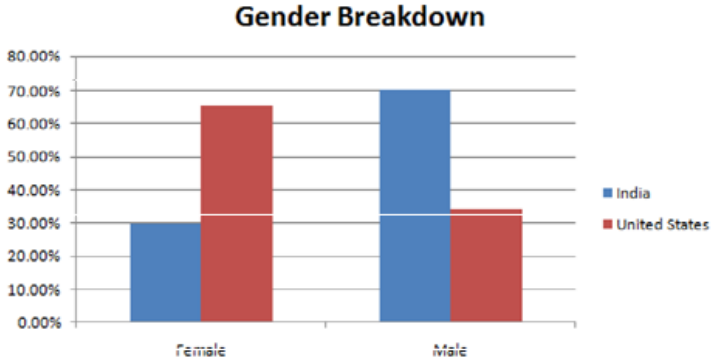
- Highly parallelizable tasks
- Every item is handled by multiple annotators
- Every annotator provides an answer
- Consolidated answers solve scientific problems

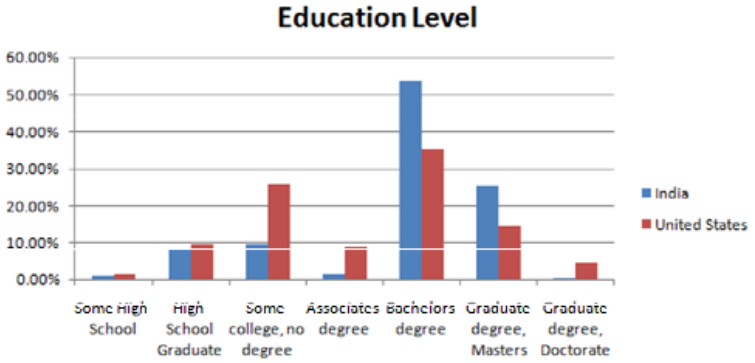


A LARGE, BUT NOT ALWAYS DIVERSE CROWD

Country of residence

- **United States: 46.80%**
- **India: 34.00%**
- **Miscellaneous: 19.20%**





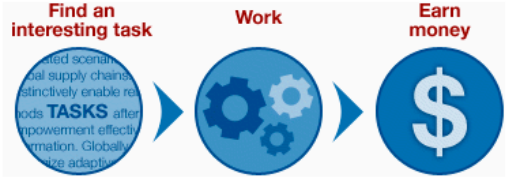
amazon mechanical turk
Artificial Artificial Intelligence

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

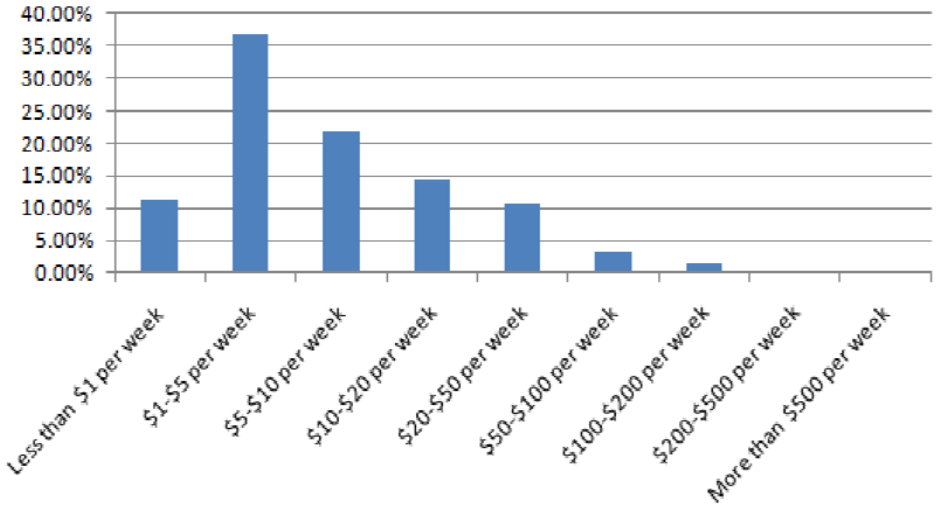
Ask workers to complete HITs - *Human Intelligence Tasks* - and get results using Mechanical Turk. Register Now

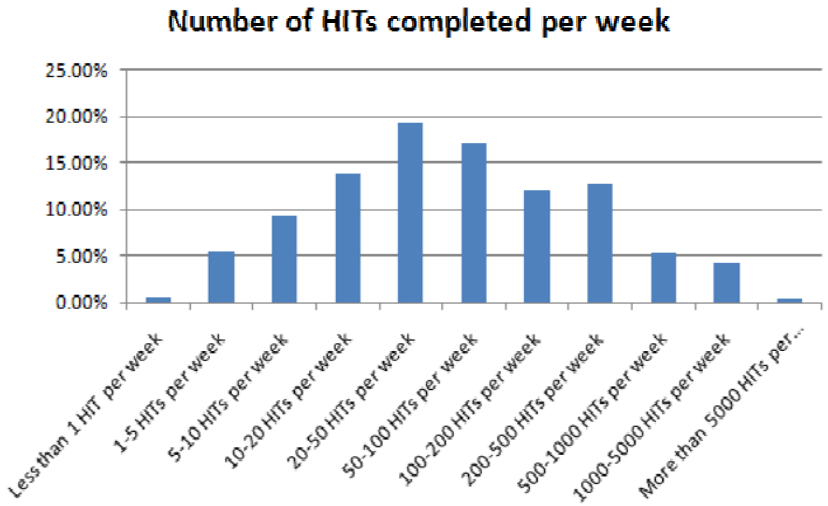
As a Mechanical Turk Requester you:

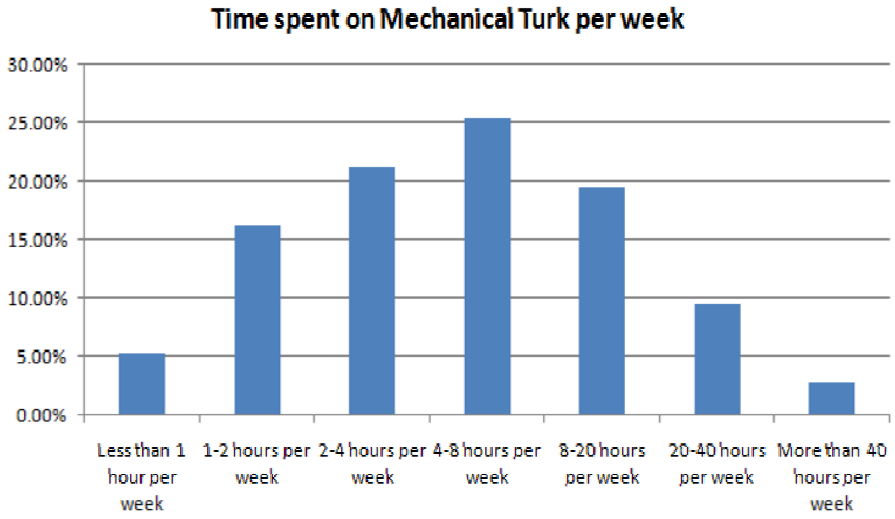
- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of HITs completed in minutes
- Pay only when you're satisfied with the results

60% OF WORKERS SPEND MORE THAN 4 HOURS A WEEK ON MTURK

http://www.mturk-tracker.com/







SIGNIFICANT AMOUNT OF RESOURCES AND TIMELY DELIVERY

CrowdFlower

BROAD RANGE OF TASKS

CrowdFlower

Business Data

Collect data on businesses at massive scale

Content Moderation and Curation

Quickly find both good and bad user generated content

Ranked

Boost conversions with better search results

Content Generation

Improve your search engine ranking with quality content

Custom solutions

We help businesses of all sizes automate really big custom projects

Customer and Lead Data Enhancement

Increase sales by knowing more about your customers

Sentiment and Opinion Analysis

Know exactly what people are saying about you

Categorize

Categorize products, businesses, videos, events, & more

Surveys

Find and interact with highly-qualified digital consumers

Builder

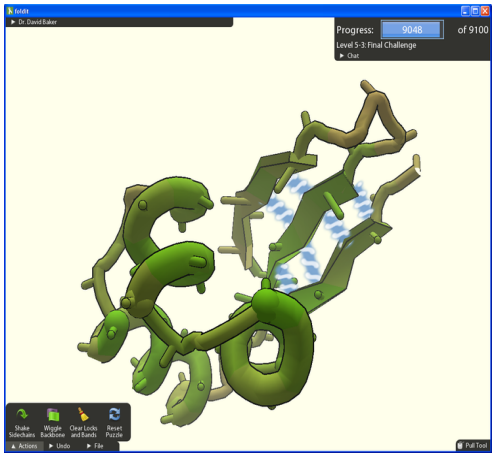
Advanced user? Developer? Build your own crowdsourcing projects

EXPLICIT VS. IMPLICIT CONTRIBUTION - AFFECTS MOTIVATION AND ENGAGEMENT

Users aware of how their input contributes to the achievement of application's goal (and identify themselves with it)

vs.

Tasks are hidden behind the application narratives. Engagement ensured through other incentives



COMPLEX WORKFLOWS CANNOT ALWAYS BE DIRECTLY IMPLEMENTED

See also [Bernstein et al., 2010]

WHAT IS OUTSOURCED

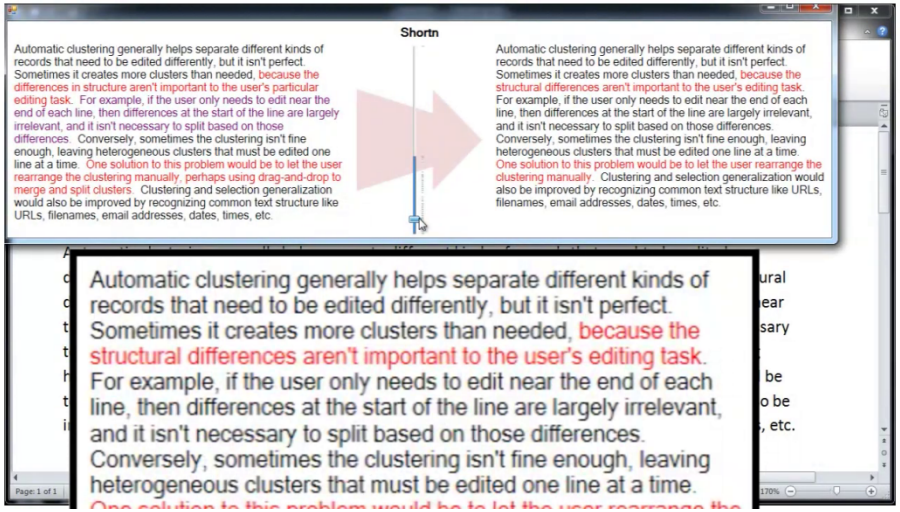
- Text shortening, proof-reading, open editing

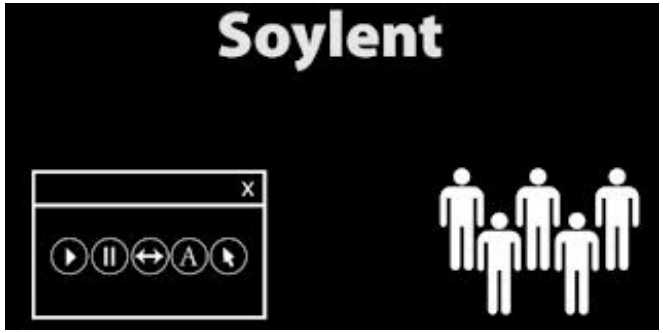
WHO IS THE CROWD

- MTurk

HOW IS THE TASK OUTSOURCED

- Text divided into paragraphs
- Select-fix-verify pattern
- Multiple workers in each step





http://www.youtube.com/watch?v=n_miZqsPwsc

DIMENSIONS OF HUMAN COMPUTATION (2)

See also [Quinn & Bederson, 2012]

HOW ARE THE RESULTS VALIDATED

- Solutions space closed vs. open
- Performance measurements/ground truth
- Statistical techniques employed to predict accurate solutions
  - May take into account confidence values of algorithmically generated solutions

HOW CAN THE PROCESS BE OPTIMIZED

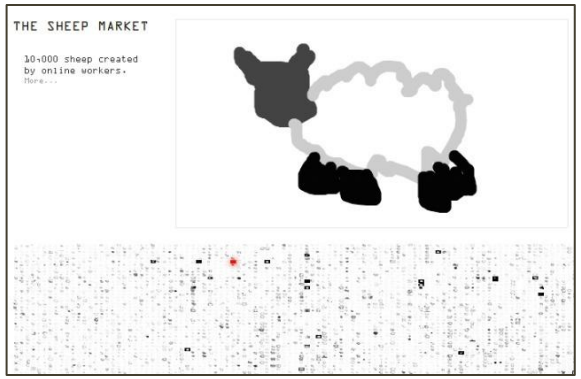
- Incentives and motivators
- Assigning tasks to people based on their skills and performance (as opposed to random assignments
- *Symbiotic combinations of human- and machine-driven computation, including combinations of different forms of crowdsourcing*

OPEN SOLUTION SPACES

Selecting the right option vs. assessing the quality of the work

The goal is to undertake much of the assessment either automatically or use the crowd for it.

'draw a sheep facing to the left.'



CloudCrowd
We're working on it. Lots of us.

Upcoming Credentialed Work

Last Updated: May 16, 2013 11:07AM PST

Start Date	Brief Description	Required Credential	Number of Tasks	Pay Rate
Current (on-going)	150 to 300-word Small Business Product Descriptions	Marketing Writer 2	100+ Remaining	5 Cents (*Per Word*)
Current (on-going)	475-word Reply! Articles	General Writer 1	400+ Remaining	3 Cents (*Per Word*)
Current (on-going)	65 to 75-word Medical Specialty Blurbs	General Writer 2	4000+	5 Cents (*Per Word*)
Current (on-going)	500- to 750-word movie reviews and analyses.	General Writer 1	36/week	3 Cents (*Per Word*)
Current (on-going)	450-word Small Business Blog Articles	Marketing Writer 2	40/month	5 Cents (*Per Word*)
Current (on-going)	300-word Marketing Content.	General Writer 1	100+	3 Cents (*Per Word*)
Current (test-run)	Costume/Accessory PDs.	Marketing Writer 1	300 remaining	3 Cents (*Per Word*)
Current (on-going)	Short Leadins.	General Writer 1	500/week	3 Cents (*Per Word*)
Current (on-going)	PPQAs	General Writer 1	20,000 (late May/early June)	3 Cents (*Per Word*)
Current (on-going)	Visa Print PDs	Marketing Writer 1	5000 Remaining	3 Cents (*Per Word*)

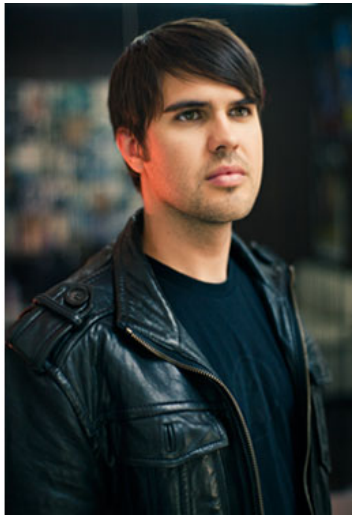
Example Categorization Questions

What type of business is this?
Bank of America
- (•) Financial Institute
- ( ) Retailer
- ( ) Resturaunt
- ( ) Other

Does this blog comment meet our guidelines?
"No way! You're crazy to think that"
- ( ) Yes
- (•) No

What emotion is shown in this picture?

- ( ) Happiness
- ( ) Anger
- (•) Sadness
- ( ) Surprise



MEASURING PERFORMANCE CAN BE CHALLENGING

WHO AND HOW

- Redundancy
- Excluding spam and obviously wrong answers
- Voting and ratings by the crowd
- Assessment by the requester
- Where does the ground truth come from and is it needed
  - Note: improving recall of algorithms

WHEN

- Real-time constraints in games
- Near-real-time microtasks, see Bernstein et al. Crowds in Two Seconds: Enabling Realtime Crowd-Powered Interfaces. In Proc. UIST 2011.

ALIGNING INCENTIVES IS ESSENTIAL

Motivation: driving force that makes humans achieve their goals

Incentives: 'rewards' assigned by an external 'judge' to a performer for undertaking a specific task

- Common belief (among economists): incentives can be translated into a sum of money for all practical purposes.

Incentives can be related to both extrinsic and intrinsic motivations.

Extrinsic motivation if task is considered boring, dangerous, useless, socially undesirable, dislikable by the performer.

Intrinsic motivation is driven by an interest or enjoyment in the task itself.

IN THIS TUTORIAL

GAMES WITH A PURPOSE

MICROTASKS

HYBRID SYSTEMS

GAMES WITH A PURPOSE (GWAP)

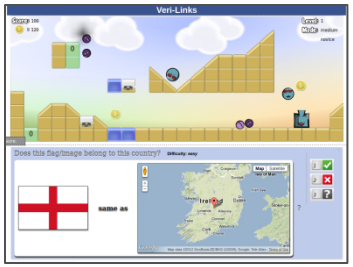
See also [van Ahn & Dabbish, 2008]

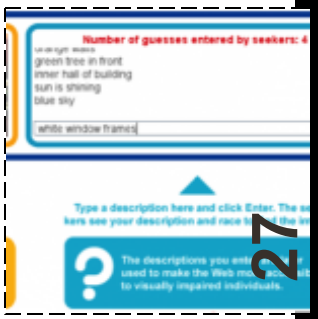
Human computation disguised as casual games

Tasks are divided into parallelizable atomic units (challenges) solved (consensually) by players

Game models

- Single vs. multi-player
- Selection agreement vs. input agreement vs. inversion-problem games

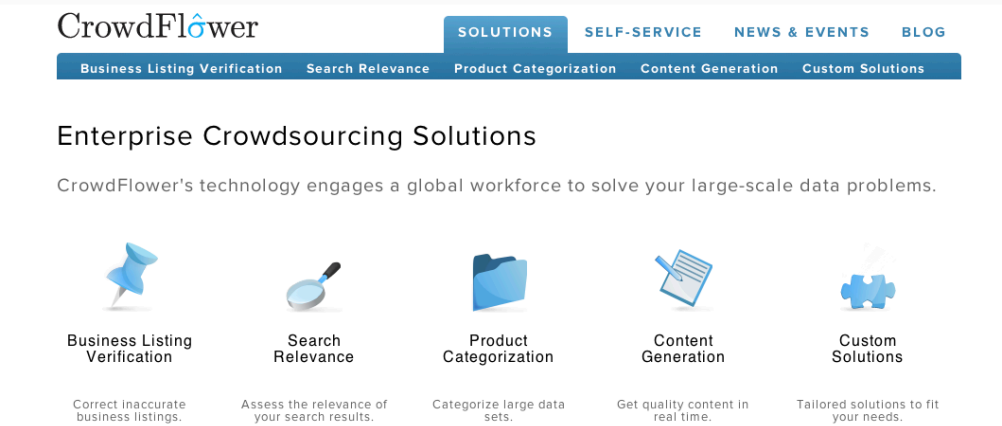


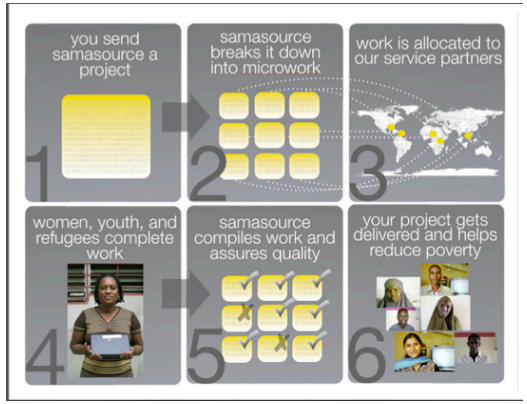


MICROTASK CROWDSOURCING

Similar types of tasks, but different incentives model (monetary reward)

Successfully applied to transcription, classification, and content generation, data collection, image tagging, website feedback, usability tests…

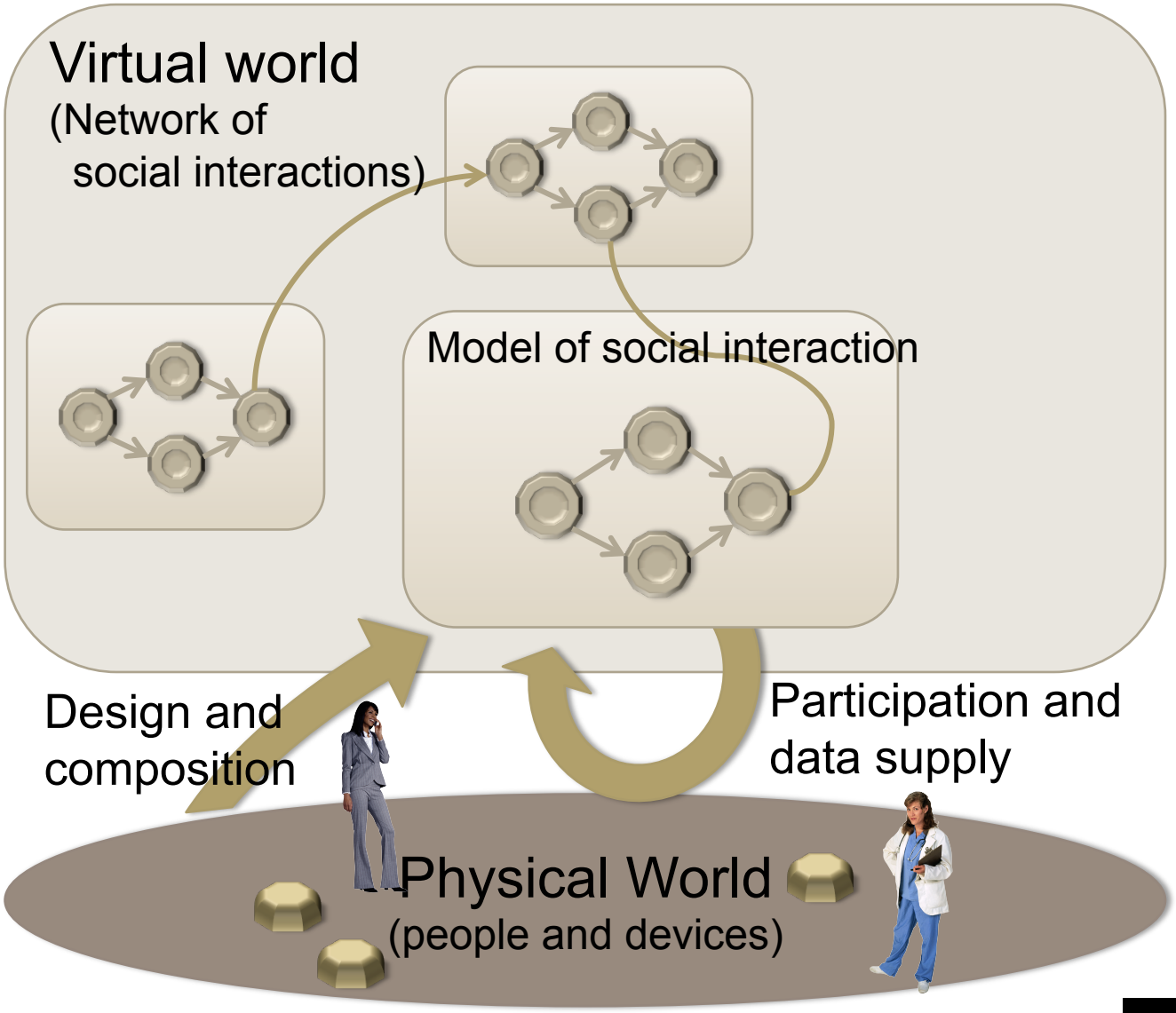




THE SAME, BUT DIFFERENT

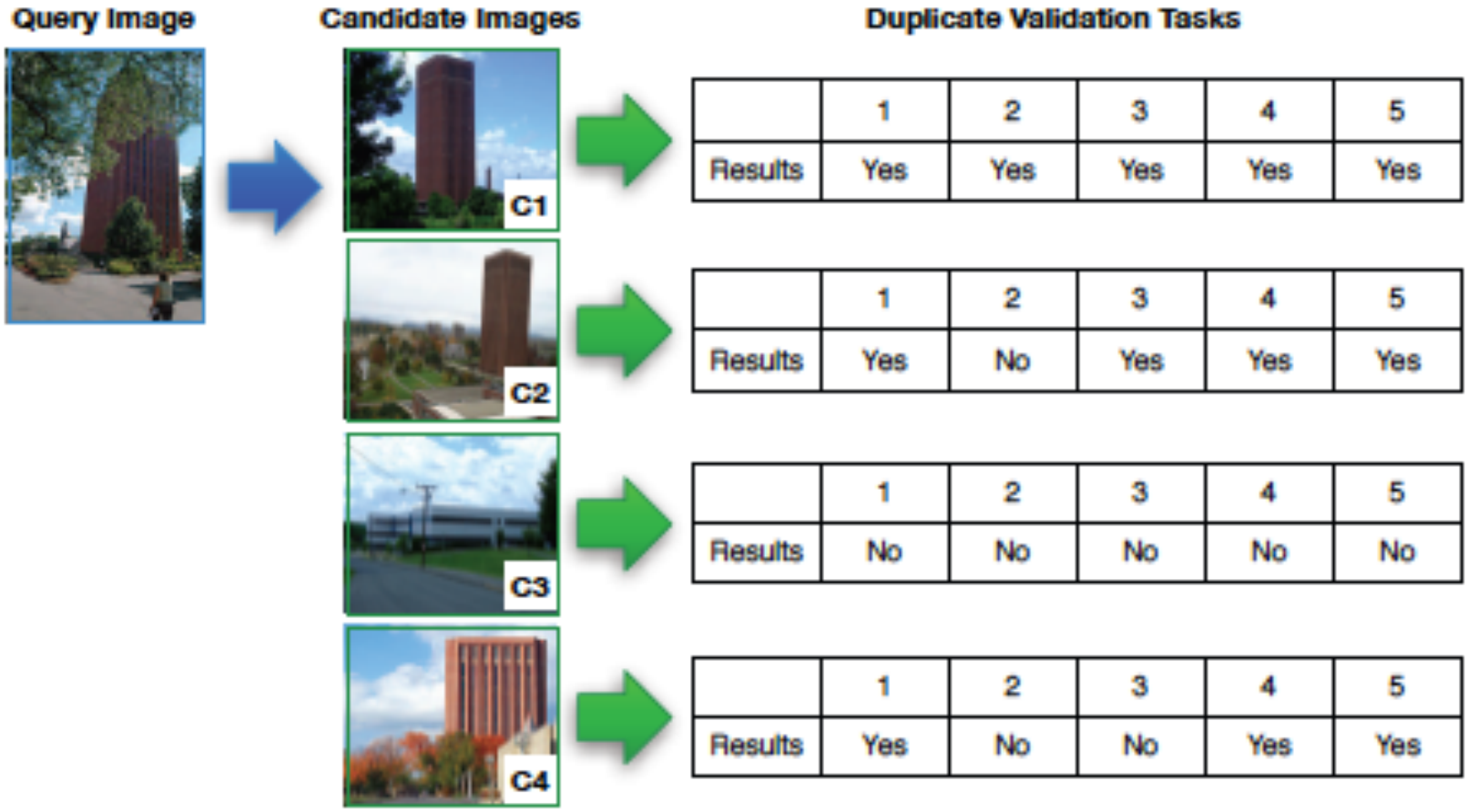
- **Tasks leveraging common human skills, appealing to large audiences**
  - Selection of domain and task more constrained in games to create typical UX
- **Tasks decomposed into smaller units of work to be solved independently**
- **Complex workflows**
  - Creating a casual game experience vs. patterns in microtasks
- **Quality assurance**
  - Synchronous interaction in games
  - Levels of difficulty and near-real-time feedback in games
  - Many methods applied in both cases (redundancy, votes, statistical techniques)
- **Different set of incentives and motivators**

HYBRID SYSTEMS



Dave Robertson

EXAMPLE: HYBRID IMAGE SEARCH



	1	2	3	4	5
Results	Yes	Yes	Yes	Yes	Yes

	1	2	3	4	5
Results	Yes	No	Yes	Yes	Yes

	1	2	3	4	5
Results	No	No	No	No	No

	1	2	3	4	5
Results	Yes	No	No	Yes	Yes

Yan, Kumar, Ganesan, CrowdSearch: Exploiting Crowds for Accurate Real-time Image Search on Mobile Phones, Mobisys 2010.

EXAMPLE: HYBRID DATA INTEGRATION

paper	conf
Data integration	VLDB-01
Data mining	SIGMOD-02

title	author	email	venue
OLAP	Mike	mike@a	ICDE-02
Social media	Jane	jane@b	PODS-05

Generate plausible matches

- paper = title, paper = author, paper = email, paper = venue
- conf = title, conf = author, conf = email, conf = venue

Ask users to verify

Does attribute paper match attribute author?

paper	conf
Data integration	VLDB-01
Data mining	SIGMOD-02

title	author	email
OLAP	Mike	mike@a
Social media	Jane	jane@b

Yes No Not sure

McCann, Shen, Doan: Matching Schemas in Online Communities. ICDE, 2008

EXAMPLE: HYBRID QUERY PROCESSING

Use the crowd to answer DB-hard queries

Where to use the crowd:

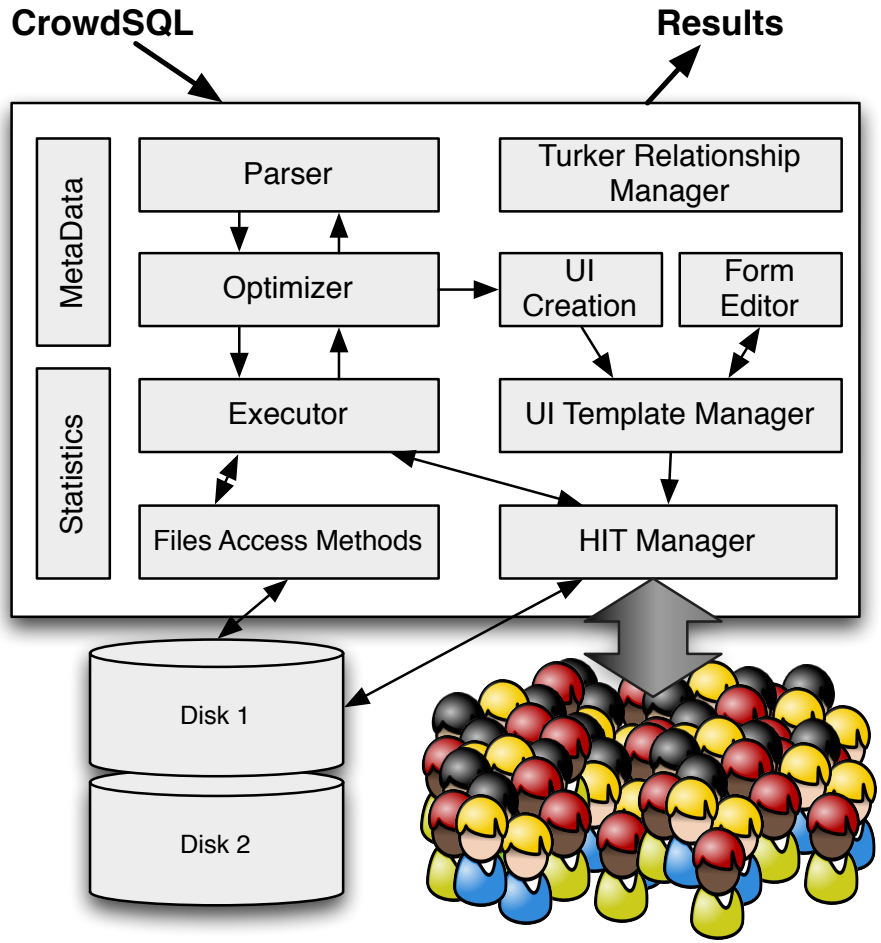
Find missing data

Make subjective comparisons

Recognize patterns

But not:

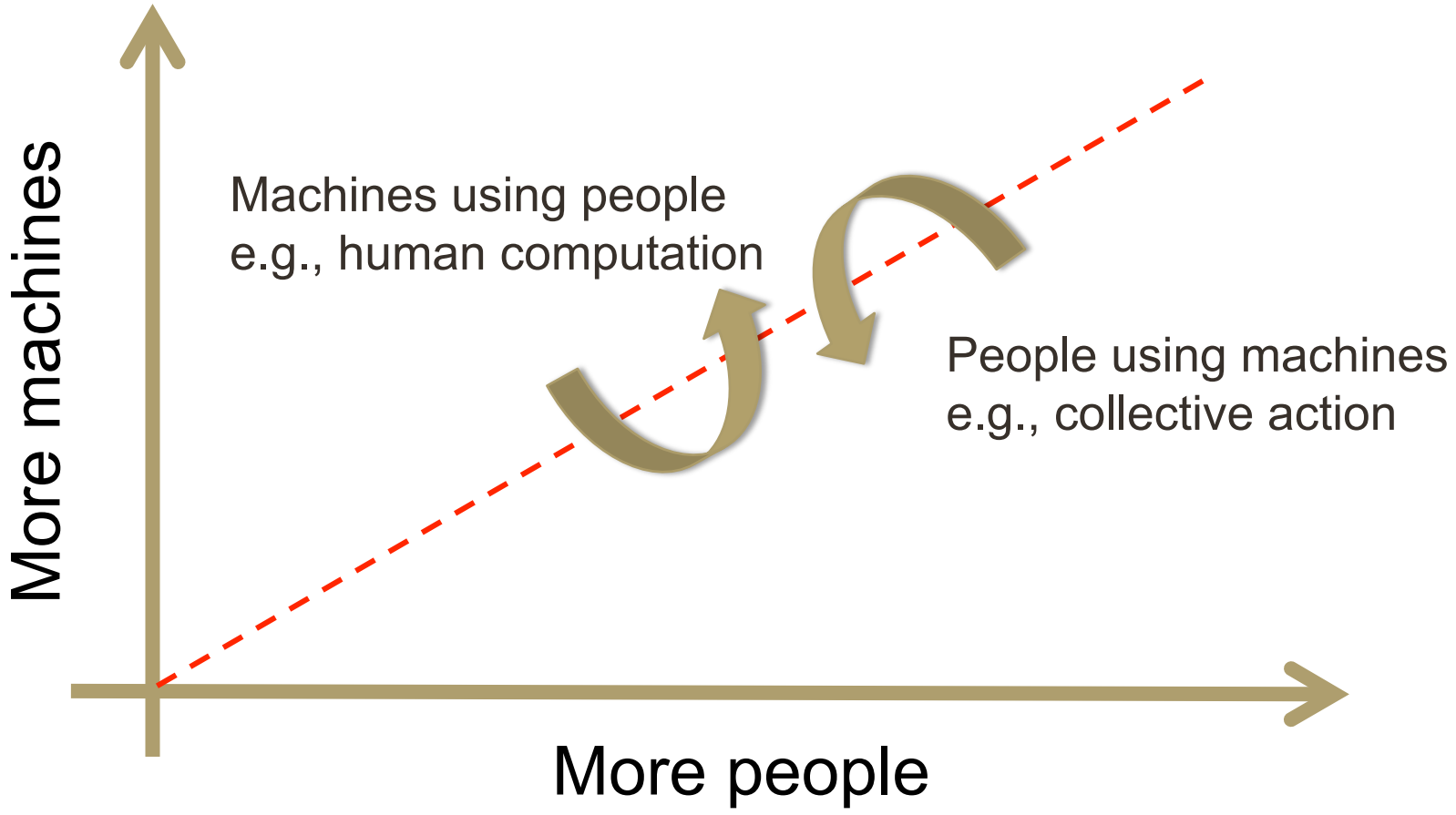
Anything the computer already does well



M. Franklin, D. Kossmann, T. Kraska, S. Ramesh and R. Xin . CrowdDB: Answering Queries with Crowdsourcing, SIGMOD 2011

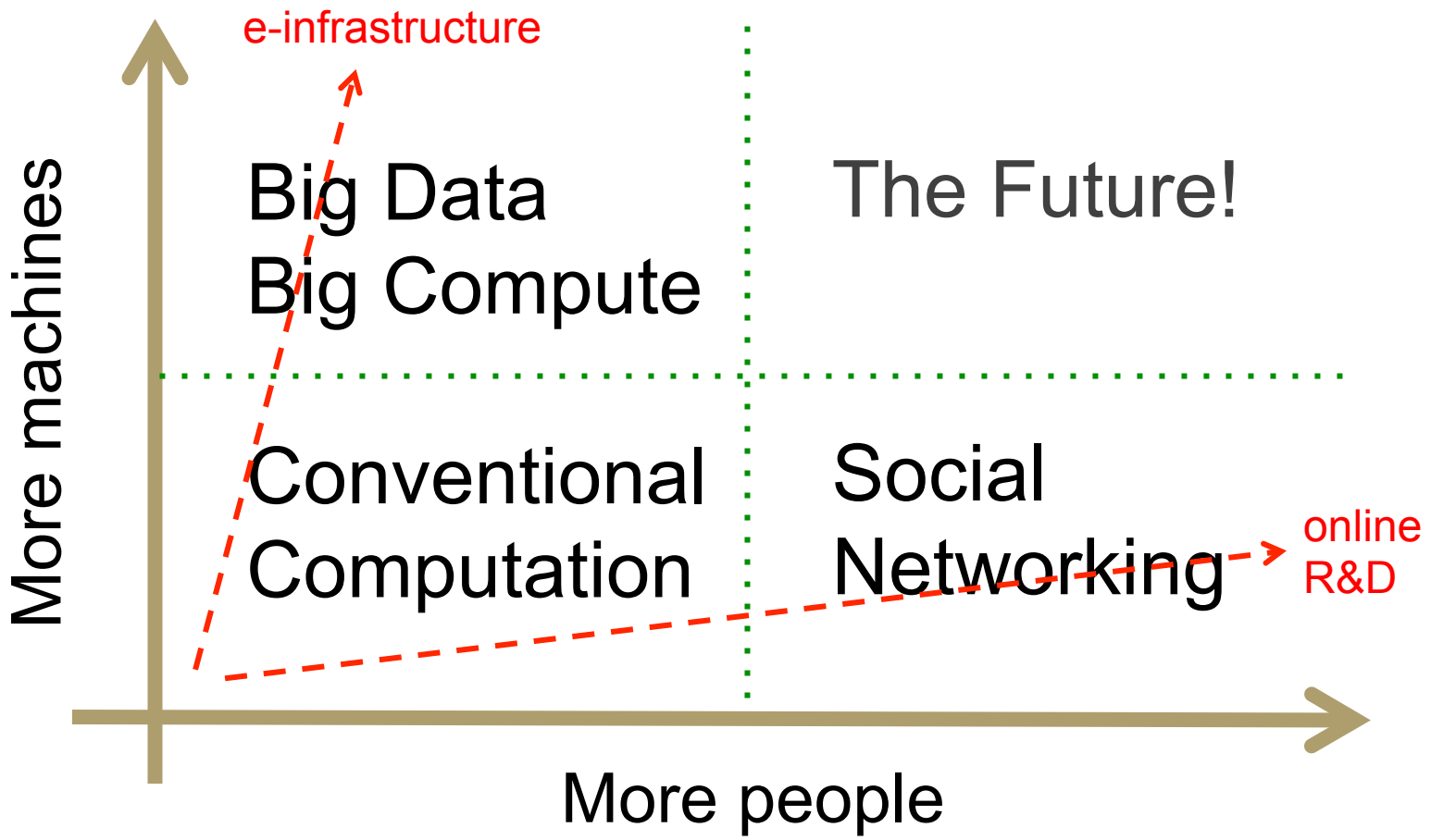
WHAT'S NEXT

THE BIGGER PICTURE



Dave de Roure

OPEN QUESTION: HOW TO BUILD SOCIAL SYSTEMS AT SCALE?



Dave de Roure

GAMES WITH A PURPOSE

ELENA SIMPERL

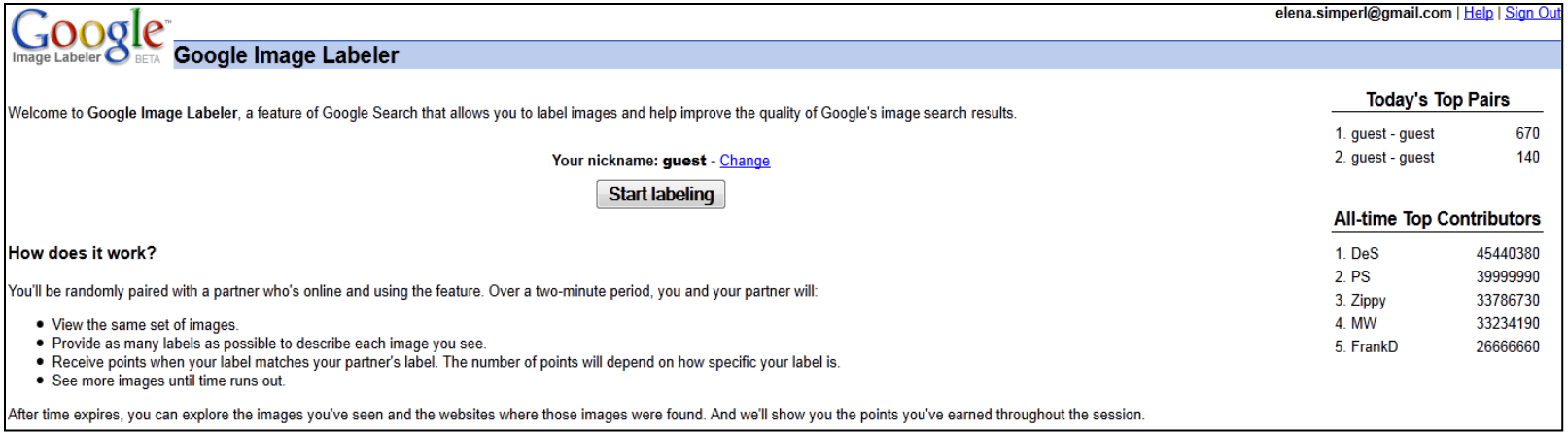
UNIVERSITY OF SOUTHAMPTON

GAMES AND GAMIFICATION

GAMES WITH A PURPOSE (GWAP)

„ a human-based computation technique in which a computational process performs its function by outsourcing certain steps to humans in an entertaining way"

[Wikipedia]



RELATED: GAMIFICATION

“use of game play mechanics for non-game applications […] in order to encourage people to adopt the applications”

[Wikipedia]

Image from http://gapingvoid.com/2011/06/07/pixie-dust-the-mountain-of-mediocrity/

HOW TO IMPLEMENT GAMIFICATION*

- **Cosmetic: adding game-like visual elements or copy (usually visual design or copy-driven)**
- **Accessory: wedging in easy-to-add-on game elements, such as badges or adjacent products (usually marketing-driven)**
- **Integrated: more subtle, deeply integrated elements like % complete (usually interaction-design driven)**
- **Basis: making the entire offering a game (usually product-driven)**

*** http://uxmag.com/design/a-gamification-framework-for-interaction-designers**

GAMIFICATION FEATURES*

- **Accelerated feedback cycles**
  - Annual performance appraisals vs immediate feedback to maintain engagement.
- **Clear goals and rules of play**
  - Players feel empowered to achieve goals vs fuzzy, complex system of rules in real-world.
- **Compelling narrative**
  - Gamification builds a narrative that engages players to participate and achieve the goals of the activity

***http://www.gartner.com/it/page.jsp?id=1629214**

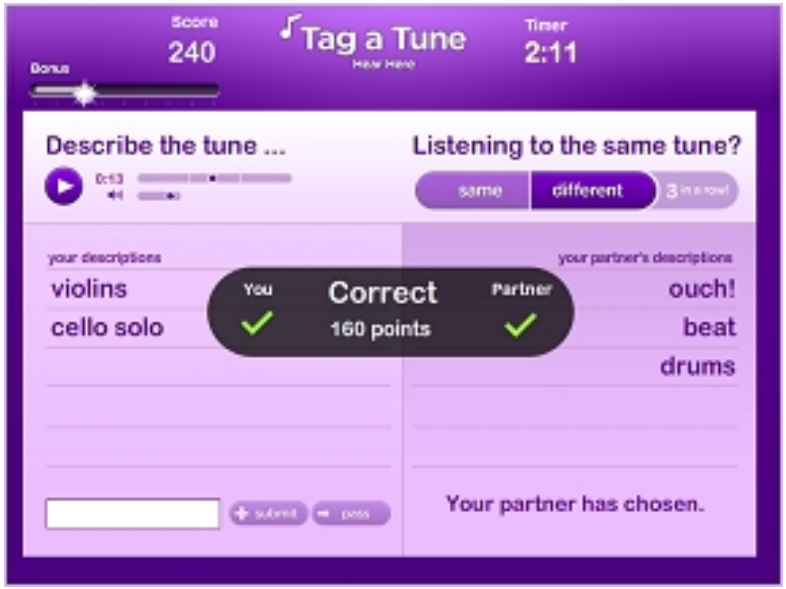
WHAT TASKS CAN BE SUBJECT TO A GAME?*

- Decomposable into simpler tasks
- Nested tasks
- Performance is measurable
- Obvious rewarding scheme
- Skills can be arranged in a smooth learning curve

*http://www.lostgarden.com/2008/06/what-actitivies-that-can-be-turned-into.html

EXAMPLE: GAME ELEMENTS IN TAG-A-TUNE

- Task is decomposable
  - Audio annotation one tune per game round
- No nested tasks
- Performance: consensus, bonus round on rating audios

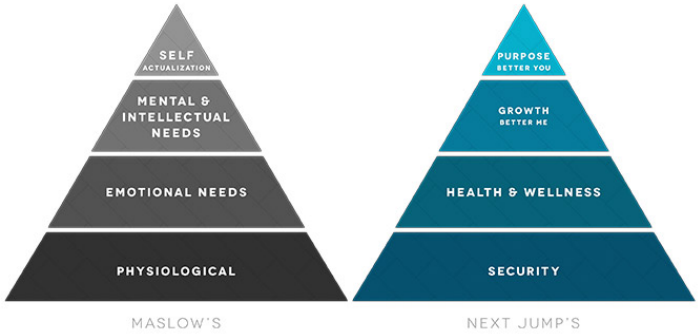


Players decide whether they listen to the same tune by exchanging free-text hints

EXAMPLE: GAMIFIED HEALTHCARE

- **eCommerce solution provider, loyality management, gamification**
- **Goal: motivate employees to exercise ; reduces health insurance costs and increases productivity**
- **How**
  - Installed gyms in offices and built custom application to check in to each workout; competition with cash prize → 12% employees
  - Fitness as a team sport: form regional teams competing against each other; leaderboards → 70% employees

nextjump



DIMENSIONS OF GWAP DESIGN

DIMENSIONS OF GWAP DESIGN

WHAT IS THE PURPOSE OF THE GAME

- **Concrete specification of the task**
  - Example: annotation of a set of 500,000 images using free labels, controlled vocabulary etc
- **Where does the input data come from? How much noise can you expect in the data?**
  - Example: validating the results of algorithms; poor input data hampers UX

HOW CAN IT BE TRANSLATED INTO DECOMPOSABLE TASKS

- **Repetitive tasks vs. player experience; see motivation**

DIMENSIONS OF GWAP DESIGN (2)

WHAT SUB-TASKS CAN YOU IDENTIFY

- **Number of interrelated steps in a casual game and granularity of tasks**

HOW DOES THE HUMAN READABLE DESCRIPTION OF THE TASK LOOK LIKE

- **See Linked Data examples**

DIMENSIONS OF GWAP DESIGN (3)

HOW TO YOU MEASURE PERFORMANCE

- **Redundancy (output-agreement games)**
- **Consensus (input agreement, cf Tag-A-Tune)**
- **Describer - guesser**

WHAT DO USERS RECEIVE POINTS FOR, WHEN, AND HOW MANY

- **Mechanism design**

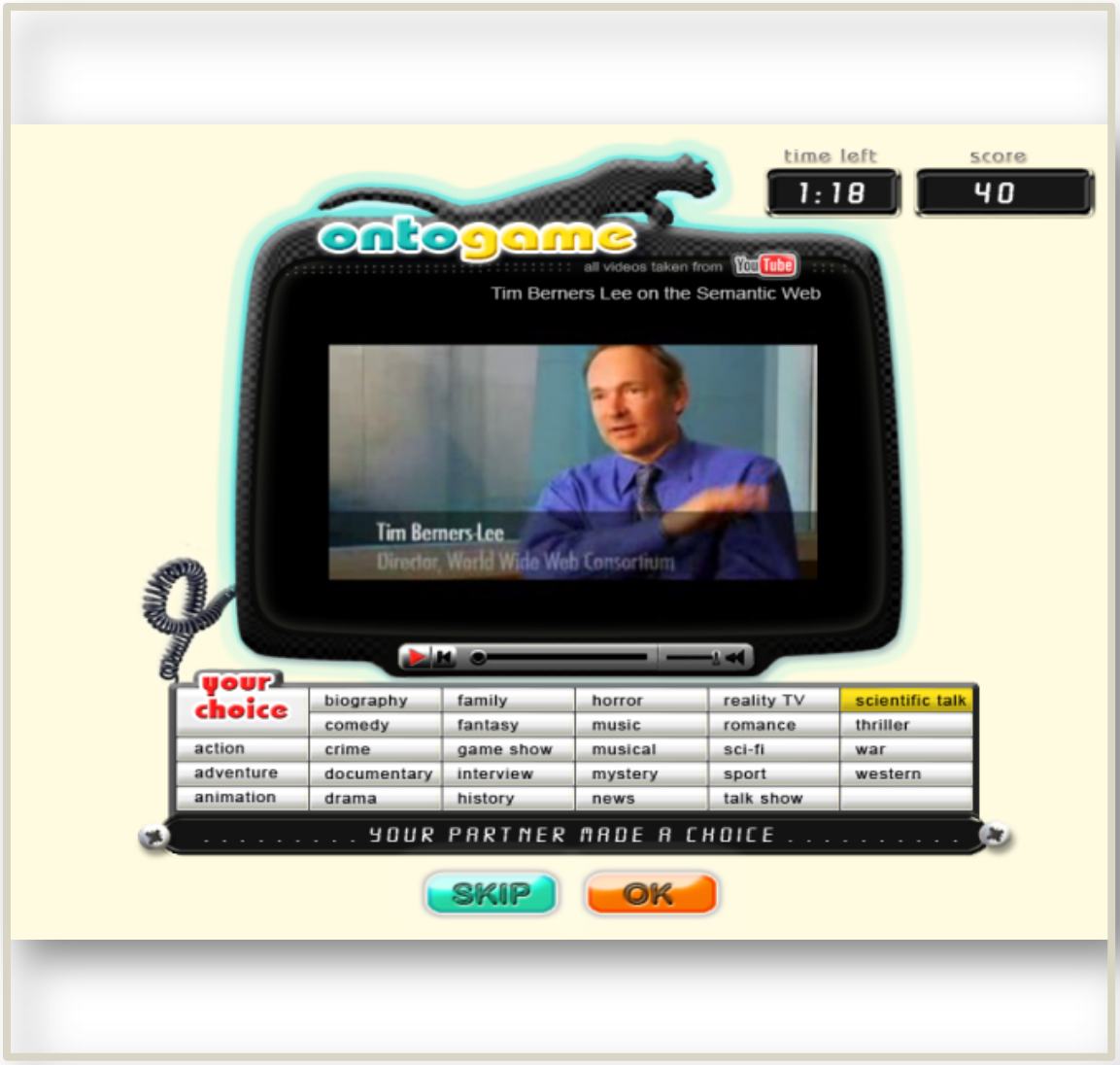
Note: tasks cannot be too difficult, otherwise the tasks feel like work; they have to be interesting and intellectually challenging, otherwise the game is boring; players should be able to get better at it during the game.

SINGLE VS. MULTI-PLAYER GAMES

- **Multi-player games**
  - UX (player appreciate social contact and intellectual challenge)
  - Consensus mechanism, less spam
  - Rapid feedback cycles
  - But: requires players' matching functionality and enough players in the system at the same time
    - Can be simulated using bots and (lots of) pre-recorded rounds
- **Single-player games**
  - Different quality assurance method (player receives reward once correct answer is determined); or
  - Training data available to build initial profile

VERBOSITY AS INVERSION PROBLEM GAME

TASKS SHOULD BE SOLVABLE



MECHANISM DESIGN

- **Area of game theory**
  - Game designer defining the structure of the game
  - Game designer is interested in specific outcomes and attempts to influence players' behavior to achieve these outcomes

- **Different reward models can be applied**
  - Pay-per-item vs winner-takes-it-all
  - Competitions among individuals and teams
  - How to price contributions

  - These parameters will change the behavior of the users in the system

DIMENSIONS OF GWAP DESIGN (4)

HOW DO YOU TRANSLATE CROWD INPUTS INTO VALIDATED ANSWERS

- When are two answers the same
- How many assignments per question
- Player's reliability, spam

HOW DO YOU ASSIGN CHALLENGES TO PLAYERS

- Random vs based on previous performance
- The same about players matching

DIMENSIONS OF GWAP DESIGN (5)

WHAT ADDITIONAL GAME ELEMENTS CAN YOU INCLUDE

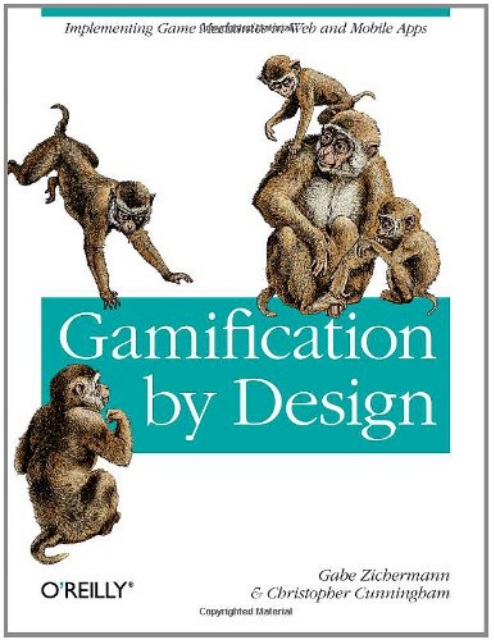
- Different types of players*
- Useful information (tabu lists)
- Levels of difficulty (requires knowledge of the problem space)
- Timing adds a sense of urgency, feels more like play and less like work
- Leader boards, badges, appointment dynamics

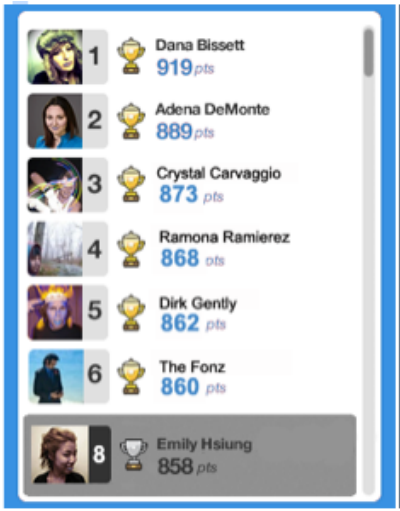
***http://www.gamasutra.com/blogs/VictorManrique/20130524/193007/Gamification_Player_Types_The_TimeEngagement_Pyramid.php**

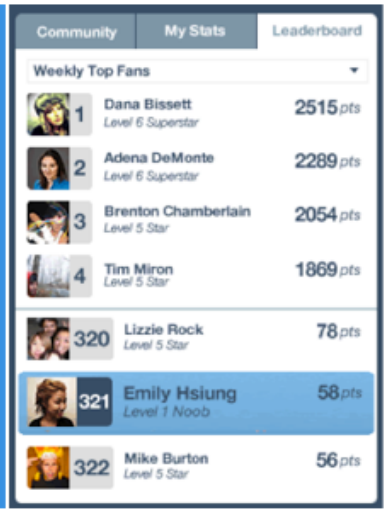
GAME MECHANICS: LEADERBOARDS*

- Global: top-players
- Local: user can find himself, has attainable objective
- Friends: user competes and interacts with players he feels emotionally attached to
- Filtered: e.g., all players in one area, creates a sense of community
- Aggregates: e.g., for teams, encourage cooperation, create social pressure

***http://blog.badgeville.com/2013/03/15/game-mechanics-leaderboards-pt1/**



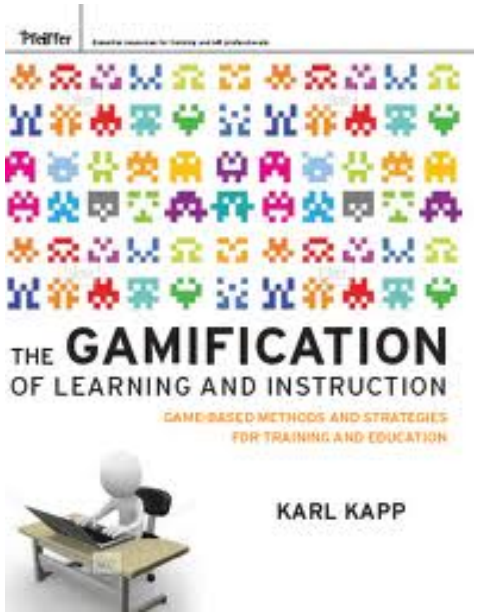


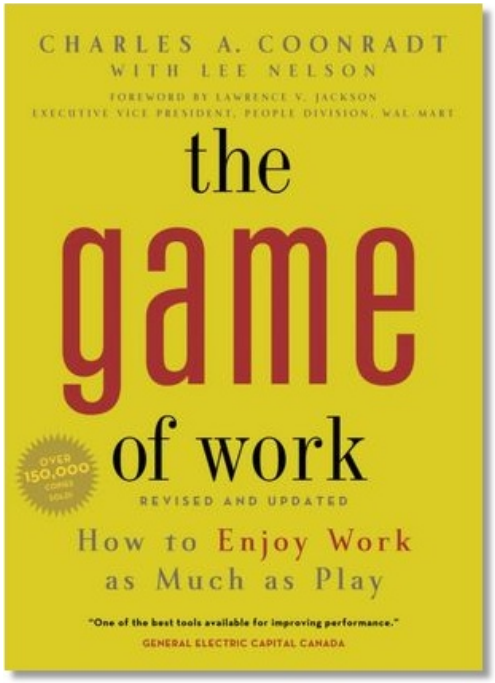


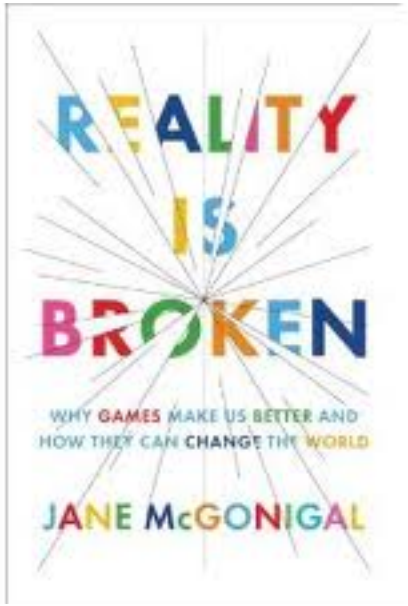
CONCLUSIONS

GWAPs are a useful concept for intellectually challenging, but solvable, fine-granular tasks, in domains appealing for mainstream casual Internet users

Game mechanics need careful tuning to be effective







THANK YOU

e.simperl@soton.ac.uk @esimperl

Micro-task Management and Automation

Gianluca Demartini

Types of Crowdsourcing Tasks

Task Granularity	Examples
Complex Tasks	• Build a website • Develop a software system • Overthrow a government?
Simple Projects	• Design a logo and visual identity • Write a term paper
Macro Tasks	• Write a restaurant review • Test a new website feature • Identify a galaxy
Micro Tasks	• Label an image • Verify an address • Simple entity resolution

Inspired by the report: “Paid Crowdsourcing”, Smartsheet.com, 9/15/2009

Outline

- Micro-task Crowdsourcing Challenges
  - Design the User Interfaces
  - Define the right Incentives
  - Task Patterns
  - Scalability
  - Quality (more in the next session)

Case-Study: Amazon MTurk

- Micro-task crowdsourcing marketplace
- On-demand, scalable, real-time workforce
- Online since 2005 (still in “beta”)
- Currently the most popular platform
- Developer’s API as well as GUI

Amazon MTurk

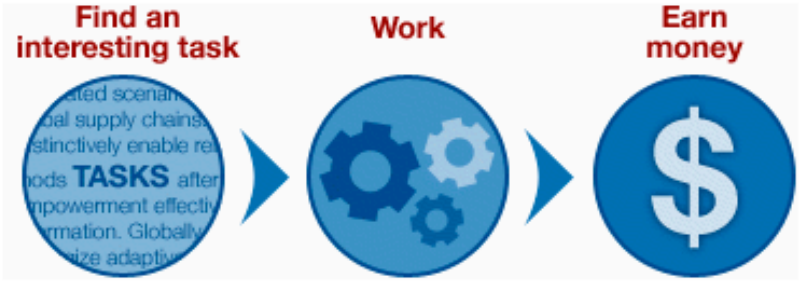
amazon mechanical turk
beta
Artificial Artificial Intelligence

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 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

Get Started

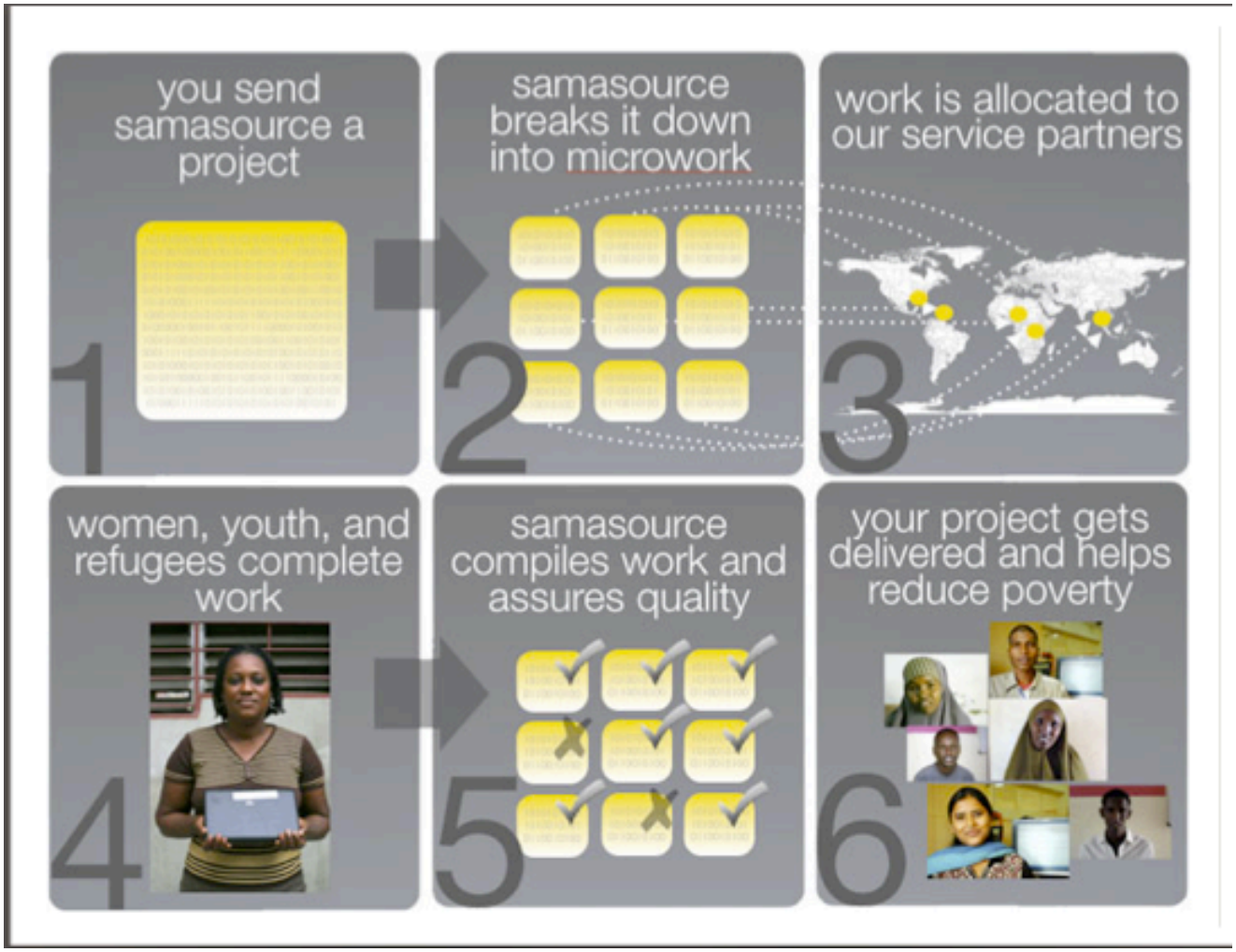
Amazon Mturk Workflow

- Requesters create tasks (HITs)
- Workers preview, accept, submit HITs
- Requesters approve, download results

Microtask Aggregators

http://www.businesswire.com/news/home/20120207005761/en/CrowdFlower-Reports-Revenue-300-Year-Year-300

Samasource.org



mobileworks

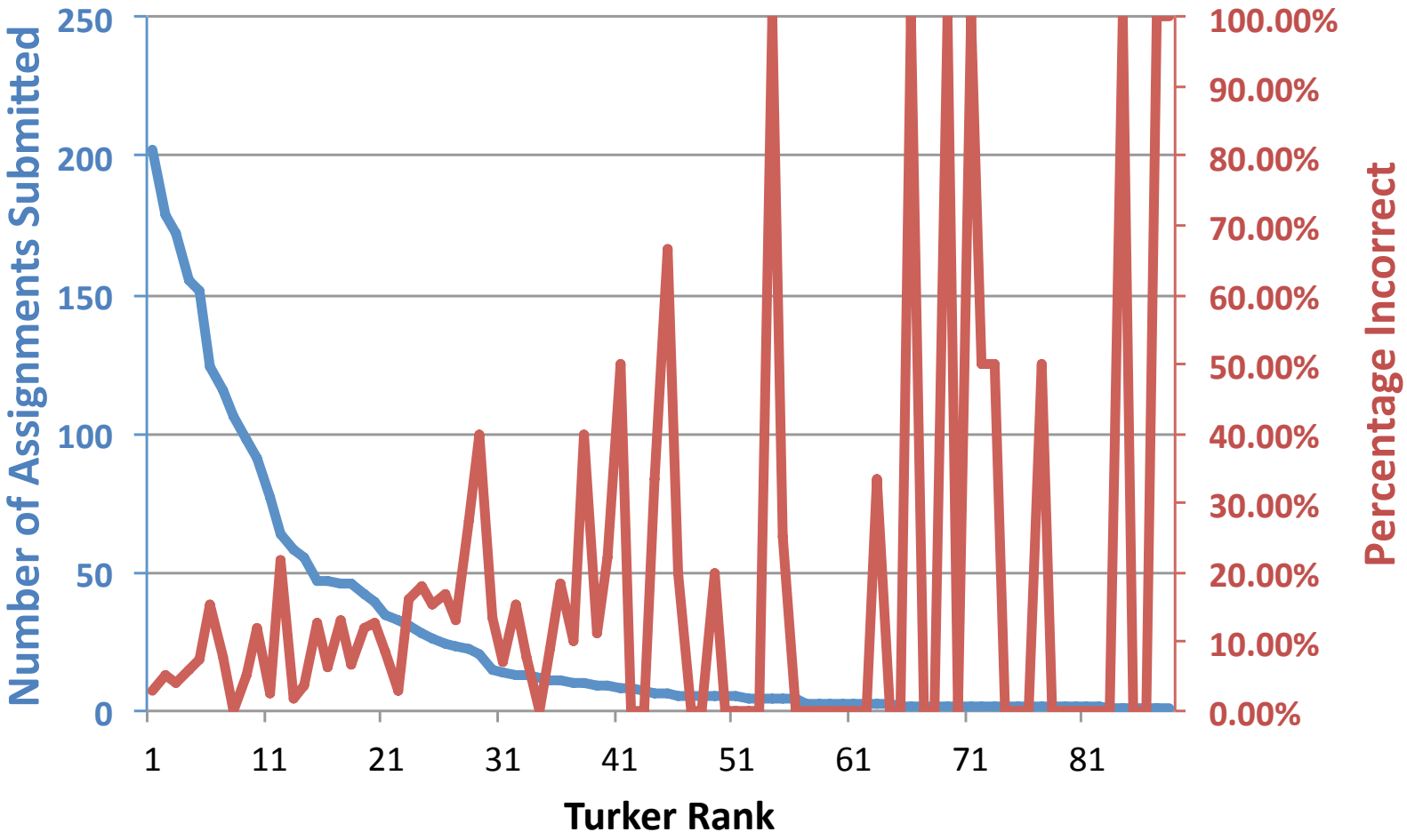
- “Accurate, qualified workers and reliable security”
- Workers from Latin America
- Reward decided by the platform based on task complexity

Design of a Task on MTurk

High-level Issues in Crowdsourcing

- Process
  - Experimental design, annotation guidelines, iteration
- Choose crowdsourcing platform (or roll your own!)
- Human factors
  - Payment / incentives, interface and interaction design, communication, reputation, recruitment, retention
- Quality Control / Data Quality
  - Trust, reliability, spam detection, consensus labeling

Turker Affinity and Errors



[Franklin, Kossmann, Kraska, Ramesh, Xin: CrowdDB: Answering Queries with Crowdsourcing. *SIGMOD*,2011]

Typical Workflow

- Define and design what to test
- Sample data
- **Design the experiment**
- Run experiment (see later session by Maribel)
- Collect data and analyze results
- **Quality control**

Task Design

- One of the most important parts
- Part art, part science
- Instructions are key
- Prepare to iterate

Task Design

- Ask the right questions
- Workers may not be experts: don't assume the same understanding in terms of terminology
- Show examples
- Hire a technical writer
  - Engineer writes the specification
  - Writer communicates

Task Design - UI

- Generic tips
  - Experiment should be self-contained.
  - Keep it short and simple. Brief and concise.
  - Be very clear with the relevance task.
  - Engage with the worker. Avoid boring stuff.
  - Always ask for feedback (open-ended question) in an input box.

Task Design - UI

- Presentation
- Document design
- Highlight important concepts
- Colors and fonts
- Need to grab attention
- Localization

Other design principles

- Text alignment
- Legibility
- Reading level: complexity of words and sentences
- Attractiveness (worker's attention & enjoyment)
- Multi-cultural / multi-lingual
- Who is the audience (e.g. target worker community)
  - Special needs communities (e.g. simple color blindness)
- Cognitive load: mental rigor needed to perform task

Bad Example

- Asking too much, task not clear, “do NOT/reject”
- Worker has to do a lot of stuff

Help us describe How-To Videos! Earn $2.50 bonus for every 25 videos entered!

Watch a how-to video, and write a keyword-friendly synopsis describing the video.

1. Click on the link to watch the **Film & Theater** how-to video ==> 332492 Get a 35mm film look with a depth of field adaptor
2. Write a description of the video linked in 4 or more sentences.
3. Be detailed in your description. Describe how the procedure is done.
4. Description should be at least 100 words.
5. Description should be fewer than 2000 characters.
6. Use the character and word counters below to help you stay within the limits.
7. You must complete **25 video descriptions** in order to earn the $2.50 bonus. Bonuses are distributed after HITs have been completed. The more HITs completed and approved, the more you will earn.
8. It is **not necessary** to repeat the headline in your entry. It will **NOT** count toward your word count.
9. Do NOT describe the following: the format, where the video comes from, or how long the video is. This information is **IRRELEVANT**.
10. Do NOT describe the video in the following manner: "She turns around to face the camera. Then she faces left." Follow the examples below.

Current Word Count: 0 Current Character Count: 0 / 2000

Criteria for **REJECTION**:

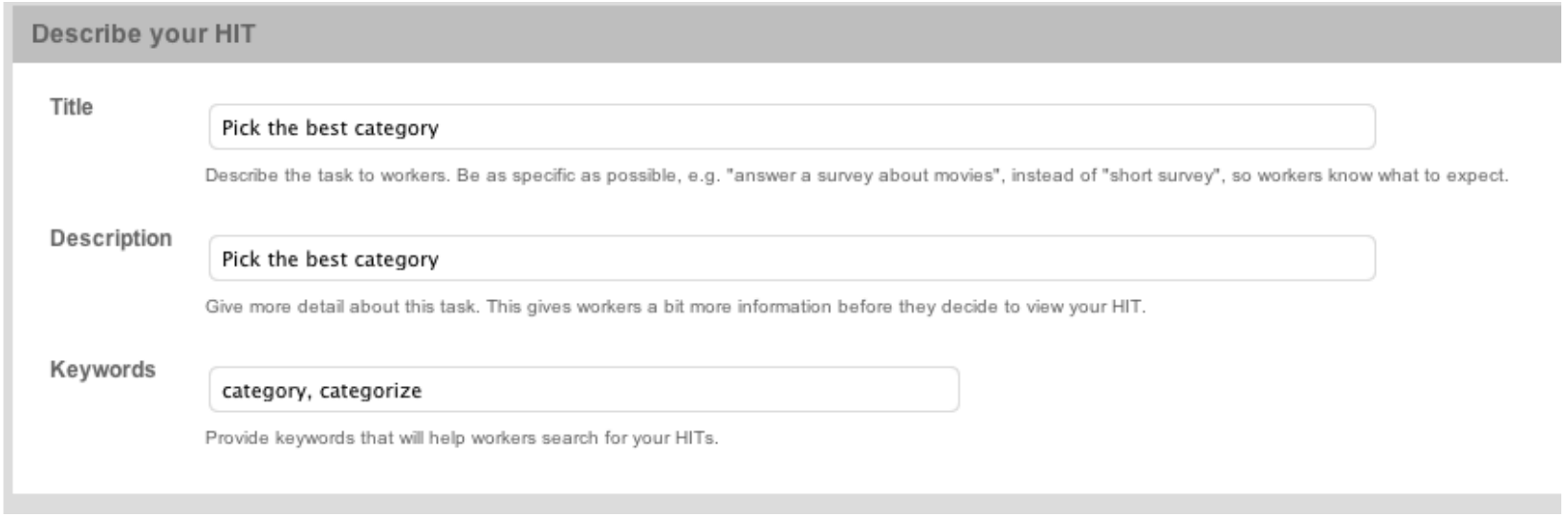
1. Entries with obvious and multiple spelling or grammatical errors will be **rejected**.
2. Entries with fewer than 100 words will be automatically **rejected**.
3. Text copied from the web or other places will be **rejected**. Multiple plagiarized answers will lead to being **BLOCKED**. You may use a quotation, but the majority of your content must be **ORIGINAL**.
4. Incomplete and blank answers will be rejected. Multiple blank answers will result in being **blocked**.
5. Tasks submitted without descriptions will be **rejected**.
6. Tasks submitted with inaccurate descriptions will be **rejected** as well.
7. Do **NOT** add any personal opinions. Entries with personal opinions or reviews will be automatically **REJECTED**.
8. If you notify us that a link is broken, we appreciate it but will not be able to accept the submission. The notification will result in **rejection**.
9. Entries that transcribe the video will be **REJECTED**.

Good Example

- All information is available
  - What to do
  - Search result
  - Question to answer

Form and Metadata

- Form with a close question (binary relevance) and open-ended question (user feedback)
- Clear title, useful keywords
- Workers need to find your task



How Much to Pay?

- Price commensurate with task effort
  - Ex: $0.02 for yes/no answer + $0.02 bonus for optional feedback
- Ethics & market-factors
  - e.g. non-profit SamaSource contracts workers refugee camps
- Uptake & time-to-completion vs. Cost & Quality
  - Too little $$, no interest or slow
  - too much $$, attract spammers
- Accuracy & quantity
  - More pay = more work, not better (W. Mason and D. Watts, 2009)

Development Framework

- Similar to a UX
- Build a mock up and test it with your team
  - Yes, you need to judge some tasks
- Incorporate feedback and run a test on MTurk with a very small data set
  - Time the experiment
  - Do people understand the task?
- Analyze results
  - Look for spammers
  - Check completion times
- Iterate and modify accordingly

Development Framework

- Introduce quality control
  - Qualification test
  - Gold answers (honey pots)
- Adjust passing grade and worker approval rate
- Run experiment with new settings & same data
- Scale on data
- Scale on workers

Summary

- Micro-task Crowdsourcing Challenges
  - Design the User Interfaces
  - Define the right Incentives
  - **Task Patterns**
  - **Scalability**
  - Quality (more in the next session)

Crowdsourcing Patterns

- Majority Vote Aggregation
  - Select the answer among a set of candidates
  - Pick the most popular answer
- Find-Fix-Verify
  - Creative process
  - Three-steps iterative crowdsourcing
- Interaction Protocol (for hybrid human-machine systems)
  - Upfront
  - Iterative

Interaction Protocol

How often can we refer to the crowd?

1. **Upfront**: Ask all the B queries at once
2. **Iterative**: Ask K queries to the crowd and use them to improve the system. Repeat this B/K times

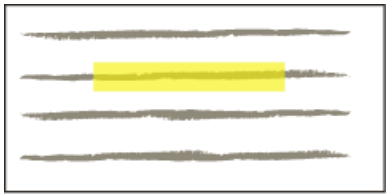
Measures Used for Selection

- **Uncertainty:** Asking hardest (most ambiguous) questions
- **Explorer:** Ask questions with potential to have largest impact on the system

Soylent: Find-Fix-Verify

Find

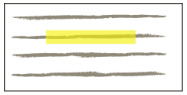
"Identify at least one area that can be shortened without changing the meaning of the paragraph."



Independent agreement to identify patches

Fix

"Edit the highlighted section to shorten its length without changing the meaning of the paragraph."



Soylent, a prototype...

Randomize order of suggestions

Verify

"Choose at least one rewrite that has style errors, and at least one rewrite that changes the meaning of the sentence."

- ☐ Soylent ~~is~~, a prototype...
- ☐ Soylent ~~is a~~ prototypes...
- ☑ Soylent is a ~~prototype~~test...

[Bernstein et al: Soylent: A Word Processor with a Crowd Inside. *UIST,* 2010]

Find-Fix-Verify

- Machine Translation example

- Find
  - Show automatically translated text
  - Ask if they are grammatically correct
- Fix
  - Ask to translate those which contain errors (multiple times)
- Verify
  - Select the best translation among the available ones

References

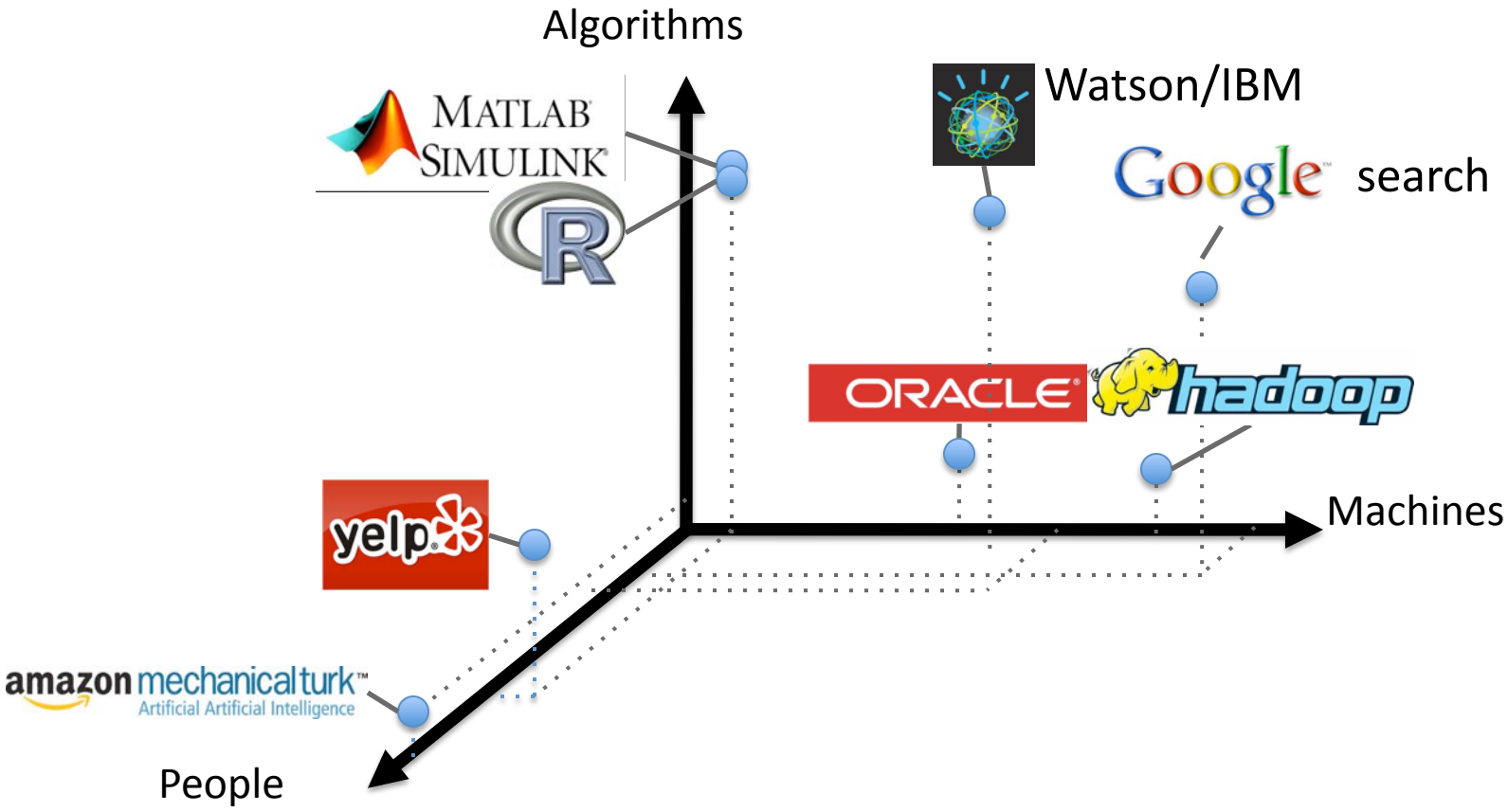
- "**Crowdsourcing for Information Retrieval: Principles, Methods, and Applications**" SIGIR 2011 Tutorial.

Micro-task Automation:
Hybrid Human Machine Systems

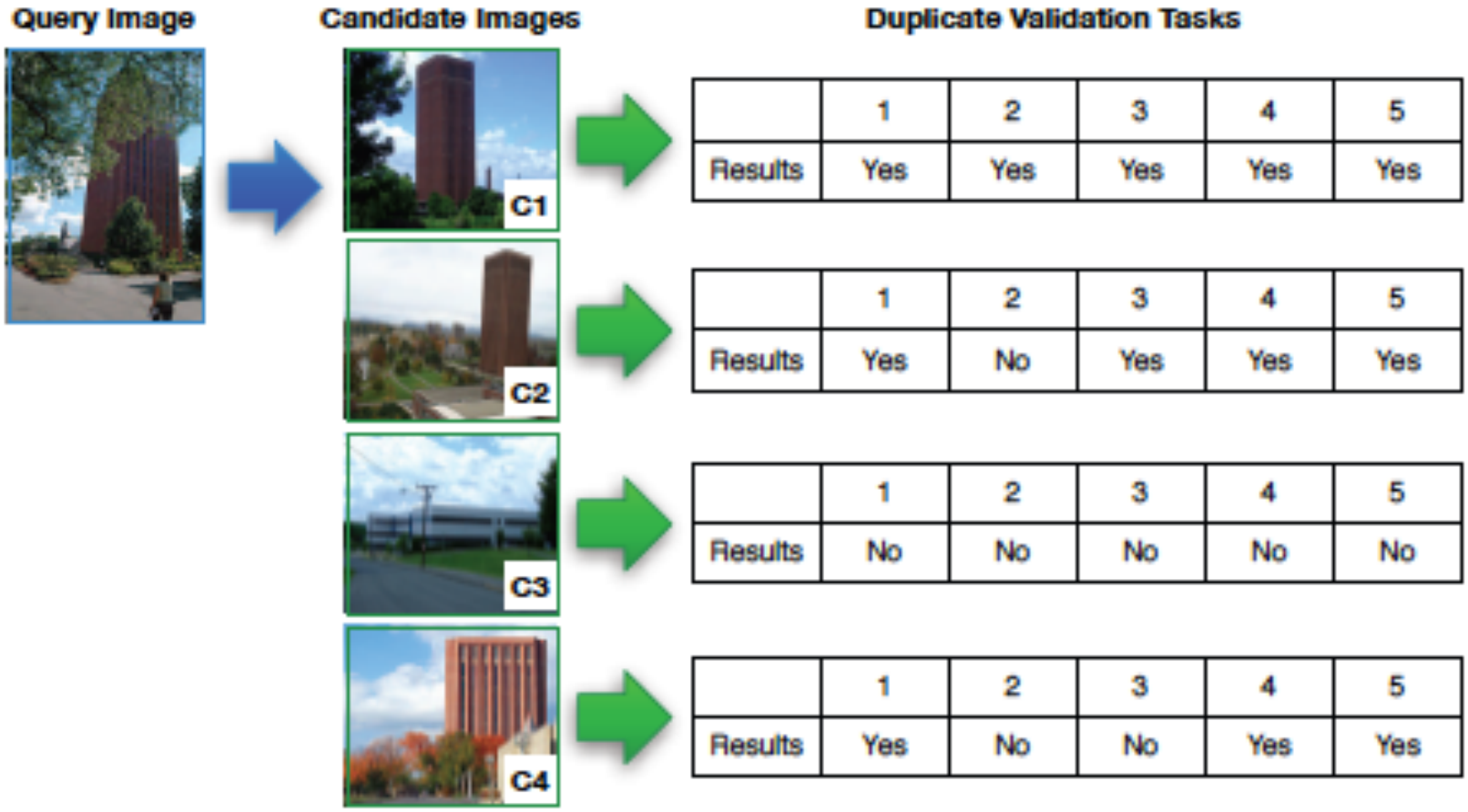
Hybrid Systems: Key Issues

- The role of machine (i.e., algorithm) and humans
  - use only humans? both? who's doing what?
- Quality control
- **Optimization: What to crowdsource**
- **Scalability: How much to crowdsource**

Thinking About Hybrid Systems



Example: Hybrid Image Search



	1	2	3	4	5
Results	Yes	Yes	Yes	Yes	Yes

	1	2	3	4	5
Results	Yes	No	Yes	Yes	Yes

	1	2	3	4	5
Results	No	No	No	No	No

	1	2	3	4	5
Results	Yes	No	No	Yes	Yes

Yan, Kumar, Ganesan, CrowdSearch: Exploiting Crowds for Accurate Real-time Image Search on Mobile Phones, Mobisys 2010.

Example: Hybrid Data Integration

paper	conf
Data integration	VLDB-01
Data mining	SIGMOD-02

title	author	email	venue
OLAP	Mike	mike@a	ICDE-02
Social media	Jane	jane@b	PODS-05

- **Generate plausible matches**
  - paper = title, paper = author, paper = email, paper = venue
  - conf = title, conf = author, conf = email, conf = venue
- **Ask users to verify**

Does attribute paper match attribute author?

paper	conf
Data integration	VLDB-01
Data mining	SIGMOD-02

title	author	email
OLAP	Mike	mike@a
Social media	Jane	jane@b

Yes No Not sure

McCann, Shen, Doan: Matching Schemas in Online Communities. ICDE, 2008

CrowdQ: Crowdsourced Query Understanding

- CrowdQ is the first system that uses crowdsourcing to
  - *Understand* the meaning of a keyword query
  - *Build* a structured (SPARQL) query template
  - *Answer* the query over Linked Open Data

Gianluca Demartini, Beth Trushkowsky, Tim Kraska, and Michael Franklin. **CrowdQ: Crowdsourced Query Understanding.** In: 6th Biennial Conference on Innovative Data Systems Research (CIDR 2013)

birthdate of the mayors of all the cities in Italy

About 124,000,000 results (0.33 seconds)

City	Mayor	Birthdate
Rome, Italy	Gianni Alemanno	March 3, 1958
Venice, Italy	Giorgio Orsoni	August 29, 1946
Milan, Italy	Giuliano Pisapia	May 20, 1949

Press to see more

Cities in Italy | Italy Travel Guide
www.**italy**logue.com/**italian-cities**
Learn about the best **cities in Italy** to visit, and some **Italian cities** you might never have heard of before. These **cities in Italy** are **all** great for visitors.

Top Ten **Cities** for Visitors to **Italy** - Top **Italian Cities** to See
go**italy**.about.com/od/planningandinformation/tp/top**cities**.htm
Italy has many beautiful and historic **cities** that are well worth a visit. Here are our picks for the ten best **cities** for visitors to **Italy**.

Italian Cities and Towns - **Italy**
en.comuni-**italian**i.it/
Information and statistics on **Italian** Regions, Provinces and Municipalities. **All Cities**, Towns and Villages in **Italy**. Official site, zip codes, phone prefix, population, ...

WHO/Europe | **Italy** - Making **cities** healthy for everyone: European ...
www.euro.who.int/.../**italy**/.../making-**cities**-healthy-for-everyo...
Jun 21, 2012 – Over 300 **mayors** and municipal health leaders met in St Petersburg, Russian Federation on ... Annual Business and Technical Conference of the WHO European Healthy **Cities** Network ... "No **all** solit" (Say no to loneliness) ...

Hybrid Human-Machine Pipeline

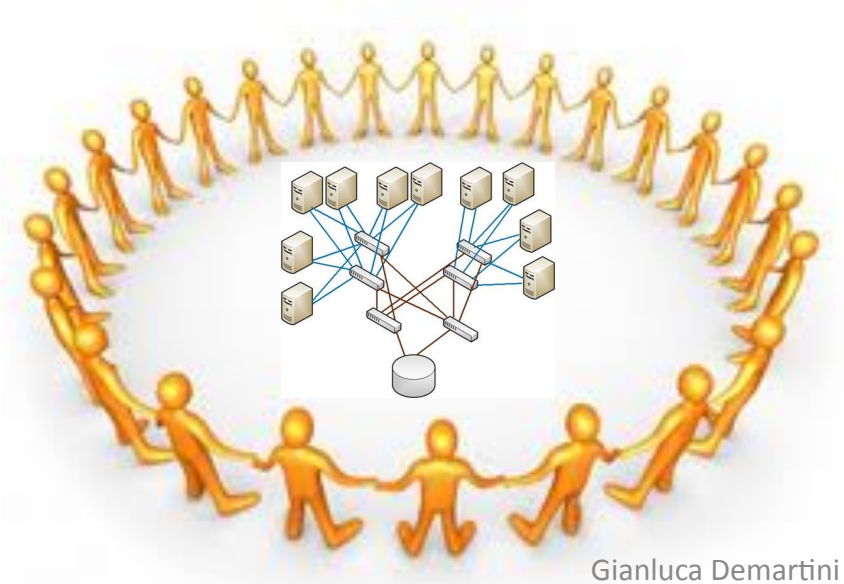
Q= birthdate of actors of forrest gump

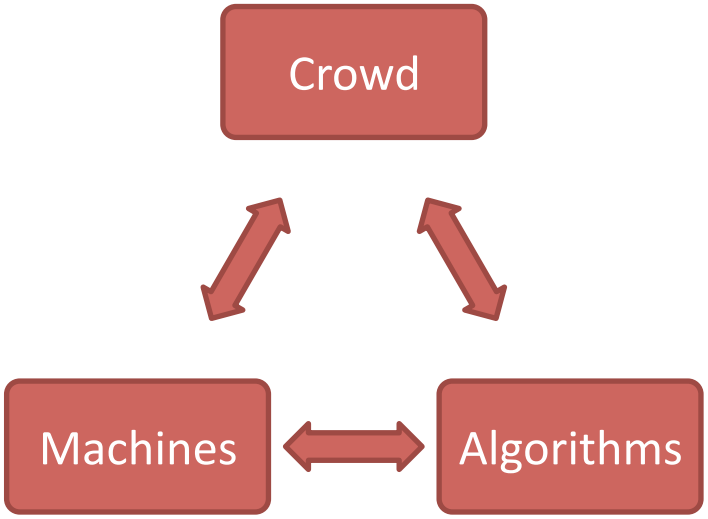
Query annotation	Noun	Noun	Named entity

Verification	Is forrest gump this entity in the query?
Entity Relations	Which is the relation between: actors and forrest gump → starring
Schema element	Starring → <dbpedia-owl:starring>
Verification	Is the relation between: **Indiana Jones – Harrison Ford** **Back to the Future – Michael J. Fox** of the same type as **Forrest Gump - actors**

ZenCrowd: Entity Linking by the Crowd

- Combine both algorithmic and manual linking
- Automate manual linking via crowdsourcing
- Dynamically assess human workers with a probabilistic reasoning framework





APRIL 9, 2012, 1:15 PM MERGERS & ACQUISITIONS

Facebook Buys Instagram for $1 Billion

BY EVELYN M. RUSLI

2:02 p.m. | Updated

Facebook is not waiting for its initial public offering to make its first big purchase.

In its largest acquisition to date, the social network has purchased Instagram, the popular photo-sharing application, for about $1 billion in cash and stock, the company said Monday.

http://dbpedia.org/resource/Facebook

http://dbpedia.org/resource/Instagram

owl:sameAs

fbase:Instagram

HTML:

```
<p>Facebook is not waiting for its initial public offering to make its first big purchase.</p><p>In its largest acquisition to date, the social network has purchased Instagram, the popular photo-sharing application, for about $1 billion in cash and stock, the company said Monday.</p>
```

RDFa enrichment

```
<p><span about="http://dbpedia.org/resource/Facebook"><cite property="rdfs:label">Facebook</cite> is not waiting for its initial public offering to make its first big purchase.</span></p><p><span about="http://dbpedia.org/resource/Instagram">In its largest acquisition to date, the social network has purchased <cite property="rdfs:label">Instagram</cite> , the popular photo-sharing application, for about $1 billion in cash and stock, the company said Monday.</span></p>
```

CNET › News › Mobile

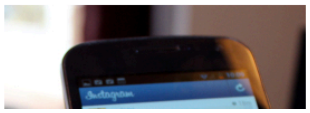
Instagram for Android is now available

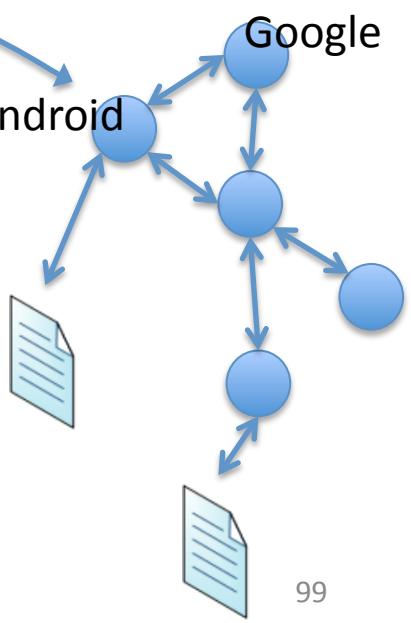
At long last, Instagram finally releases the Android version of its app.

by Jason Cipriani | April 3, 2012 10:07 AM PDT

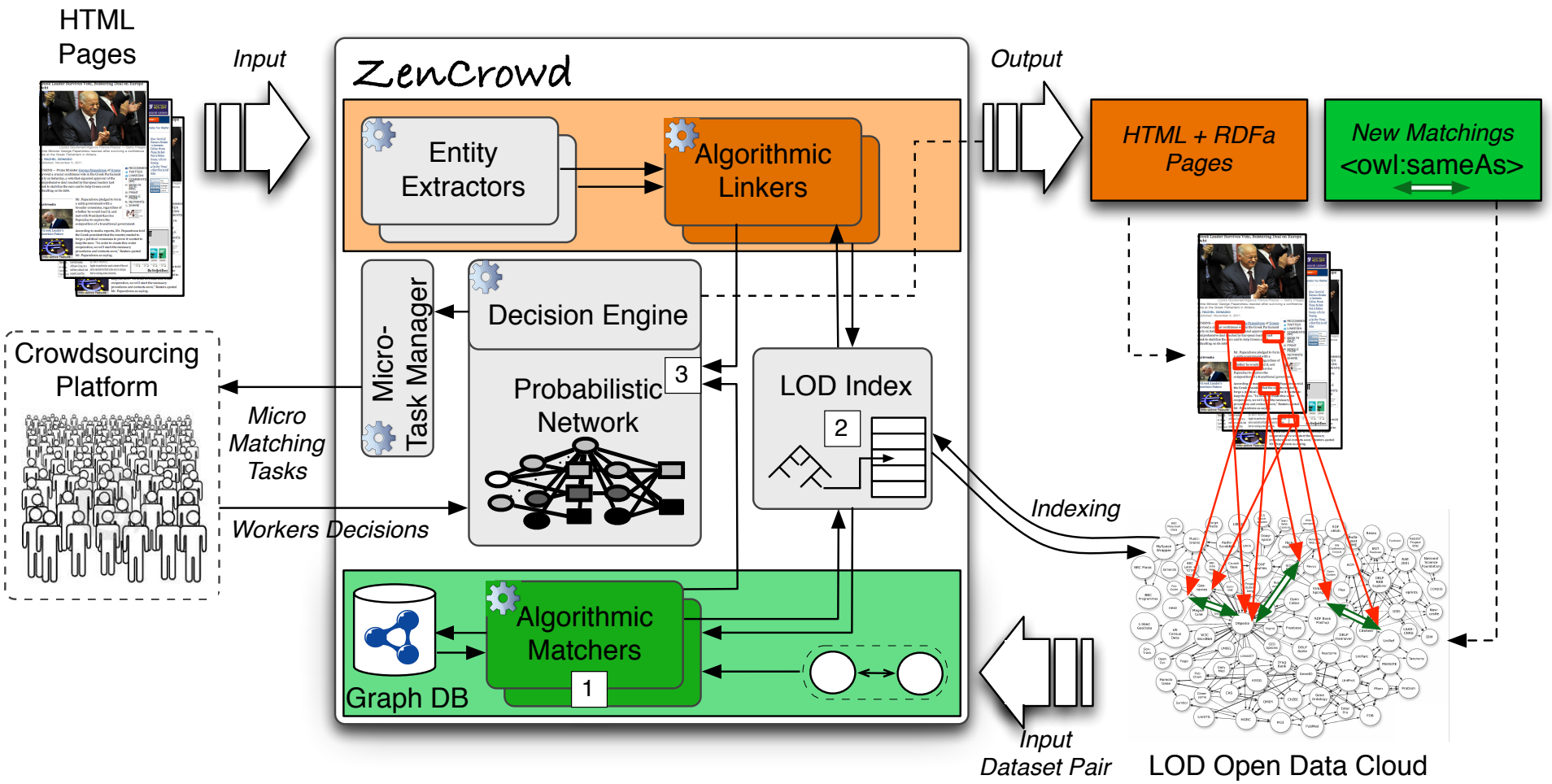
Follow

Instagram has been around since 2010, available only to iOS devices. Android users have been waiting patiently, with repeated promises of an Android version arriving soon.





ZenCrowd Architecture



Gianluca Demartini, Djellel Eddine Difallah, and Philippe Cudré-Mauroux. **ZenCrowd: Leveraging Probabilistic Reasoning and Crowdsourcing Techniques for Large-Scale Entity Linking.** In: 21st International Conference on World Wide Web (WWW 2012)

Crowdsourcing Scalability

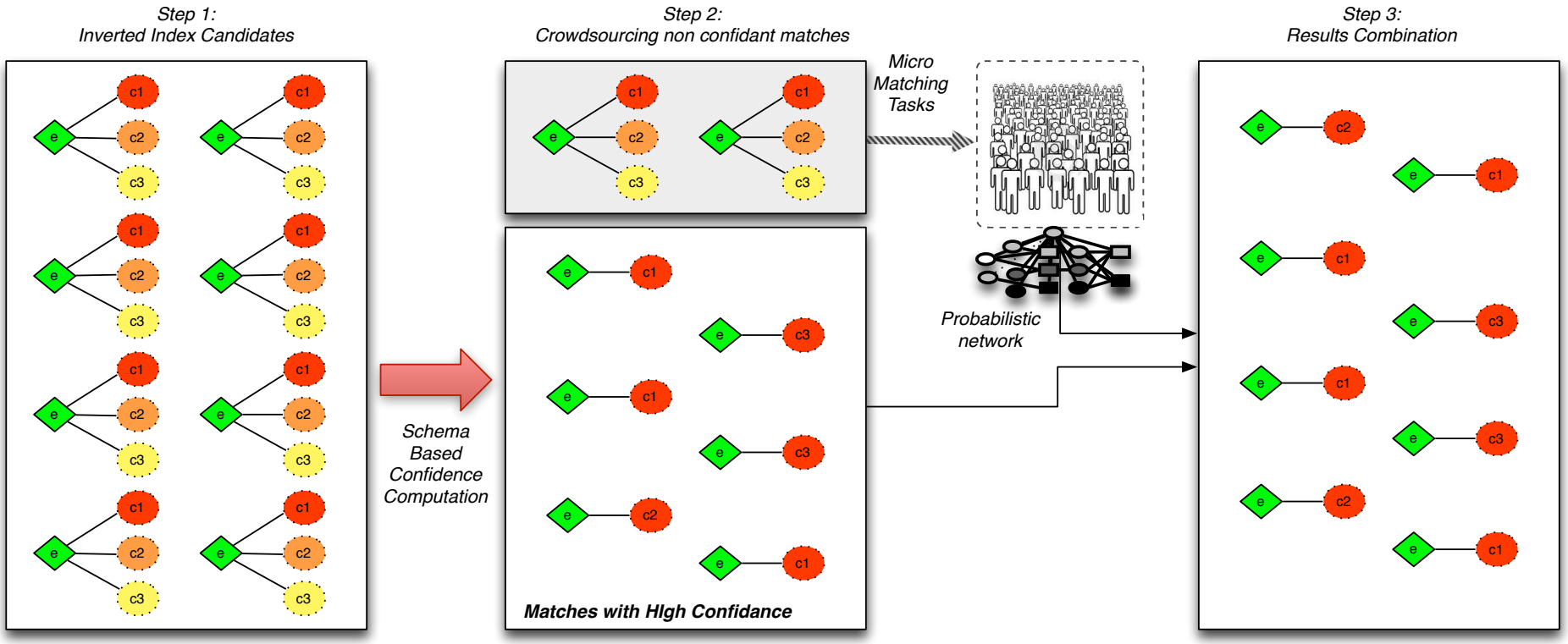
- Crowd-sourcing is becoming an indispensible method of collecting labeled data, e.g., Machine Learning
- BUT crowd-sourcing can be expensive, slow, and noisy
- All Human Intelligent Tasks (HIT) *are NOT equally difficult for the machine*
- To achieve scalability, we need to know when and how to use machines along with humans

Blocking for Instance Matching

- Find the instances about the same real-world entity within two datasets
- Avoid Comparison of all possible pairs
  - Step 1: cluster similar items using a cheap similarity measure
  - Step 2: n*n comparison within the clusters with an expensive measure

3-steps Blocking with the Crowd

- Crowdsourcing as the most expensive similarity measure



Conclusions

- Carefully design the User Interface
- Define the right Incentives
- Use Task Patterns
- Enable Scalability
- Quality (more in the next session)

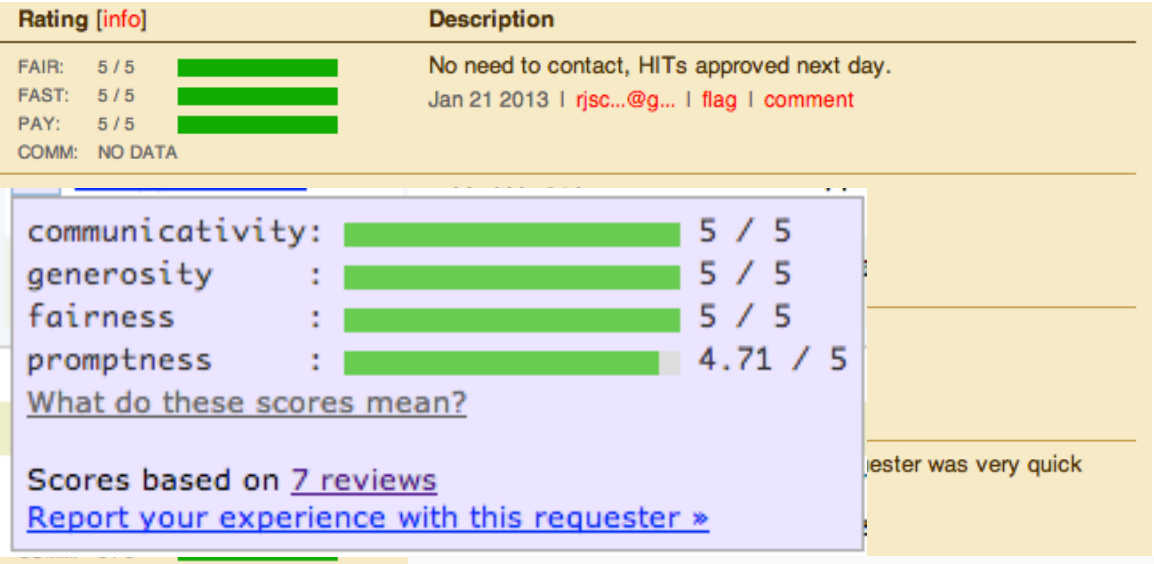
Micro-task Crowdsourcing Quality Control

Gianluca Demartini

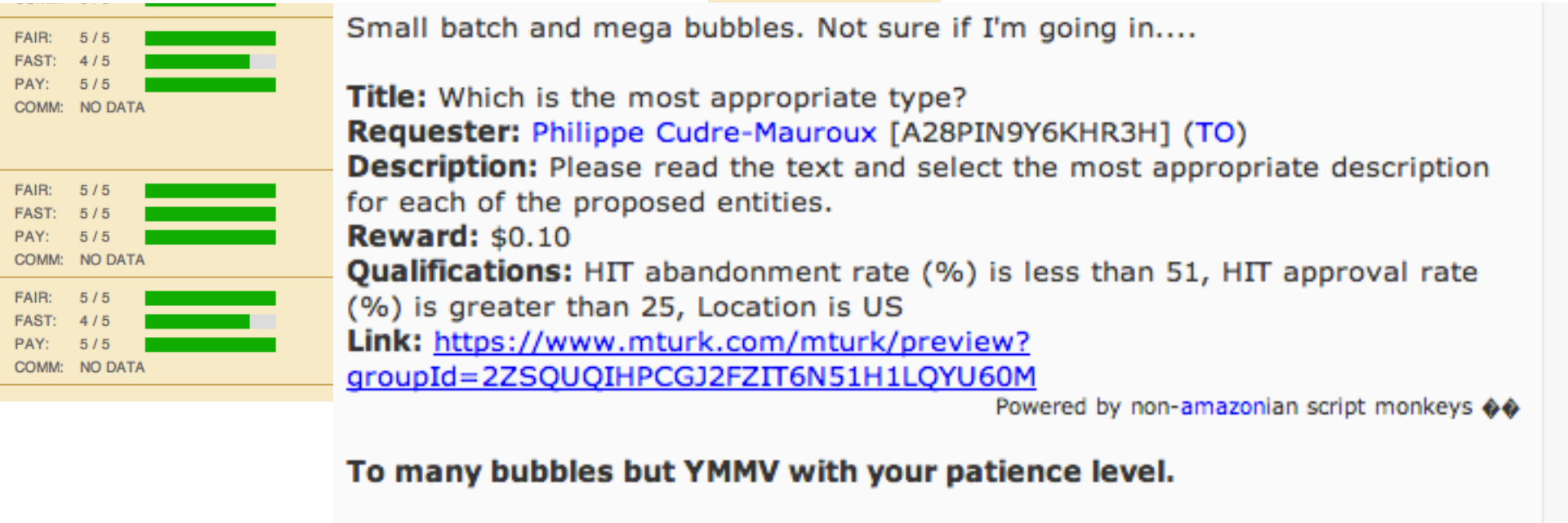
Quality Control

- Extremely important part of the experiment
- Approach as “overall” quality; not just for workers
- Bi-directional channel
  - You may think the worker is doing a bad job.
  - The same worker may think you are a lousy requester.
  - Do check the worker forums!

Crowd Worker Communities



Turkopticon.com
Mturkforum.com
Turkernation.com



Quality Control

- Approval rate: easy to use, & just as easily defeated
- Mechanical Turk Masters (since June 2011)
  - Recent addition, only for specific tasks
- Qualification test
  - Pre-screen workers' ability to do the task (accurately)
- Assess worker quality as you go
  - Trap questions with known answers ("honey pots")
  - Measure inner-annotator agreement between workers

Qualification tests: pros and cons

- Advantages
  - Great tool for controlling quality
  - Adjust passing grade
- Disadvantages
  - Extra cost to design and implement the test
  - May turn off workers, hurt completion time
  - Refresh the test on a regular basis
  - Hard to verify subjective tasks like judging relevance
- Try creating task-related questions to get worker familiar with task *before* starting task in earnest

Methods for measuring agreement

- What to look for
  - Agreement, reliability, validity
- Inter-agreement level
  - Agreement between judges
  - Agreement between judges and the gold set
- Some statistics
  - Percentage agreement
  - Cohen's kappa (2 raters)
  - Fleiss' kappa (any number of raters)
- With majority vote, what if 2 say relevant, 3 say not?
  - Use expert to break ties
  - Collect more judgments as needed to reduce uncertainty

Quality Control & Assurance

- Filtering
  - Approval rate (built-in but defeatable)
  - Geographic restrictions (e.g. US only, built-in)
  - Worker blocking
  - Qualification test
    - Con: slows down experiment, difficult to "test" relevance
    - Solution: create questions to let user get familiar *before* the assessment
  - Does not guarantee success
- Identify workers that *always* disagree with the majority
- Ask workers to rate the difficulty of a task

Other quality heuristics

- Justification/feedback as quasi-captcha
  - Should be optional
  - Automatically verifying feedback was written by a person may be difficult (classic spam detection task)
- Broken URL/incorrect object
  - Leave an outlier in the data set
  - Workers will tell you
  - If somebody answers “excellent” for a broken URL => *probably* spammer

Dealing with bad workers

- Pay for “bad” work instead of rejecting it?
  - Pro: preserve reputation, admit if poor design at fault
  - Con: promote fraud, undermine approval rating system
- Use bonus as incentive
  - Pay the minimum $0.01 and $0.01 for bonus
  - Better than rejecting a $0.02 task
- If spammer “caught”, block from future tasks
  - May be easier to always pay, then block as needed

Build Your Reputation as a Requestor

- Word of mouth effect
  - Workers trust the requester (pay on time, clear explanation if there is a rejection)
  - Experiments tend to go faster
  - Announce forthcoming tasks (e.g. tweet)

Answer justification

- Why settle for a label?
- Let workers justify answers
- INEX (Initiative for the Evaluation of XML Retrieval)
  - 22% of assignments with comments
- Has to be optional for good feedback

Gamification of IR Evaluation

- GeAnn: http://www.geann.org/

- Relevance judgments with Gamification:
  - Text relevance
  - Image relevance

Quality through Flow and Immersion: **Gamifying Crowdsourced Relevance Assessments**. Eickhoff, C., C. G. Harris, A. P. de Vries, and P. Srinivasan. SIGIR 2012.

Summary

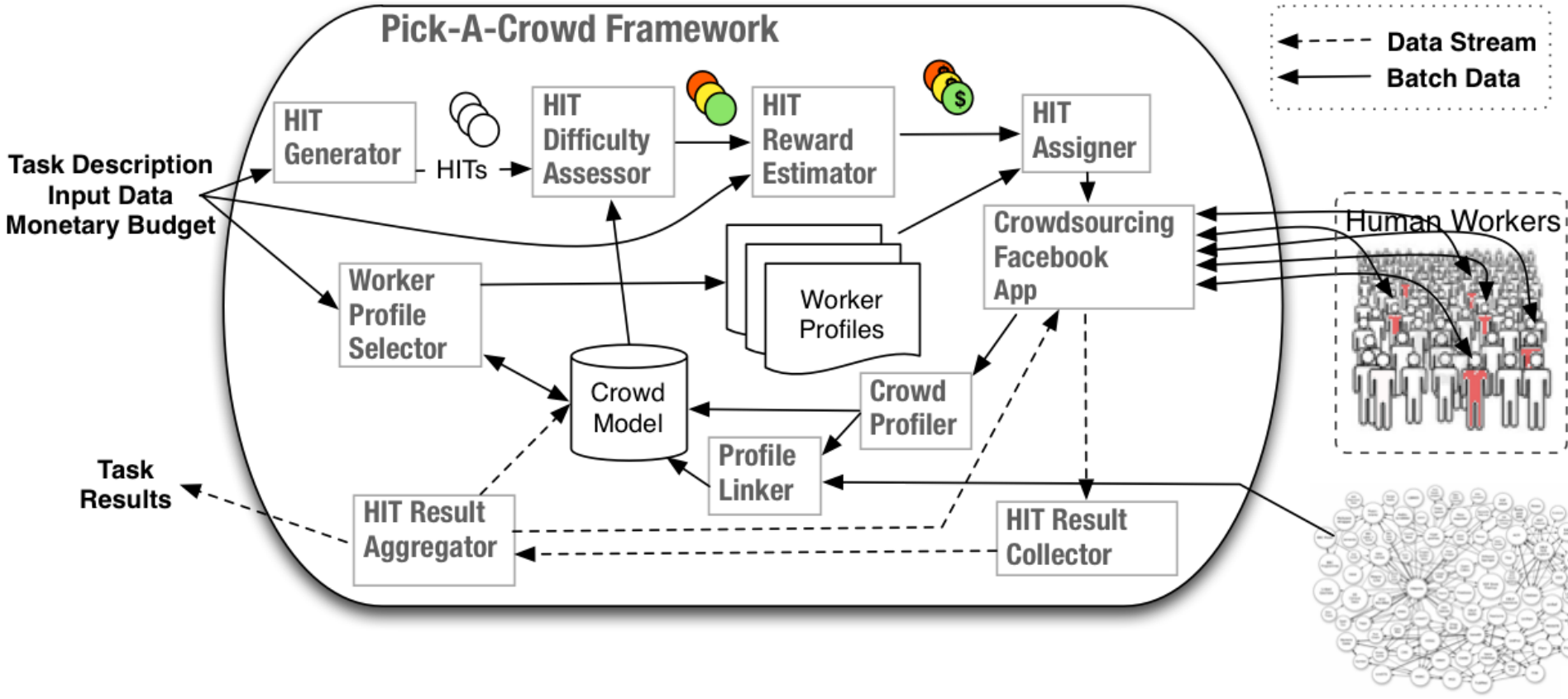
- Things that work
  - Qualification tests
  - Honey-pots
  - Good content and good presentation
  - Economy of attention
- Things to improve
  - Manage workers in different levels of expertise including spammers and potential cases.
  - Mix different pools of workers based on different profile and expertise levels.

References

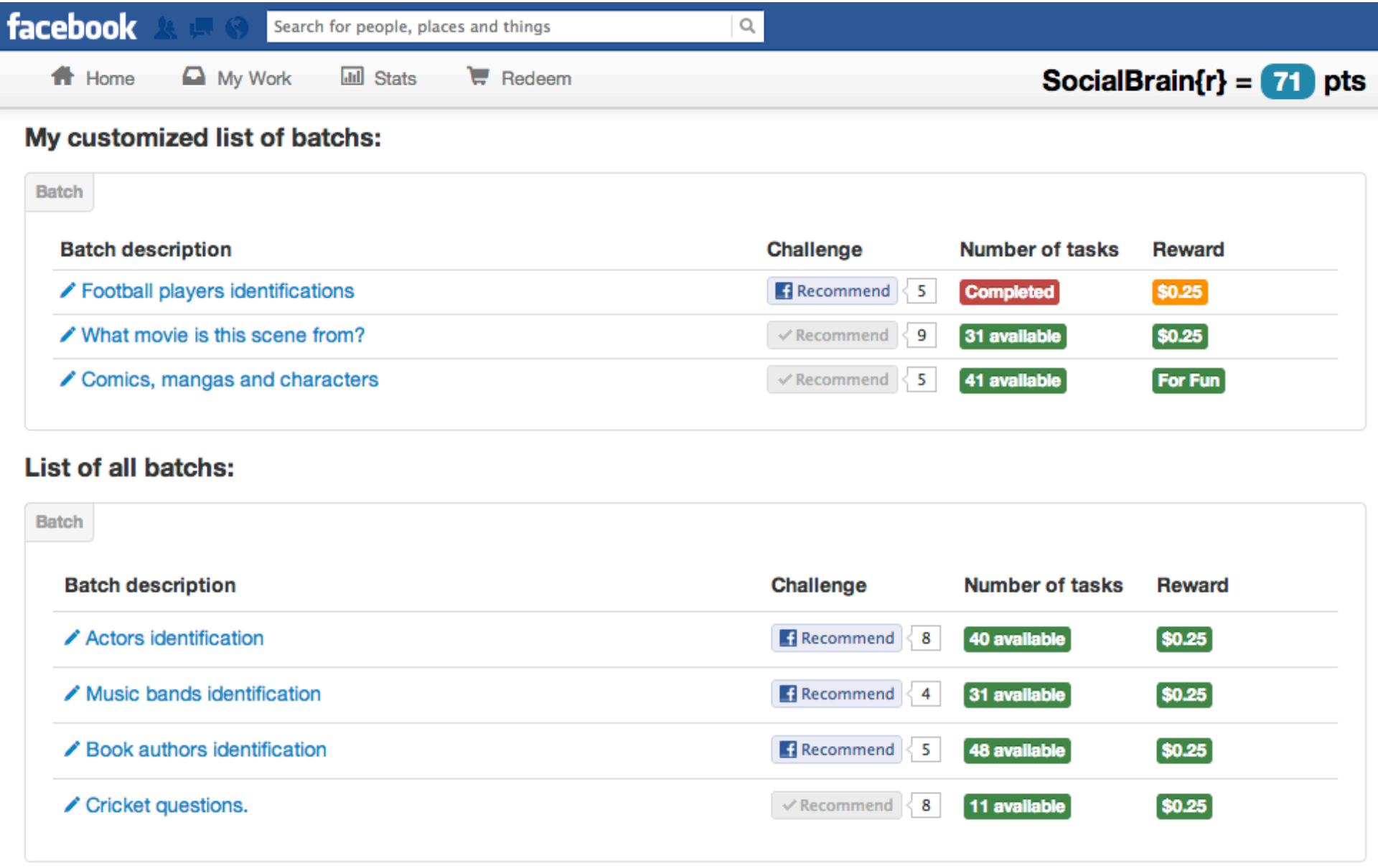
- "**Crowdsourcing for Information Retrieval: Principles, Methods, and Applications**" SIGIR 2011 Tutorial.
- "**Crowdsourcing for Search Evaluation and Social-Algorithmic Search**" SIGIR 2012 Tutorial.
- *Crowdsourcing Applications and Platforms: A Data Management Perspective.* A. Doan, M. J. Franklin, D. Kossmann, T. Kraska, VLDB 2011 (Tutorial).

Modeling Crowd Workers (via Social Network Profiles)

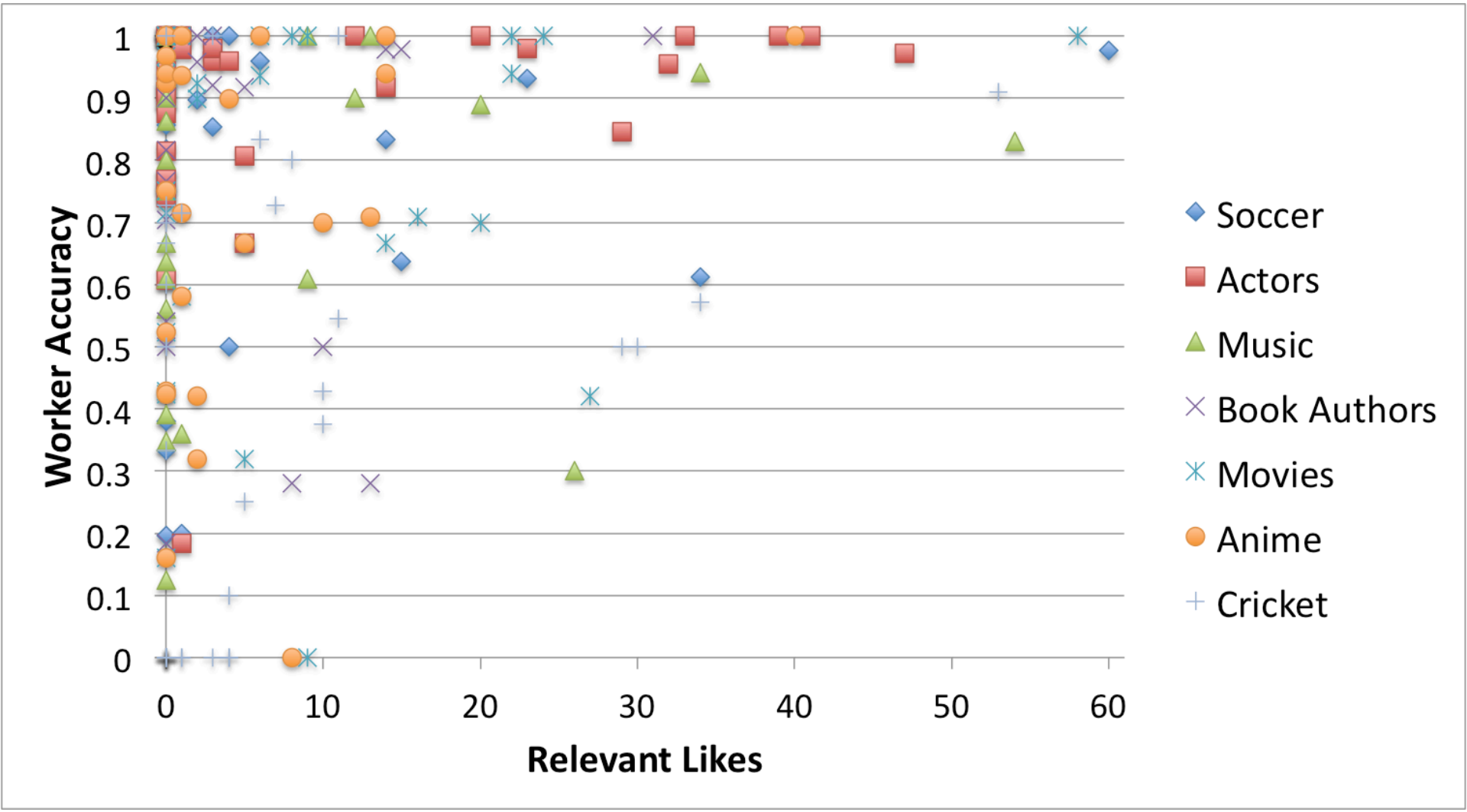
Pick-A-Crowd



Djellel Eddine Difallah, Gianluca Demartini, and Philippe Cudré-Mauroux.
Pick-A-Crowd: Tell Me What You Like, and I'll Tell You What to Do.
In: 22nd International Conference on World Wide Web (WWW 2013)


facebook
Search for people, places and things
Home
My Work
Stats
Redeem
SocialBrain{r} = 71 pts
My customized list of batchs:
Batch
Batch description
Challenge
Number of tasks
Reward
Football players identifications
Recommend 5
Completed
$0.25
What movie is this scene from?
Recommend 9
31 available
$0.25
Comics, mangas and characters
Recommend 5
41 available
For Fun
List of all batchs:
Batch
Batch description
Challenge
Number of tasks
Reward
Actors identification
Recommend 8
40 available
$0.25
Music bands identification
Recommend 4
31 available
$0.25
Book authors identification
Recommend 5
48 available
$0.25
Cricket questions.
Recommend 8
11 available
$0.25

Like vs Accuracy



Launching soon on…

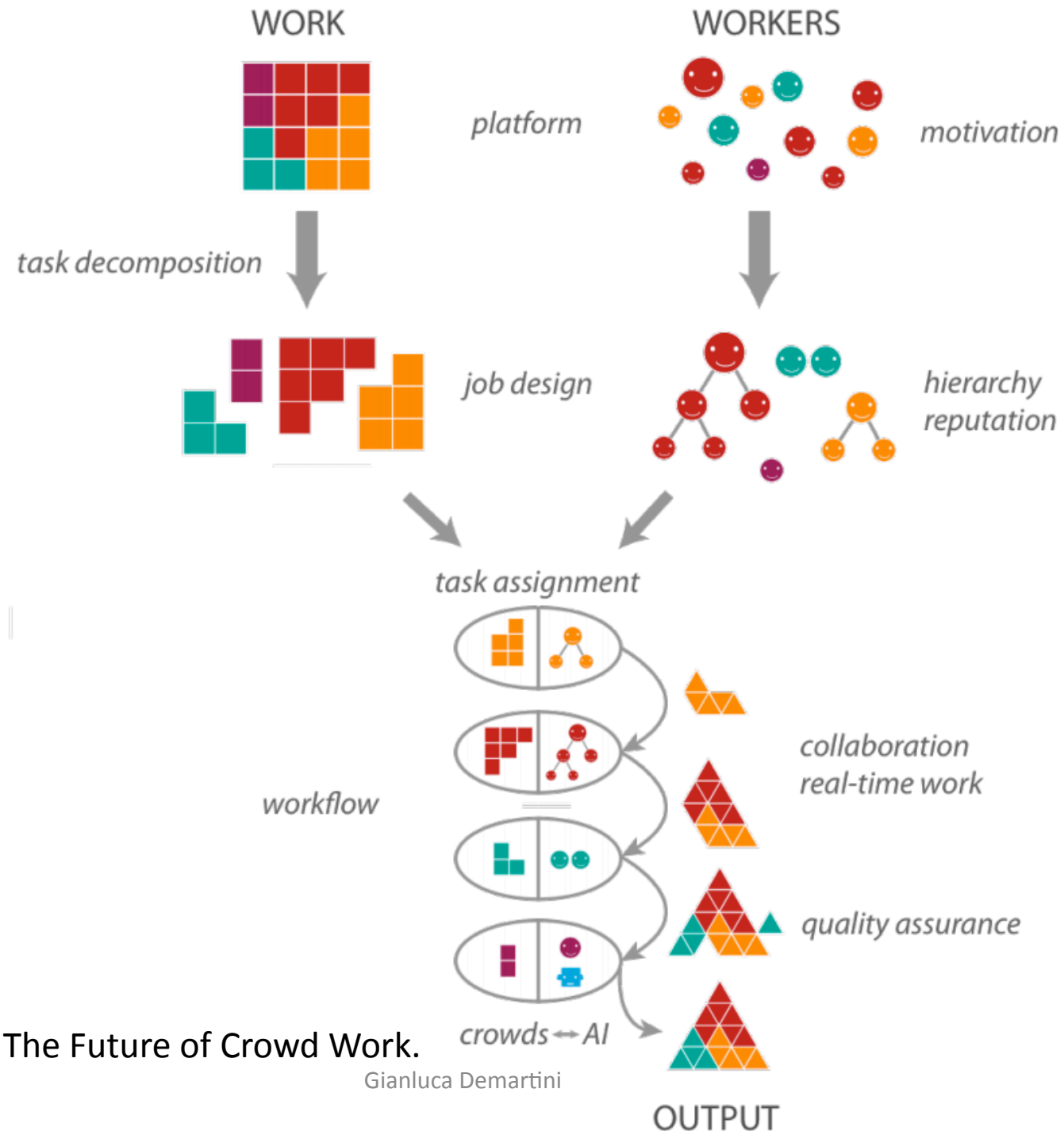
openturk.com

The Future of Crowd Work

How to obtain quality work

on the long term

(aka Open Research Directions)



Aniket Kittur et al. The Future of Crowd Work. CSCW 2013.

Recommendations

- Reputation system for workers
- More than financial incentives
- Recognize worker potential (badges)
  - Paid for their expertise
- Train less skilled workers (tutoring system)

Recommendations

- Promote workers to management roles
  - Create gold labels
  - Manage other workers
  - Make task design suggestions (first-pass validation)
- Career trajectory (based on reputation):
  1. Untrusted worker
  2. Trusted worker
  3. Hourly contractor
  4. Employee
- Platforms logs
  - Which kind of tasks attract skilled workers

Summary

- Enforce Quality:
  - Task design
  - Iterate
  - Crowd incentives
  - Know your crowd: Model workers

HUMAN COMPUTATION AND THE SEMANTIC WEB

ELENA SIMPERL

UNIVERSITY OF SOUTHAMPTON, UK

WHAT IS DIFFERENT ABOUT SEMANTIC SYSTEMS?

Semantic Web tools vs. applications

- Intelligent (specialized) Web sites (portals) with improved (local) search based on vocabularies and ontologies
- X2X integration (often combined with Web services)
- Knowledge representation, communication and exchange

WHAT DO YOU WANT YOUR USERS TO DO?

- **Semantic applications**
  - Context of the actual application
  - Need to involve users in knowledge acquisition and engineering tasks?
    - Incentives are related to organizational and social factors
    - Seamless integration of new features
- **Semantic tools (e.g., Linked Data publishing, ontology editing)**
  - Game mechanics
  - Paid crowdsourcing (integrated)
- **Using results of games with a purpose**

THE LEVEL OF TASKS FOUND IN METHODOLOGIES NEEDS FURTHER REFINEMENT

Crowdsource very specific tasks that are (highly) divisible

- Labeling (in different languages)
- Finding relationships
- Populating the ontology
- Aligning and interlinking
- Ontology-based annotation
- Validating the results of automatic methods
- …

Think about the context of the application (social structure) and about how to hide tasks behind existing practices and tools

INTERPLAY OF INCENTIVES AND MOTIVATION ACHIEVES MAXIMAL RESULTS

Focus on the actual goal and incentivize related actions

- Write posts, create graphics, annotate pictures, reply to customers in a given time…

Build a community around the intended actions

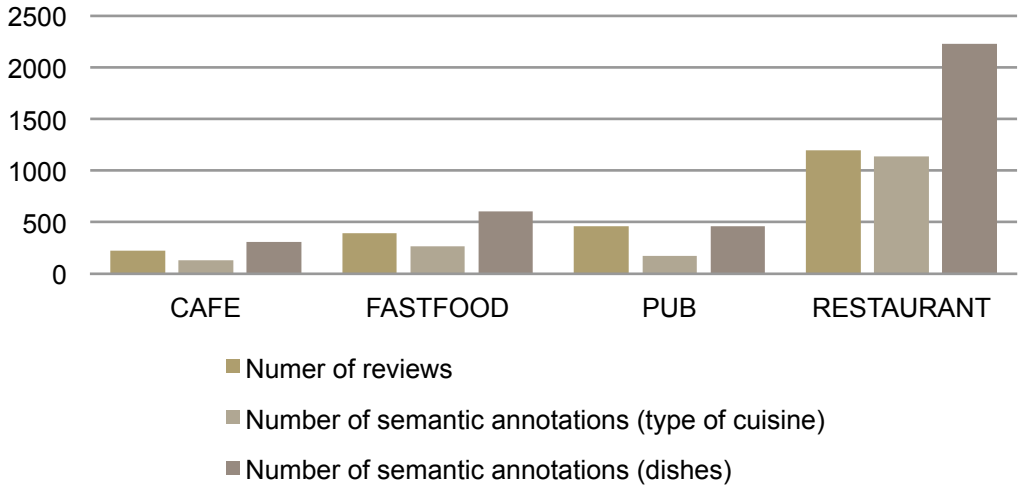
- Reward helping each other in performing the task and interaction
- Reward recruiting new contributors

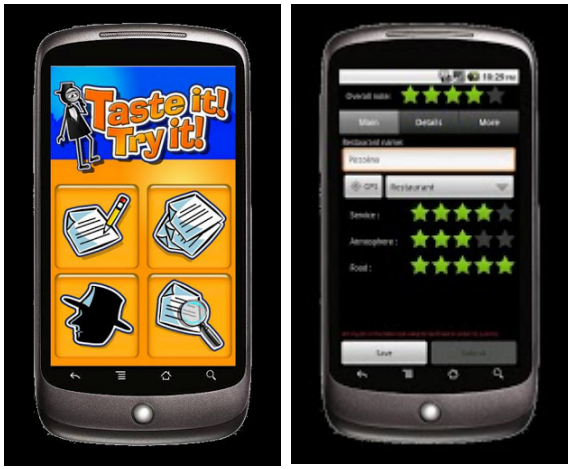
Reward repeated actions

- Actions become part of the daily routine

TASTE IT! TRY IT!

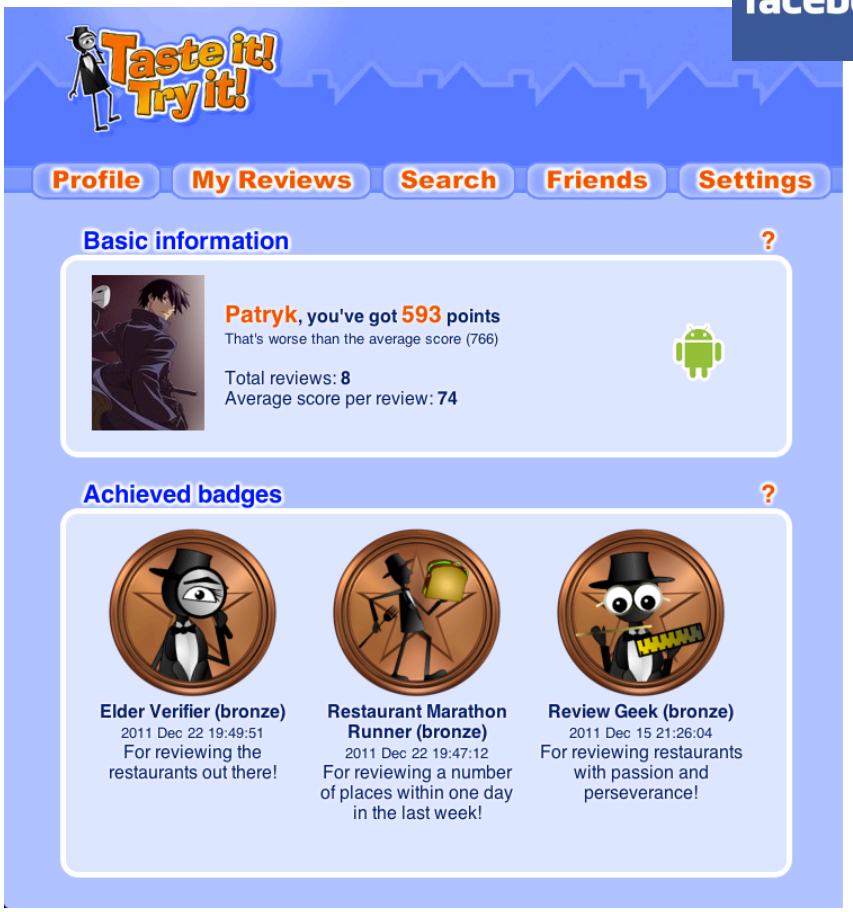
- **Restaurant review Android app developed in the Insemtives project**
- **Uses Dbpedia concepts to generate structured reviews**
- **Uses mechanism design/gamification to configure incentives**
- **User study**
  - 2274 reviews by 180 reviewers referring to 900 restaurants, using 5667 Dbpedia concepts





https://play.google.com/store/apps/details?id=insemtives.android&hl=en

SOCIABILITY DESIGN ASPECTS



MECHANISM DESIGN EXPERIMENTS

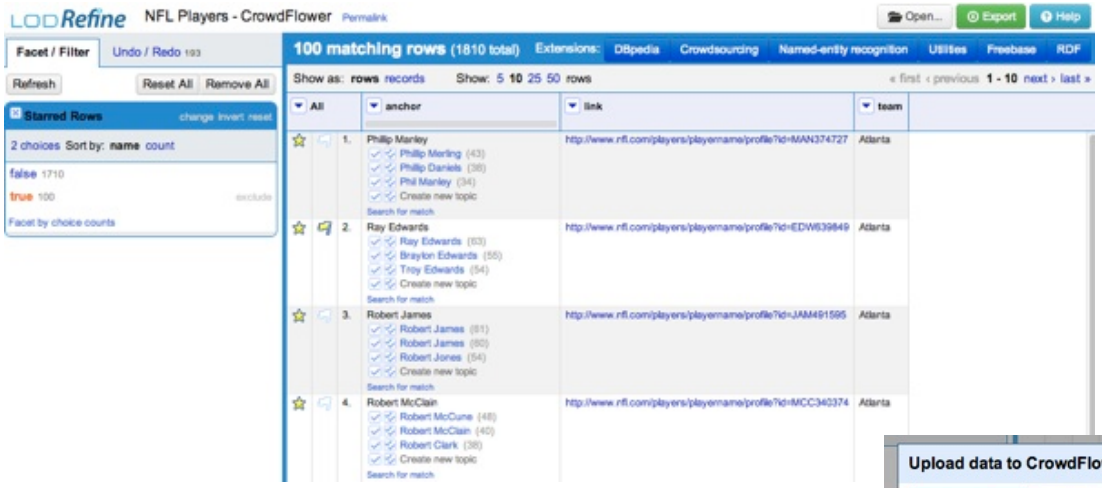
Two experiments: 150 and 30 students

- Points vs. badges
- No information about others vs. information about others (neighborhood, median, full leaderboard)

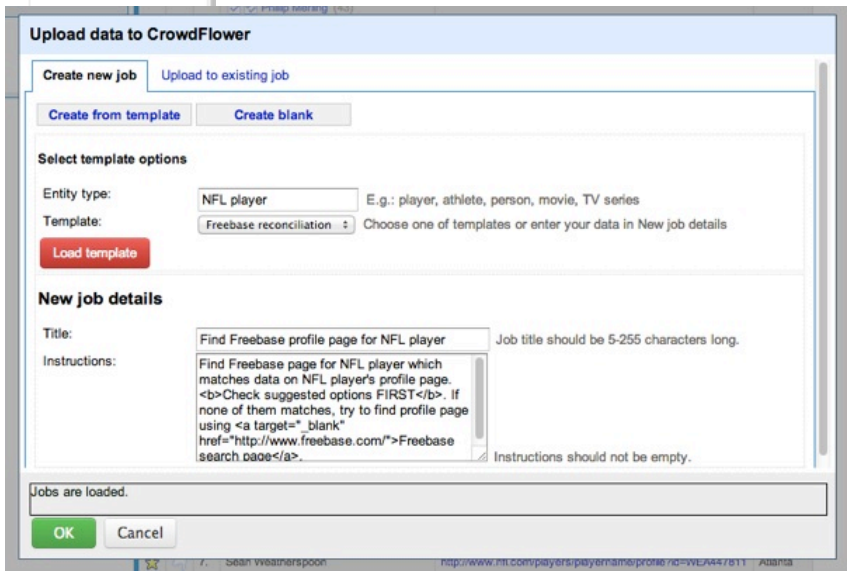
Findings

- Presenting information on performance of peers helps to increase the number of reviews
- Within the treatments with badges individuals tend to contribute more compared to treatments without assignment of badges

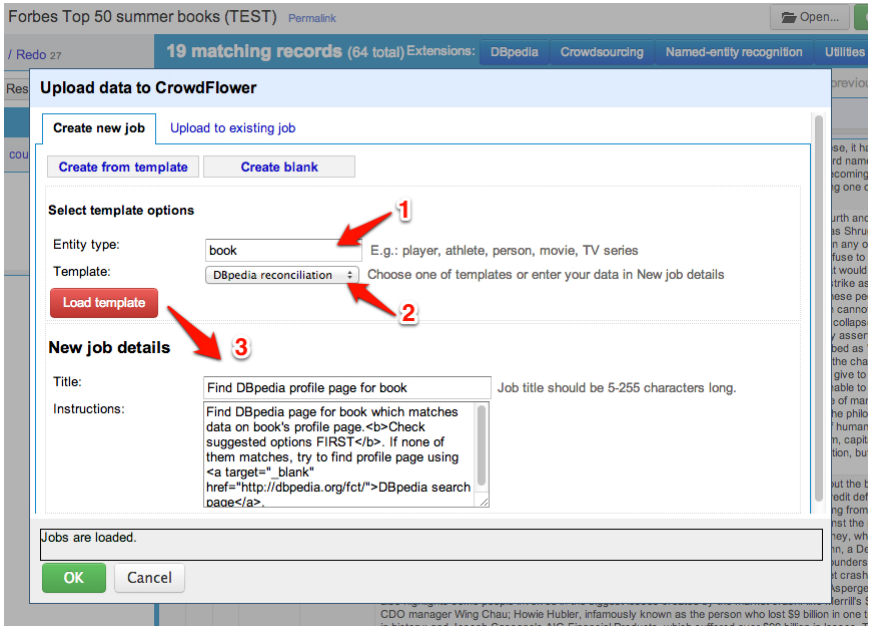
LODREFINE

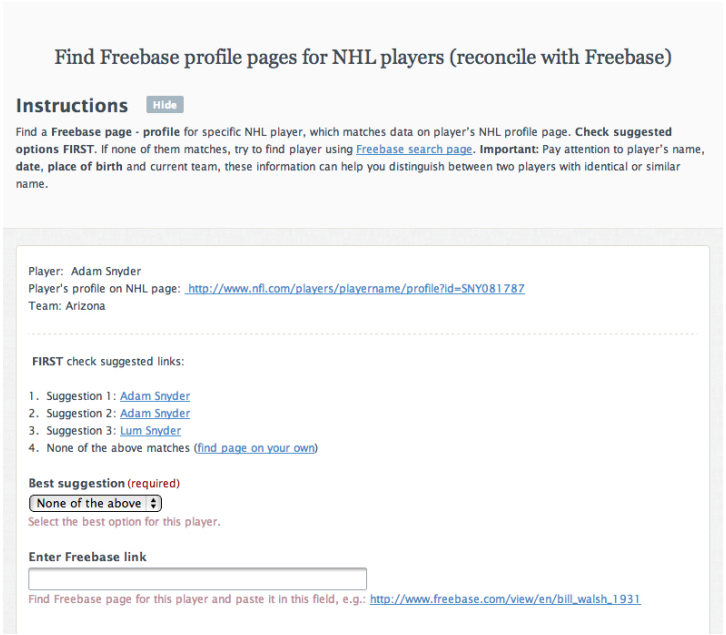


Extension of LODRefine to enhance automatic data reconciliation algorithms using CrowdFlower



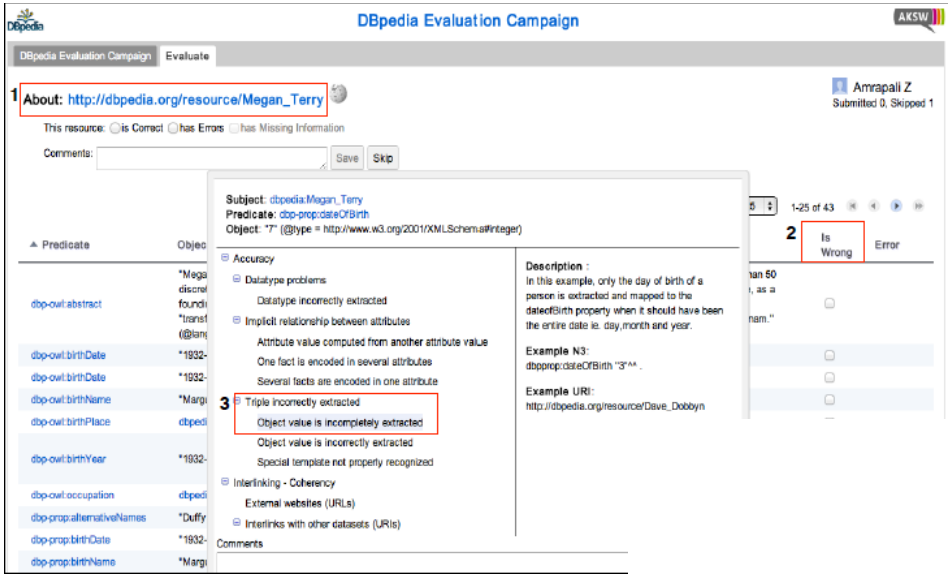
LODREFINE (2)

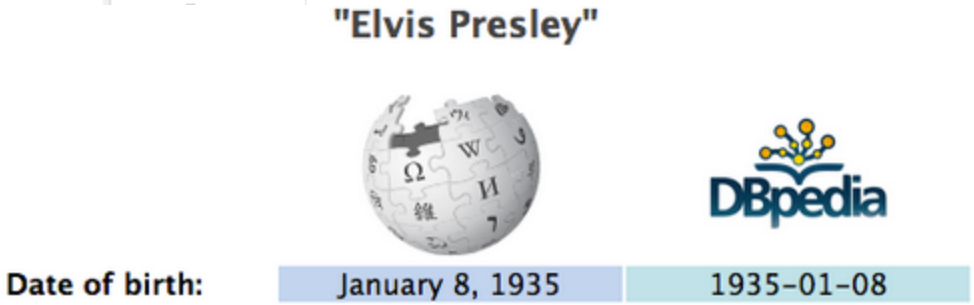




http://research.zemanta.com/crowds-to-the-rescue/

DBPEDIA CURATION



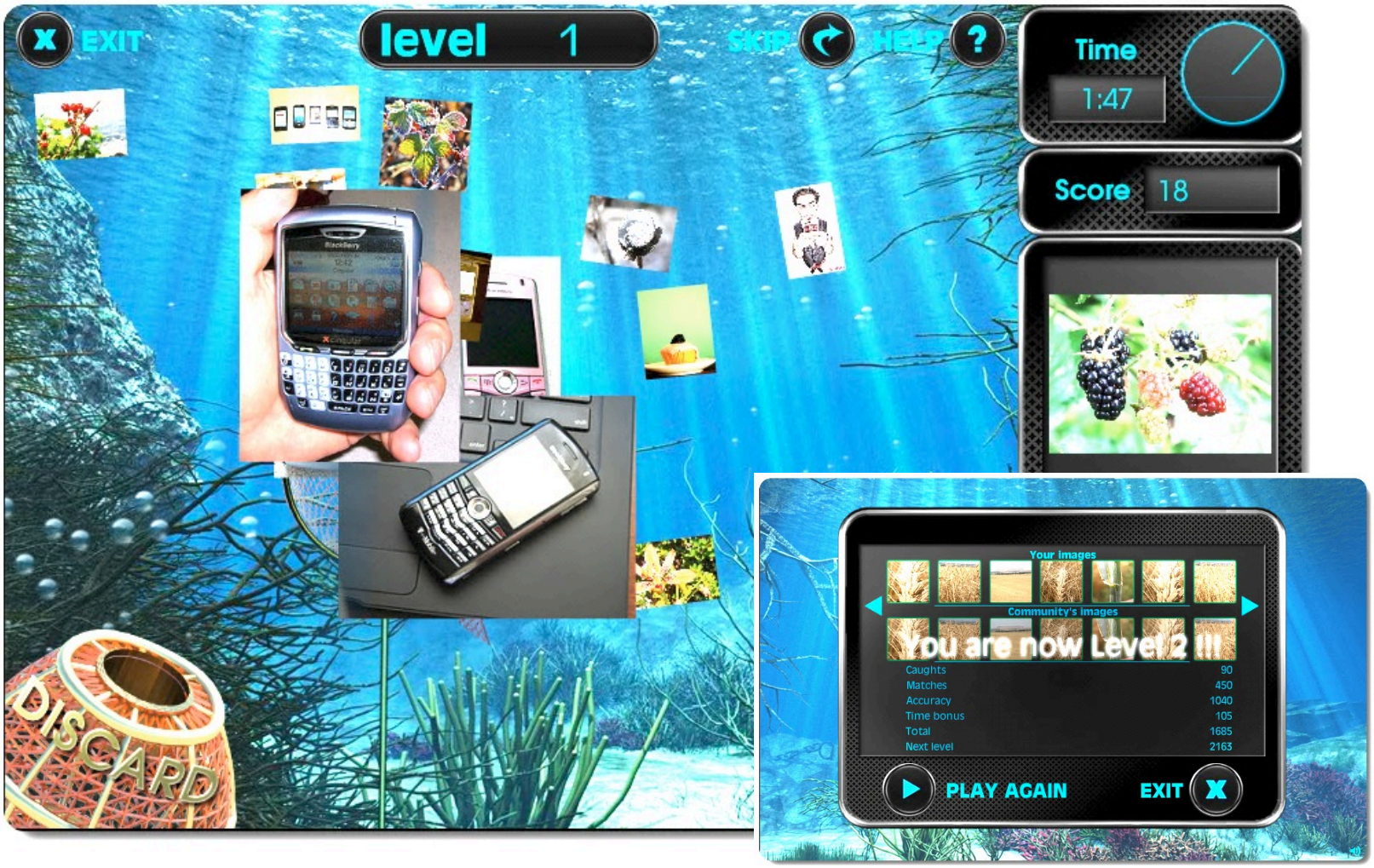


http://aksw.org/Projects/TripleCheckMate.html

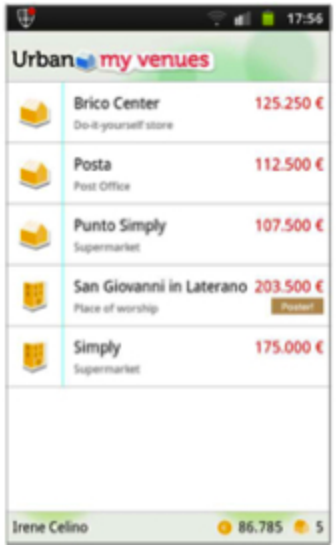
ONTOGY BUILDING

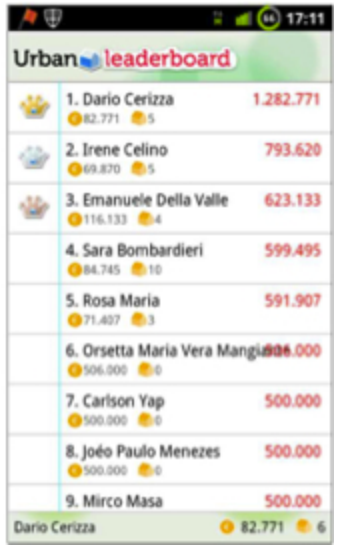
RELATIONSHIP FINDING

MULTIMEDIA INTERLINKING

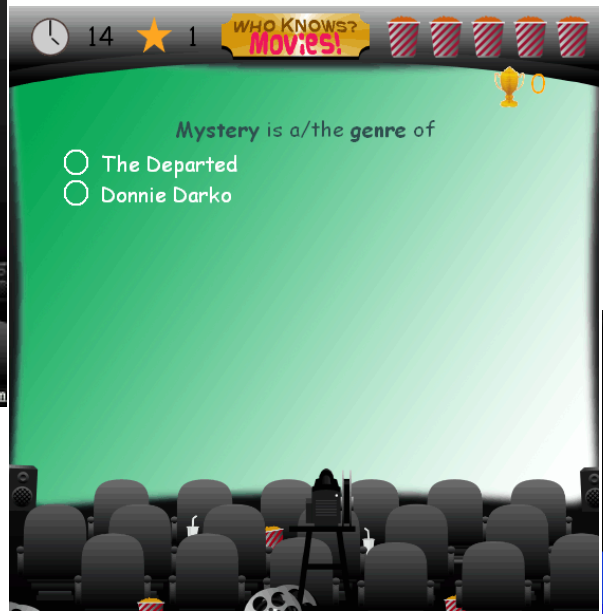


LINKED DATA CURATION





ENTITY SUMMARIZATION



REUSING CROWDSOURCING RESULTS

- **Ongoing work:**
  - Vocabulary to describe and exchange crowdsourcing results
  - Including
    - Type of crowdsourcing approach
    - Crowd
    - Inputs and outputs
    - Confidence values
    - Quality assurance method applied
    - ...