Introduction to Social Computing
Human Computation

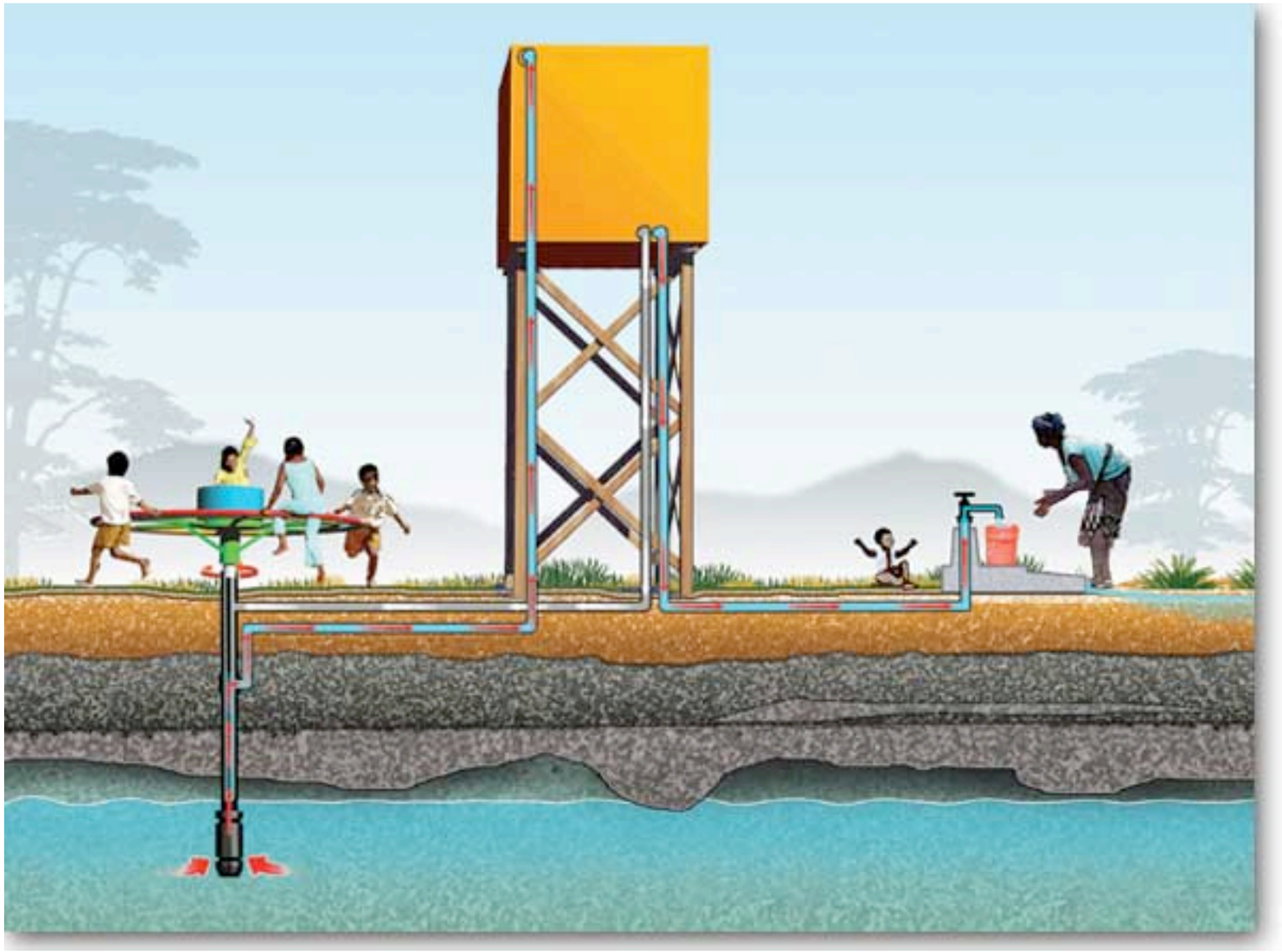
Irwin King

ATT Labs, Research
&
Department of Computer Science and Engineering
The Chinese University of Hong Kong

king@cse.cuhk.edu.hk
http://www.cse.cuhk.edu.hk/~king

Playing/Having Fun ⟷? Work/Computation

Idea of Human Computation



- Take advantage of people's desire to be entertained and perform useful tasks as a side effect

Motivations

- To describe the categorization of Human Computation Systems (HCS)
- To describe each category of HCS and present the previous work on each category
- To summarize the current state-of-the-art HCS

Why Is It Important?

- Some statistics (July 2008)
  - 200,000+ players have contributed 50+ million labels.
  - Each player plays for a total of 91 minutes.
  - The throughput is about 233 labels/player/hour (i.e., one label every 15 seconds)
- Idea behind
  - Solve some problems which are difficult to be solved by computers.
  - Take advantage of people's desire to be entertained.
  - Produce useful metadata as a by-product.

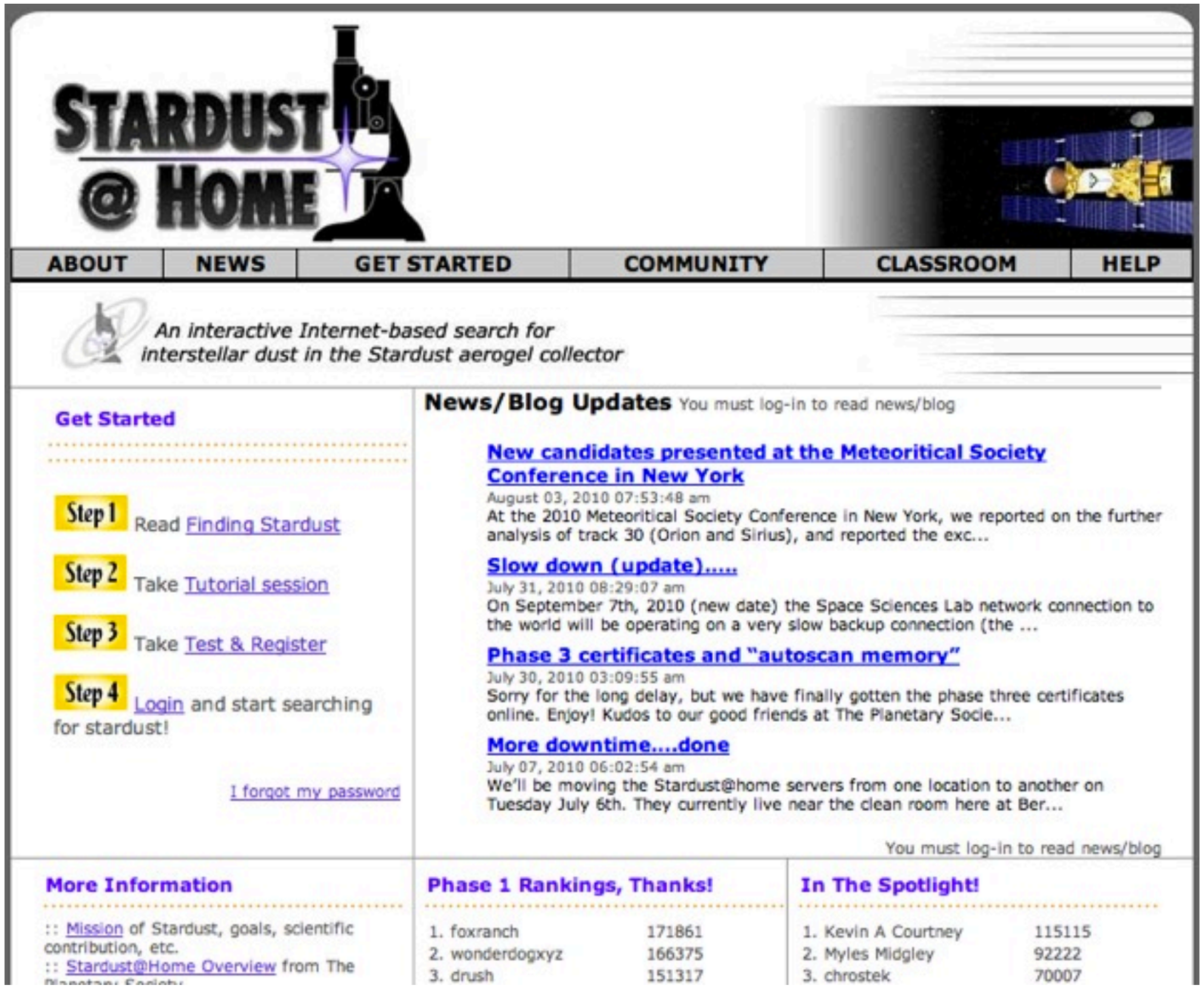
Outline

- Motivation and Background
- Types of Human Computation
  - Initiatory Human Computation
  - Distributed Human Computation
    - Social Game-based Human Computation with volunteers or paid engineers
    - Social Game-based Human Computation with online players
- Properties of Social Games
- Future Work and Final Remarks

SETI


SETI INSTITUTE
Welcome to SETI.org!
Search:
GO
Home || About us || SETI || Astrobiology || Education || Publications || Calendar || Support us || TeamSETI || Contact us
THE MISSION OF THE
SETI INSTITUTE
IS TO EXPLORE, UNDERSTAND AND EXPLAIN
THE ORIGIN, NATURE AND PREVALENCE
OF LIFE IN THE UNIVERSE.
Learn more >
Thank you, Team!
You're invited...
TeamSETI Appreciation Party
2010 Ice Cream Social
October 31, 2010
1:00pm - 3:00pm
Register Today
Come see our new headquarters!
189 N. Bernardo Ave, suite 100
Mountain View CA 94043
Contribute to the Search
setiquest
Join the Quest
TeamSETI
Support our science
Become a member
SETI
ASTROBIOLOGY
EDUCATION & OUTREACH
ALOHA ASTRONOMY
RADIO SHOW
TEAM SETI
JOIN
DONATE
LOGIN
Newsletter
Facebook
twitter
ADOPT A SCIENTIST
COLLOQUIUM SERIES
KEPLER MISSION
LIFE AT THE INSTITUTE
News
Meet our scientists: Paul Estrada — From Dust to Planets

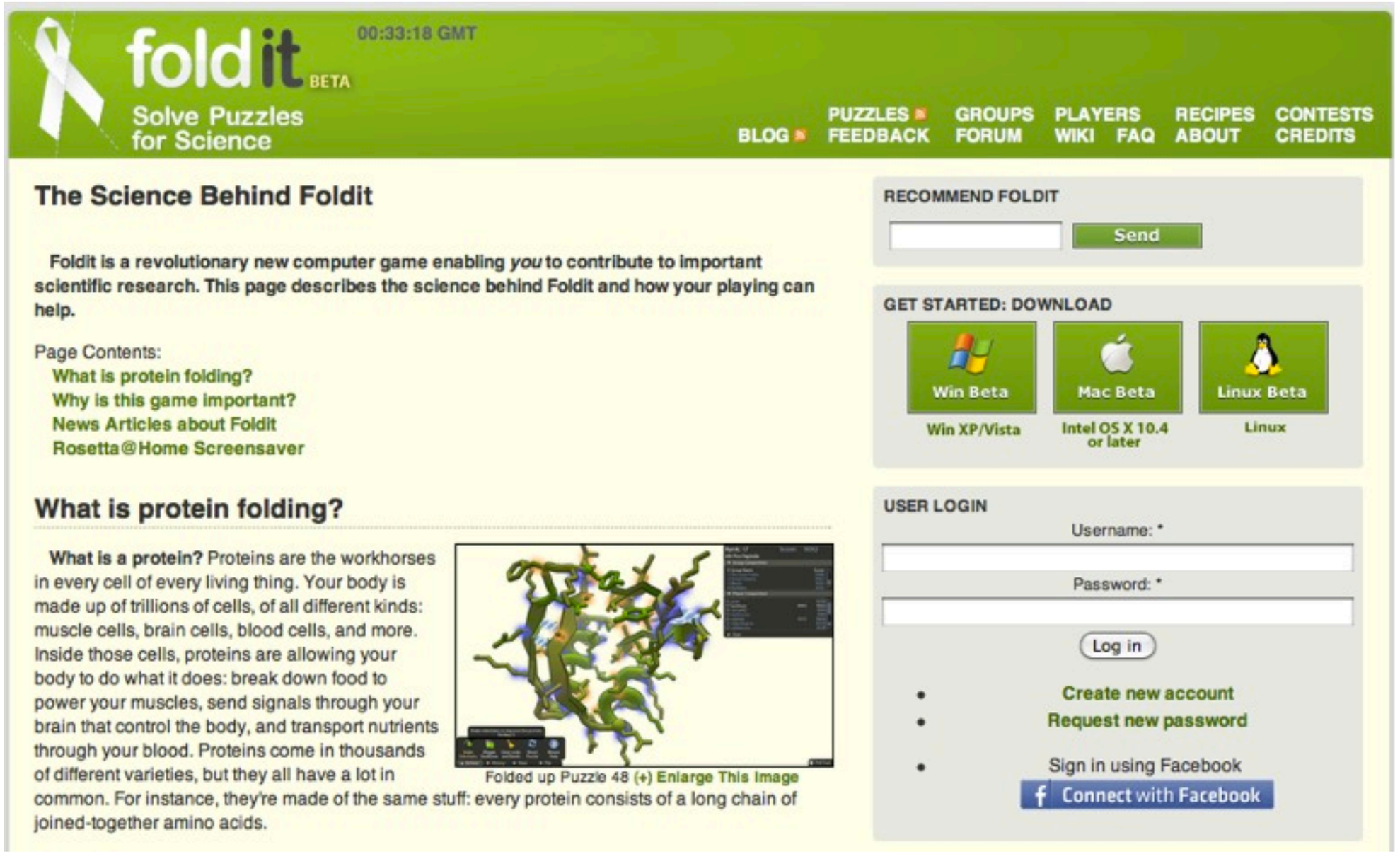
Startdust@home



Galaxy Zoo


EN · Galaxy Zoo is a ZOONIVERSE project
...just like MOON ZOO
GALAXY ZOO
HUBBLE
Home
The Story So Far
The Science
How To Take Part
Classify Galaxies
Forum
Zoo Media
Blog
FAQ
Contact Us
Pictures
Welcome to Galaxy Zoo, where you can help astronomers explore the Universe
Galaxy Zoo: Hubble uses gorgeous imagery of hundreds of thousands of galaxies drawn from NASA's Hubble Space Telescope archive. To understand how these galaxies, and our own, formed we need your help to classify them according to their shapes — a task at which your brain is
Classifier Log In
Click here to log in
Register
Forgotten Password?

Foldit: Protein Folding Game



[Cooper et al, Nature 466, 756-760 (5 August 2010)]

Background

- Human Computation Systems (HCS) aim to solve Artificial Intelligence (AI) problems through the human human interactions

- In order to ensure the collected information to be useful, we have to:

1. guarantee the quality of collected information
2. attract more people to contribute information

Types of HCS

- The categories of the human computation systems are:
  1. Initiatory Human Computation
  2. Distributed Human Computation
  3. Social Game-based Human Computation with volunteers or paid engineers
  4. Social Game-based Human Computation with online players

Initiatory Human Computation (1)

- Objective: To complete some tasks that are natural for humans but difficult for computers even computation power increased rapid recently

- Example (1): CAPTCHA
  - A computer generated challenge-response test
  - Objective: To distinguish humans from computers using a common sense problem

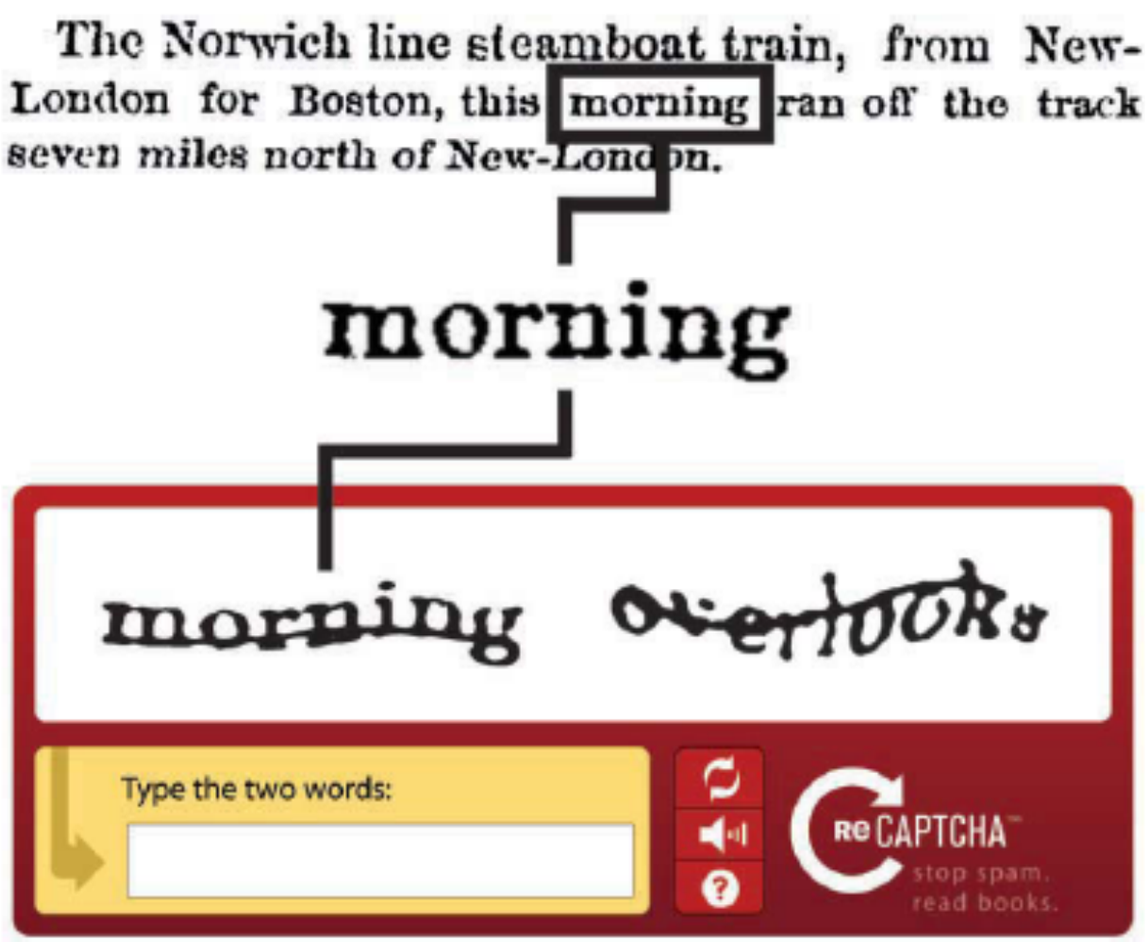
The Yahoo! CAPTCHA.

Initiatory Human Computation (2)

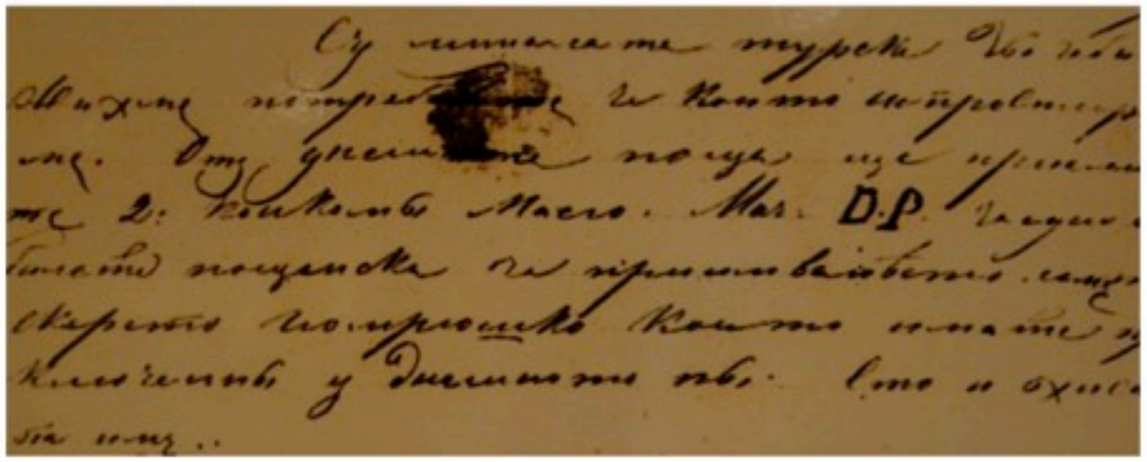
- Example (2): reCAPTCHA
  - Objective: To produce valuable common sense knowledge to improve the OCR quality in digitizing books
  - Combining two words: one identified word; and one unidentified word
  - If a user recognizes the identified word, the answer to the unidentified word is assumed to be correct

Initiatory Human Computation (3)

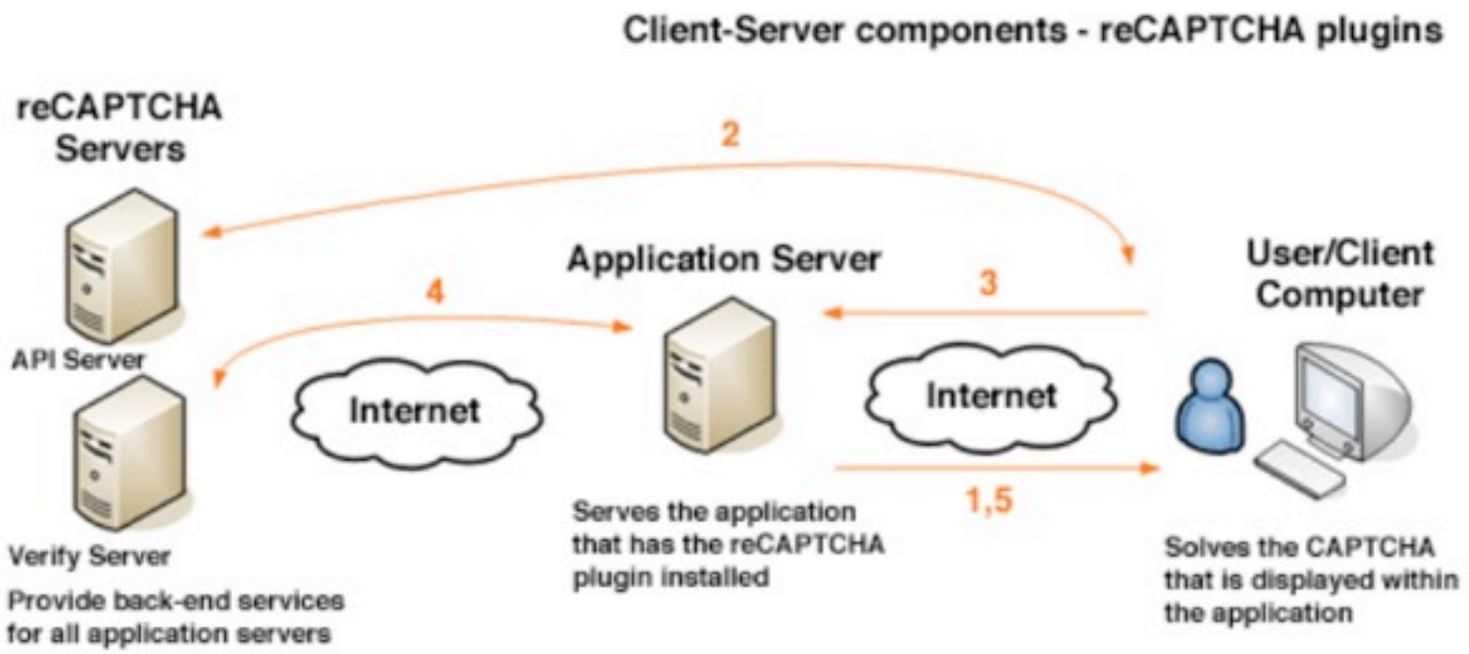
- Example (2): reCAPTCHA



reCAPTCHA

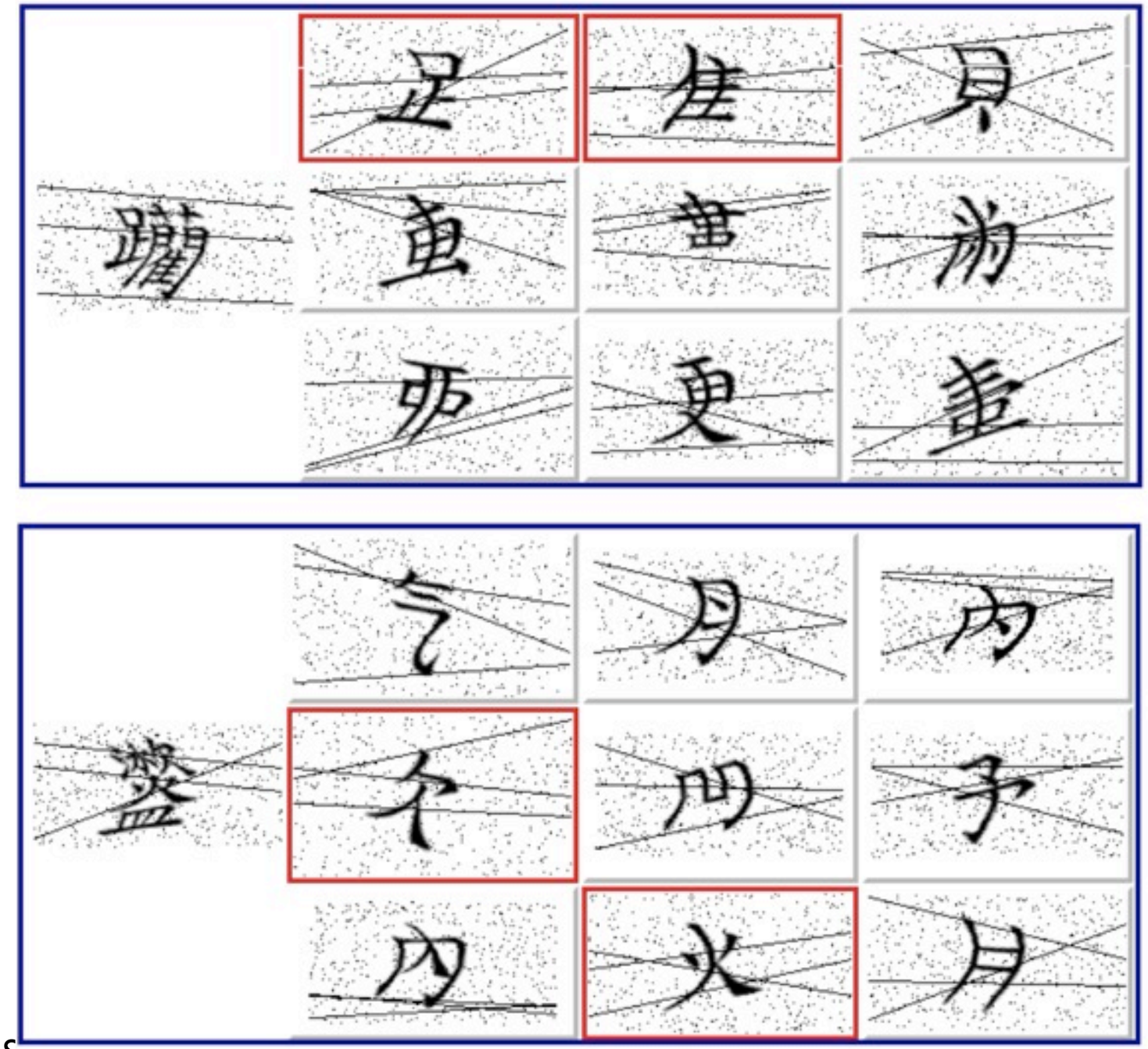



olcott
have
Type the two words:
reCAPTCHA™
stop spam.
read books.
Submit
The words above come from scanned books.
By typing them, you help to digitize old texts.


Client-Server components - reCAPTCHA plugins
reCAPTCHA Servers
API Server
Verify Server
Provide back-end services for all application servers
2
4
Internet
Application Server
Serves the application that has the reCAPTCHA plugin installed
3
Internet
1,5
User/Client Computer
Solves the CAPTCHA that is displayed within the application

Chinese CAPTCHA

Ling-Jyh Chen, Institute of Information Science, Academia Sinica, Taipei, Taiwan

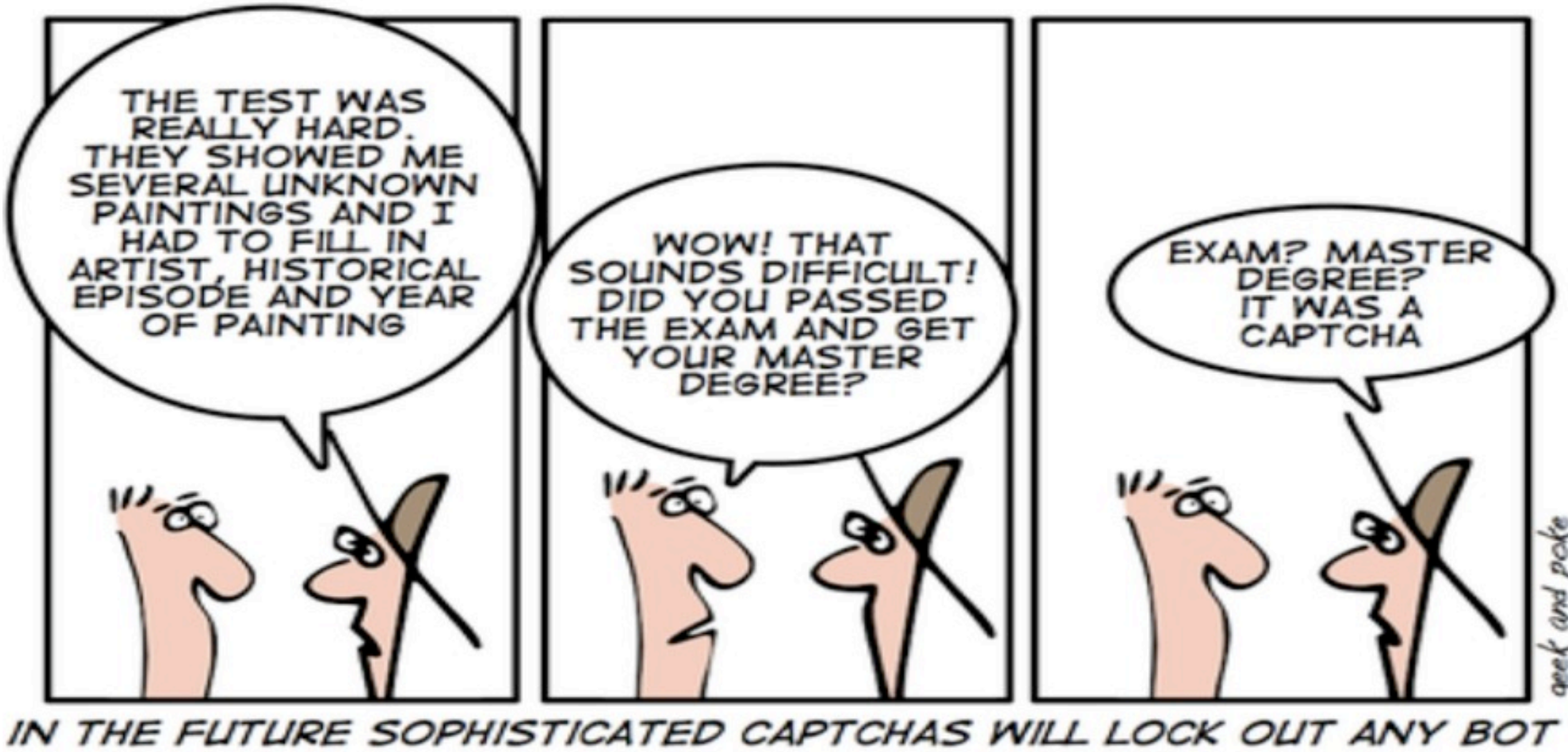


System Design Issues

- Centralized vs. distributed systems
- Single vs. multiple players per round
- Single vs. multiple outcomes per round
- Pure vs. computer-aided HCOMP
- Stationary vs. mobile players
- “just enough” incentives
- Not “just another” HCOMP system

Initiatory Human Computation (4)

- Example (3): KA-CAPTCHA
  - Objective: To collect every correct answer submitted by humans to the CAPTCHA test as a solution to a problem that computers are unable to solve
  - CAPTCHA solvers are highly interested in providing a valid response to the CAPTCHA test (because they want to access the protected resource)
  - Knowledge acquisition mechanism: To strategically asking for a solution to a particular open problem that is of interest to the CAPTCHA designer.


THE TEST WAS REALLY HARD. THEY SHOWED ME SEVERAL UNKNOWN PAINTINGS AND I HAD TO FILL IN ARTIST, HISTORICAL EPISODE AND YEAR OF PAINTING
WOW! THAT SOUNDS DIFFICULT! DID YOU PASSED THE EXAM AND GET YOUR MASTER DEGREE?
EXAM? MASTER DEGREE? IT WAS A CAPTCHA
geek and poke
IN THE FUTURE SOPHISTICATED CAPTCHAS WILL LOCK OUT ANY BOT

Distributed Human Computation (1)

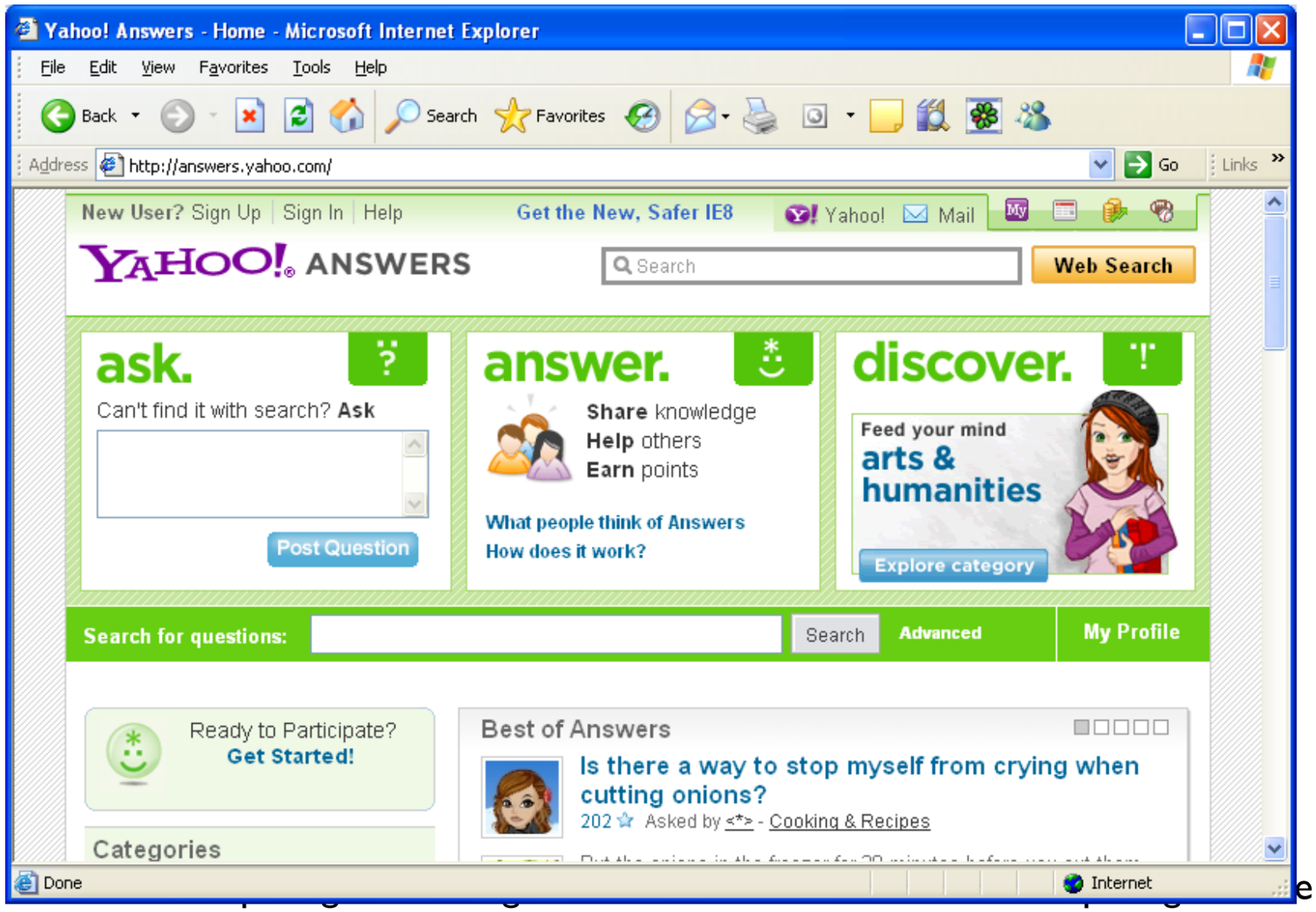
- Objective: To encourage a huge population of Internet users to contribute to solve the difficult AI problems
- Example (1): Razor
  - To use human votes to determine if a given email is spam (anti-spam mechanism)
- Example (2): Proofreader
  - To give a (small) portion of the image file and corresponding text (generated by OCR) side-by-side to a human proofreader

Distributed Human Computation (2)

- Example (3): Wikipedia
  - The collective knowledge is distributed in that essentially almost anyone can contribute to the Wiki

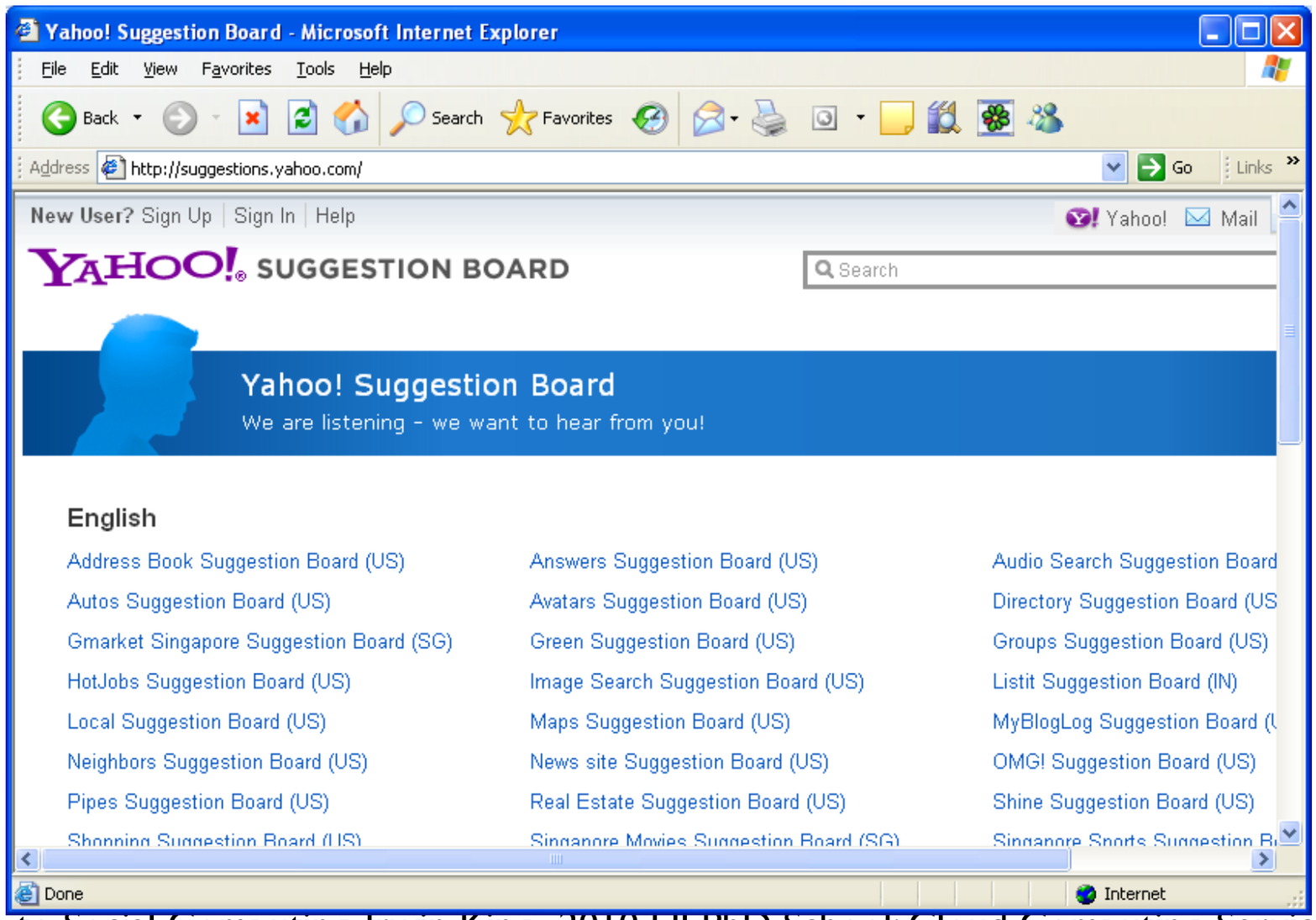
Distributed Human Computation (3)

- Example (4): Yahoo! Answers
  - To provide automated collection of human reviewed data at Internet-scale



Distributed Human Computation (4)

- Example (5): Yahoo! Suggestion Board
  - An Internet-scale feedback and suggestion system



Distributed Human Computation (5)

- Example (6): Amazon Mechanical Turk
  - It provides monetary rewards for tasks
- Example (7): LabelMe
  - A web-based tool for image annotation
  - Anybody can annotate image using it. You can only have access to the database once you have annotated a certain number of images.
- Example (8): 43Things
  - To collect goals from users and help them to find other users who have similar goals
- Example 9: MajorMiner
  - Music annotation game

Amazon Mechanical Turk

amazonmechanical turk beta
Artificial Artificial Intelligence

Your Account | HITs | Qualifications

Already have an account?
Sign in as a Worker | Requester

Introduction | **Dashboard** | **Status** | **Account Settings**

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

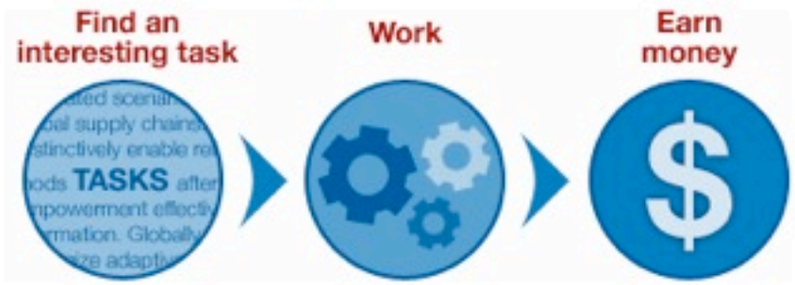
26,113 HITs available. View them now.

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



Find HITs Now

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

Get Started

or learn more about being a **Requester**

FAQ | Contact Us | Careers at Amazon | Developers | Press | Policies

An amazon.com company

Example of Mechanical Turk

Answer a short survey

1. What is your gender?

- ○ Male
- ○ Female

2. What is your age?

3. Which of the following best describes your highest achieved education level?

Some High School

4. What is the total income of your household?

Less than $12,500
$12,500 - $24,999
$25,000 - $37,499
$37,500 - $49,999

5. What is your favorite type of TV Show? (select all that apply)

- ☐ Sports
- ☐ Situational Comedies
- ☐ Drama
- ☐ News
- ☐ Music Videos

Find the Website Address for this Restaurant

- For this restaurant below, enter the website address for the official website of the restaurant
- Include the full address, e.g. http://www.thecheesecakefactory.com
- Do not include URLs to city guides and listings like Citysearch.

Restaurant Name: **${name}**

Address: **${address}**

Phone Number: **${phone}**

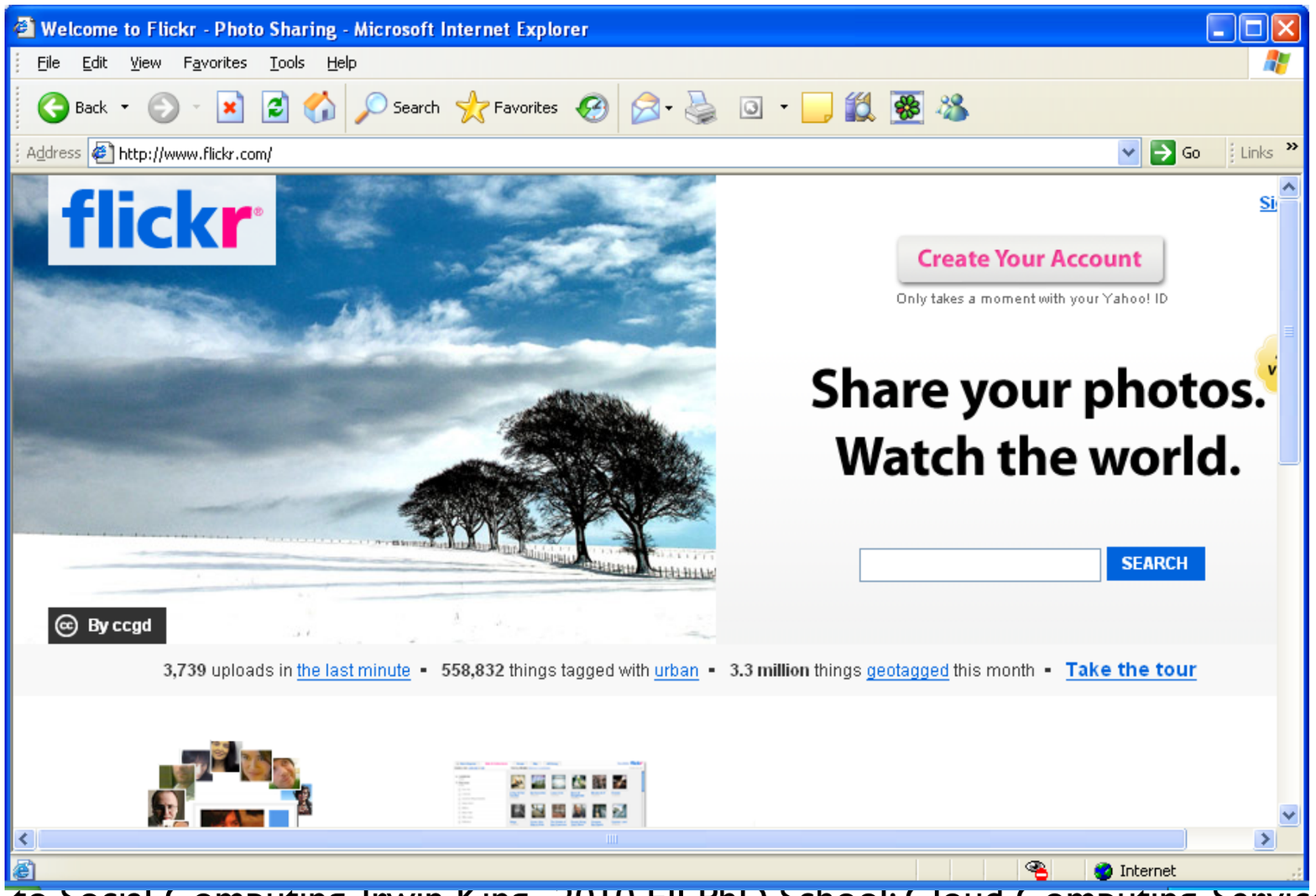
Website:

Please provide any comments you may have below, we appreciate your input!

Submit

Distributed Human Computation (6)

- Example (10): Yahoo's flickr
  - It is a photo-sharing site with captions being used as photo tags



Social Game-based Human Computation with volunteers or paid engineers (1)

- Recently social games were proposed to collect accurate information from players as a side effect of their playing
- The players are volunteers or paid engineers
- Disadvantages:
  - Rely on online volunteers or paid engineers to enter information explicitly
  - Unable to scale up the system due to high cost
  - No validation mechanism to guarantee that the information collected is accurate

Social Game-based Human Computation with volunteers or paid engineers (2)

- Most of the games at early stage aimed to collect commonsense knowledge.
- Example (1): Cyc
  - To collect information from the input by paid knowledge engineers
- Example (2): Open Mind
  - To collect common sense knowledge from people to develop intelligent software
  - Shortcoming: was too reliant on the unpaid volunteers to donate their time to contribute information

Social Game-based Human Computation with volunteers or paid engineers (3)

- Example (2): Open Mind

Social Game-based Human Computation with volunteers or paid engineers (4)

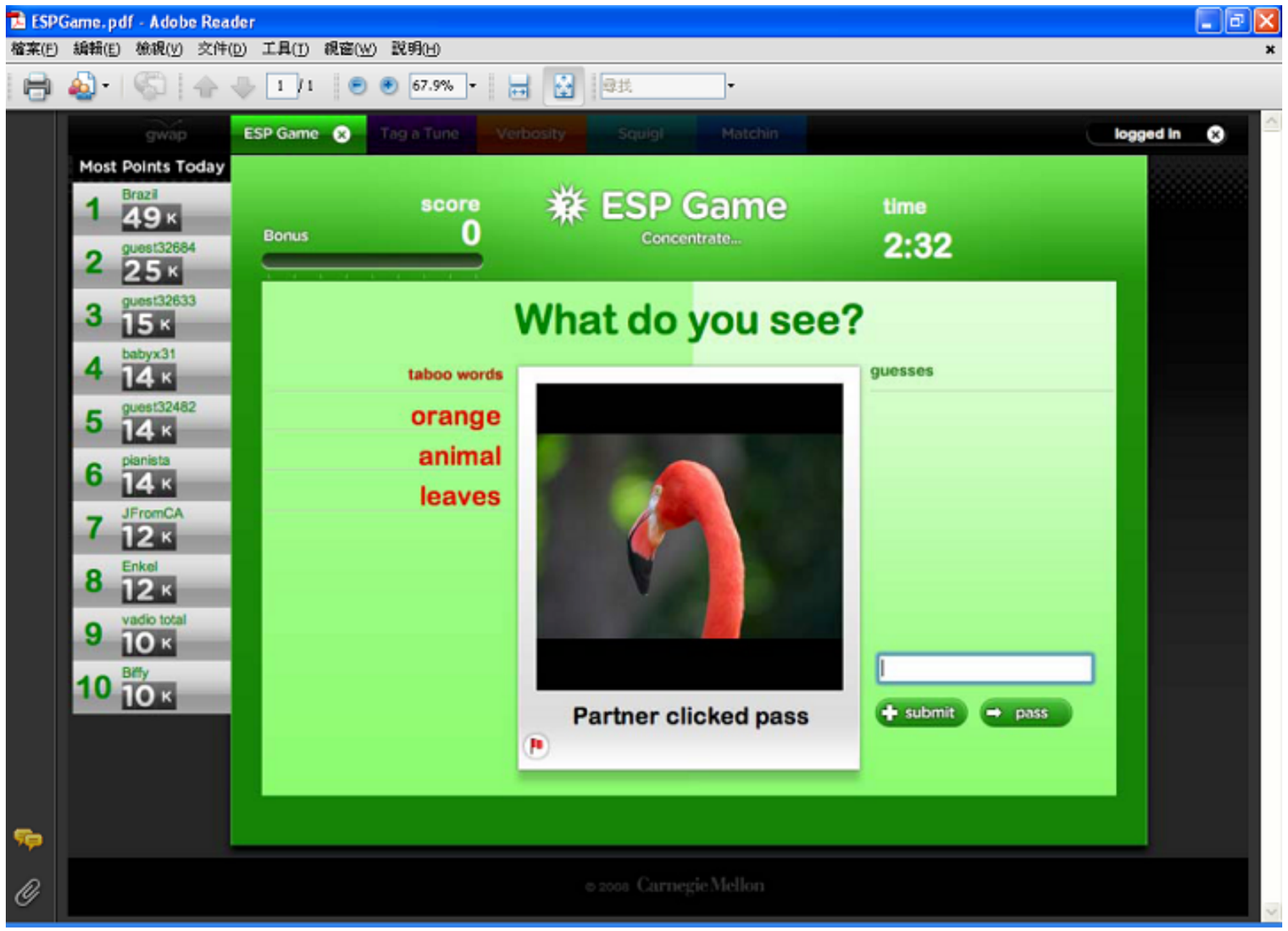
- Example (3): Mindpixel
  - Reward those Internet users who consistently validate a fact inline with the other users
  - Shortcoming: the cost is high!
- Example (4): Wildfire wally
  - To solve the maximum clique problem
  - Shortcoming: rely on unpaid volunteers to donate their time to contribute information

Social Game-based Human Computation with online players (1)

- Later, social games were proposed to collect information from the players as a side effect of their playing
- Advantage:
  - It encouraged more Internet users to contribute information to solve the AI problems because of the increasingly popularity of online game
- TWO important factors for collecting information effectively from players through a social game:
  - Guarantee the quality of collected information
  - Maintain the enjoyment of players in the game

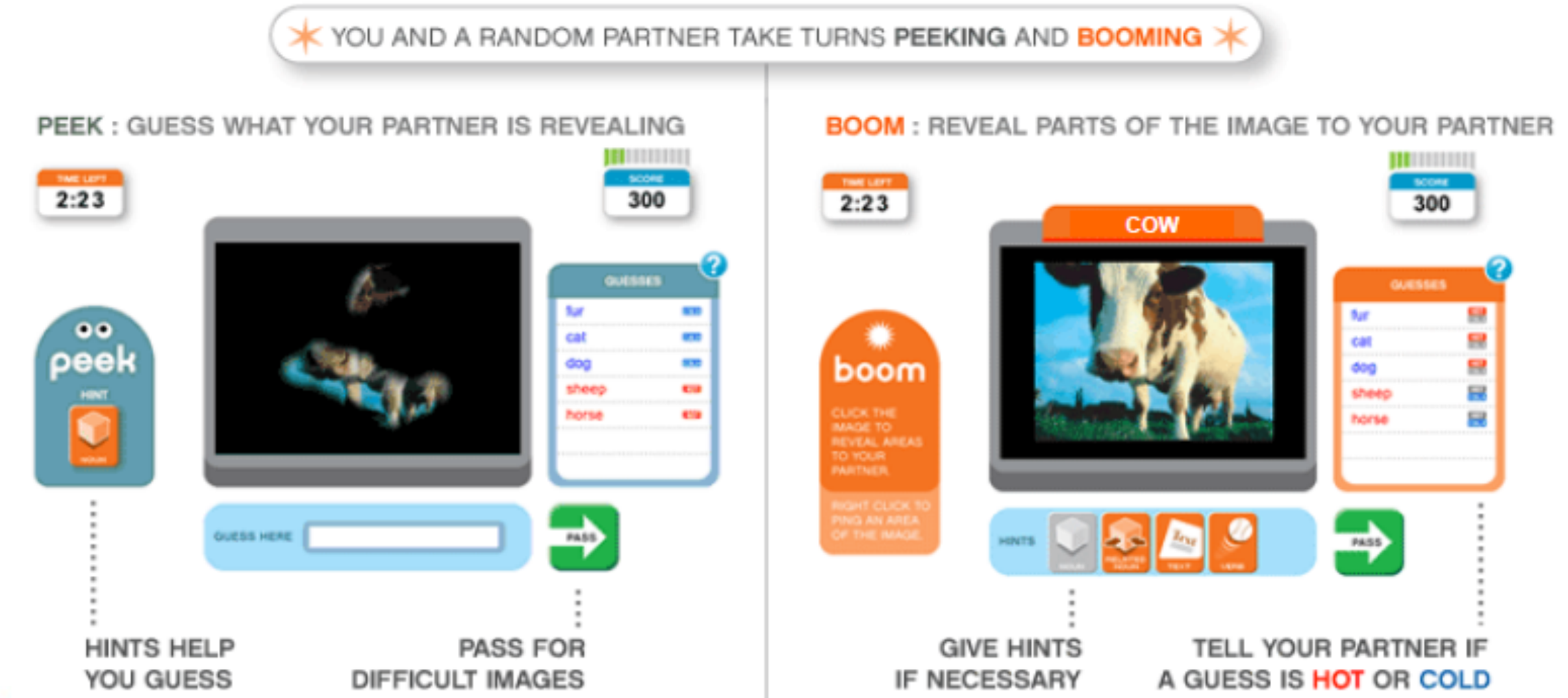
Social Game-based Human Computation with online players (2)

- To collect text information from images
  - Examples (1): ESP game



Social Game-based Human Computation with online players (3)

- To collect text information for images:
  - Examples (2): Peekaboom



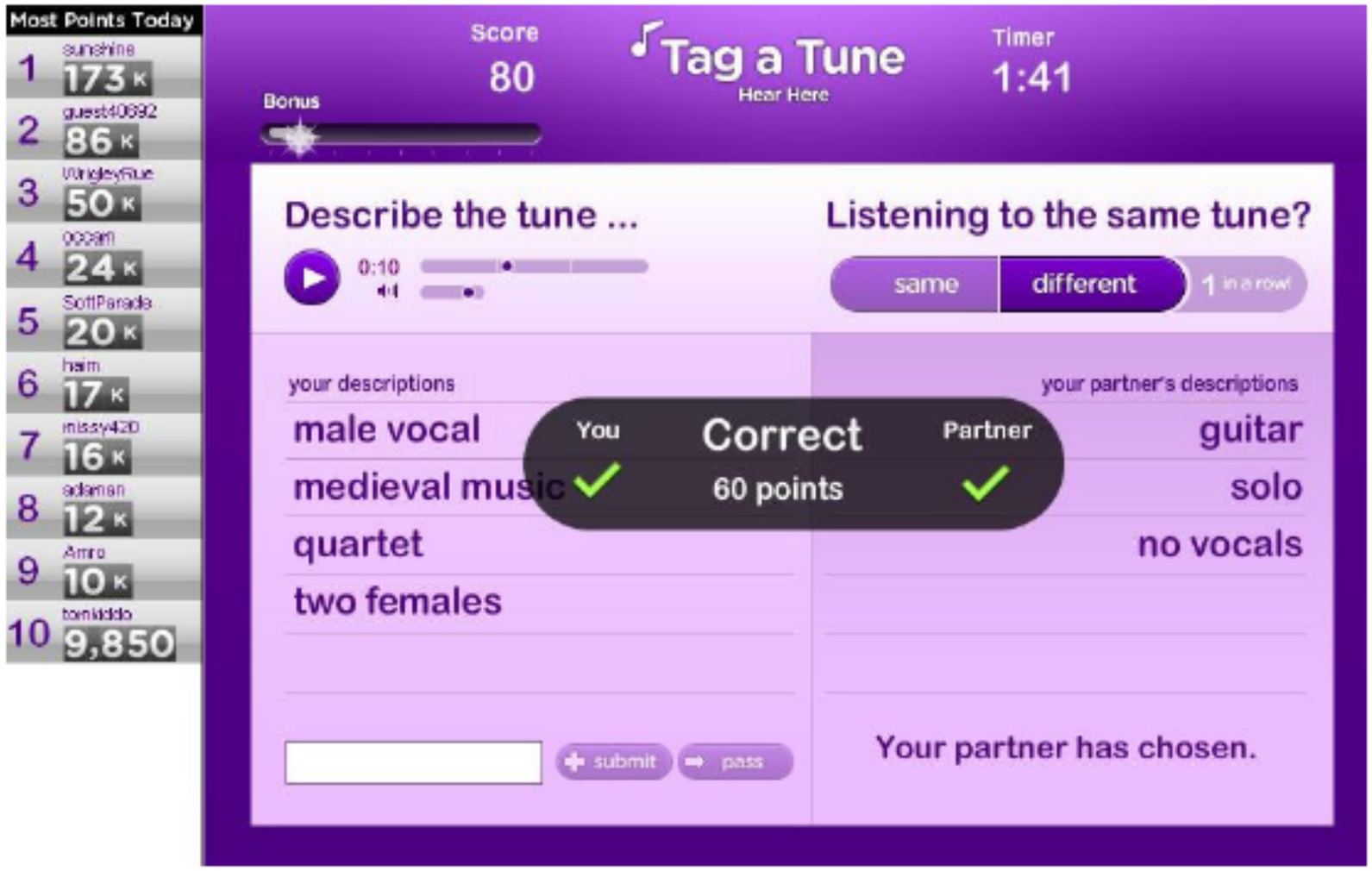
Social Game-based Human Computation with online players (4)

- To collect commonsense knowledge:
  - Examples (3): Verbosity

Figure 1. Part of the Narrator's screen.

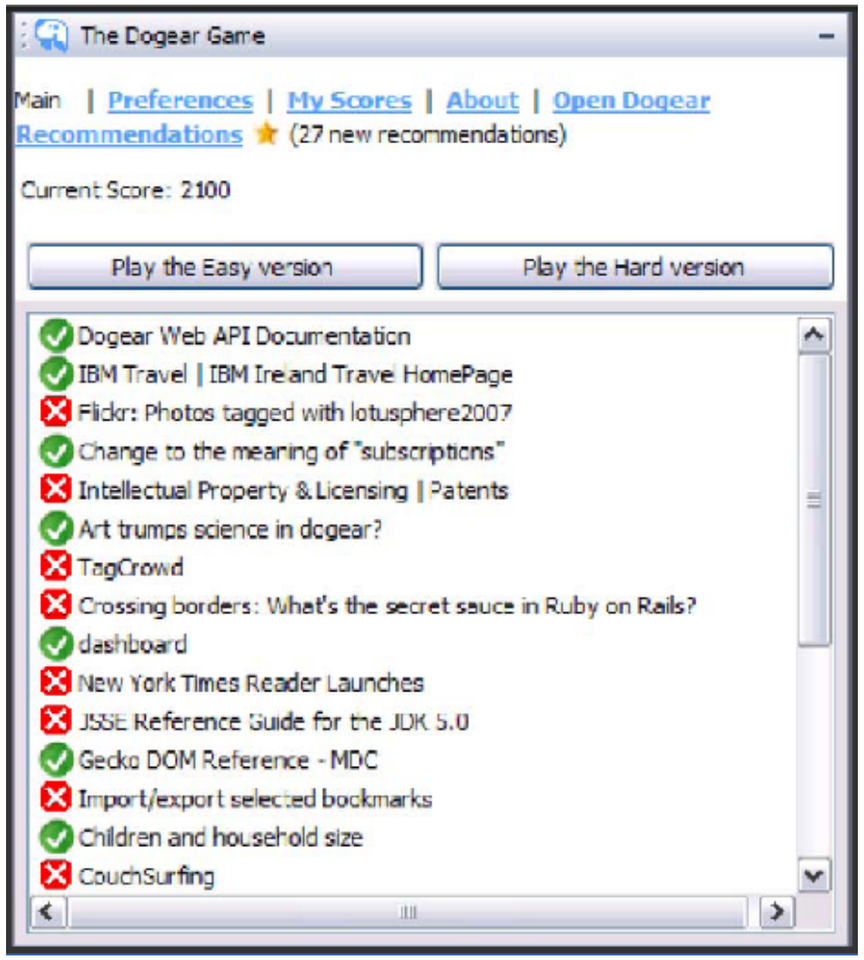
Social Game-based Human Computation with online players (5)

- To collect subjective descriptions of sounds and music:
  - Example (4): Tagatune



Social Game-based Human Computation with online players (6)

- To learn colleagues' bookmarks in an organizational goal:
  - Example (5): Dogear Game



Social Game-based Human Computation with online players (7)

- To tag locations in the real world through gameplay in mobile social games:
  - Example (6): Gopher guessing game

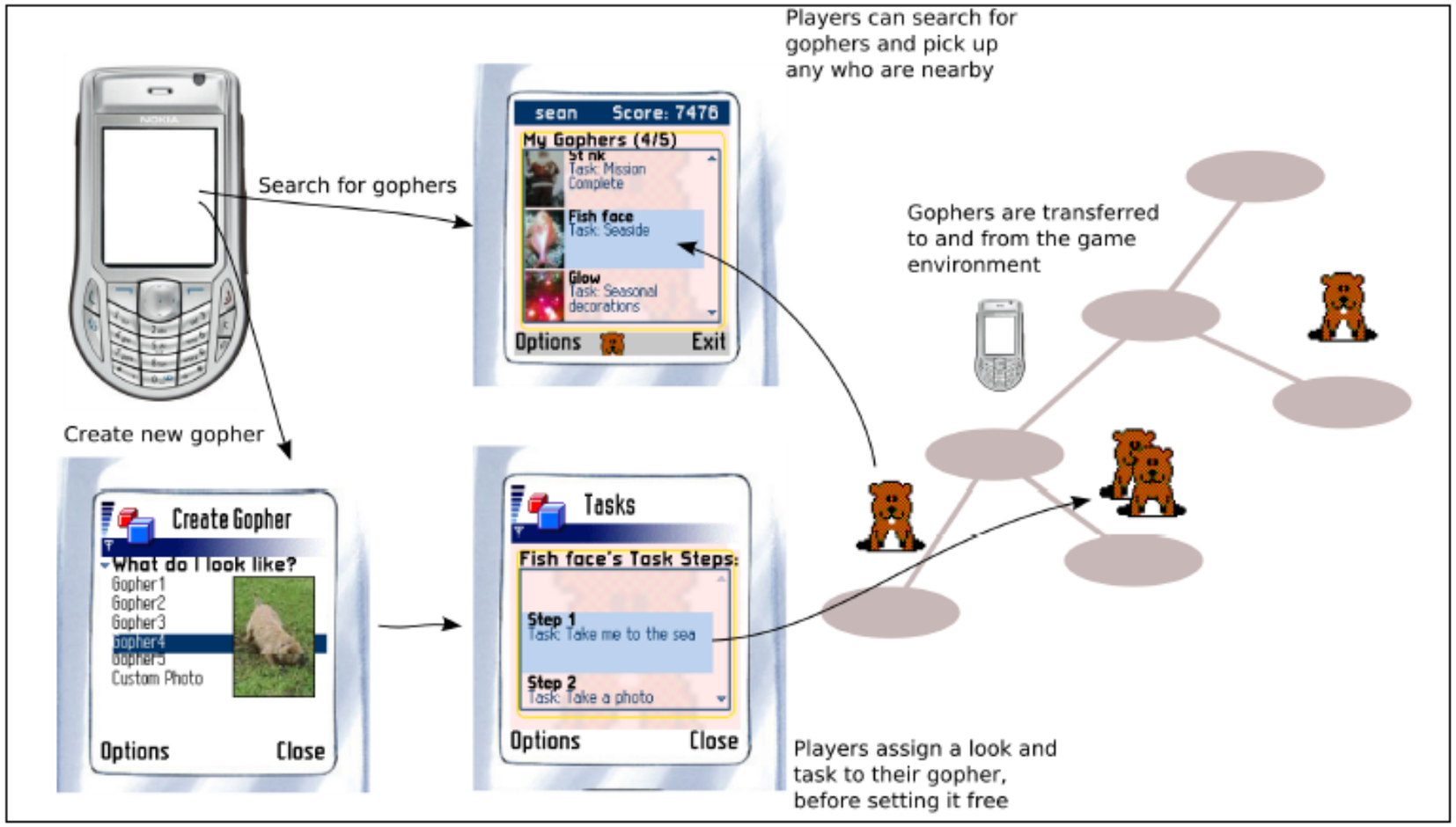


Figure 1. Real world experience, acquiring gophers

Social Game-based Human Computation with online players (8)

- To tag locations in the real world through gameplay in mobile social games:
  - Example (7): Gopher guessing game

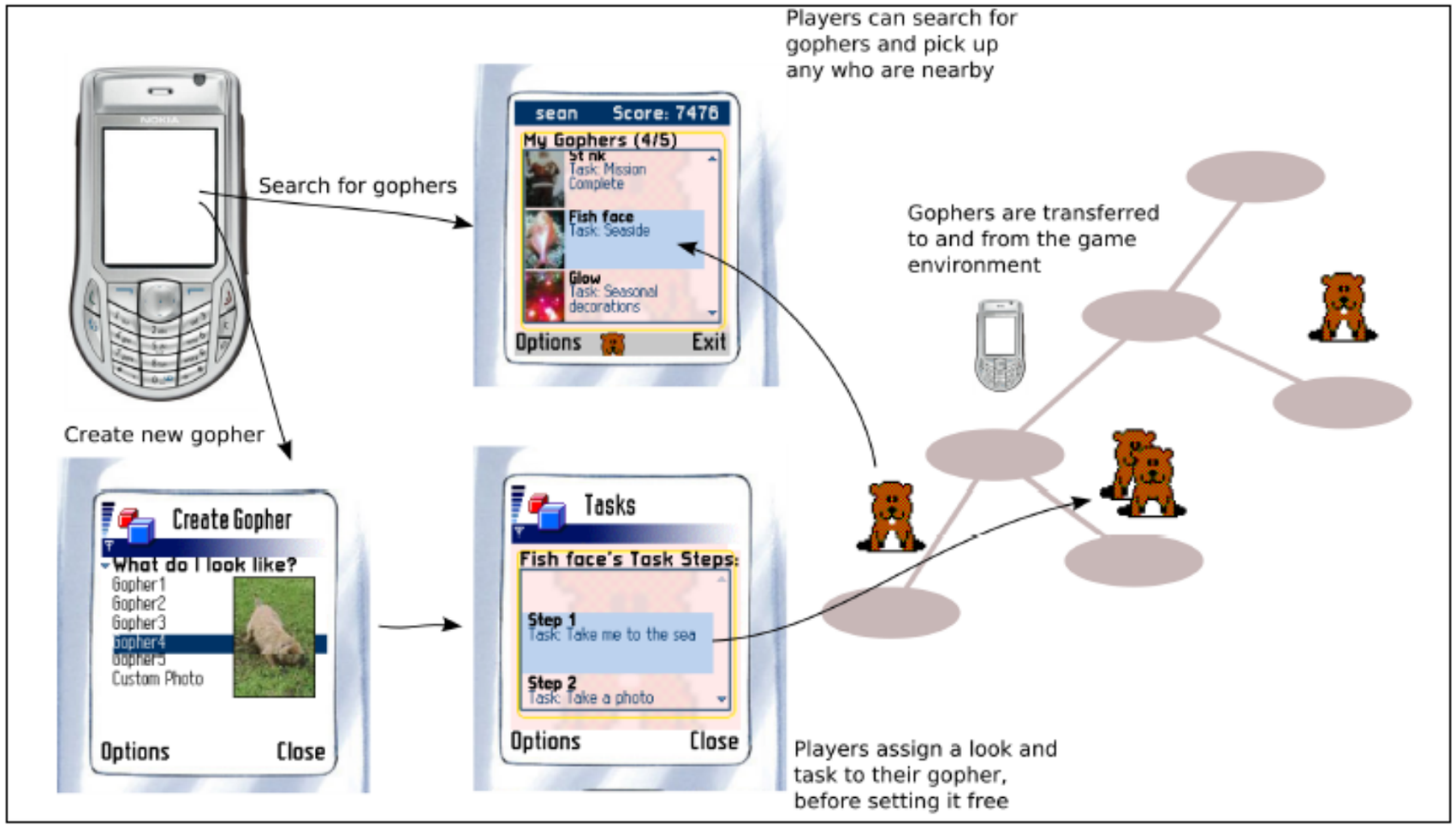
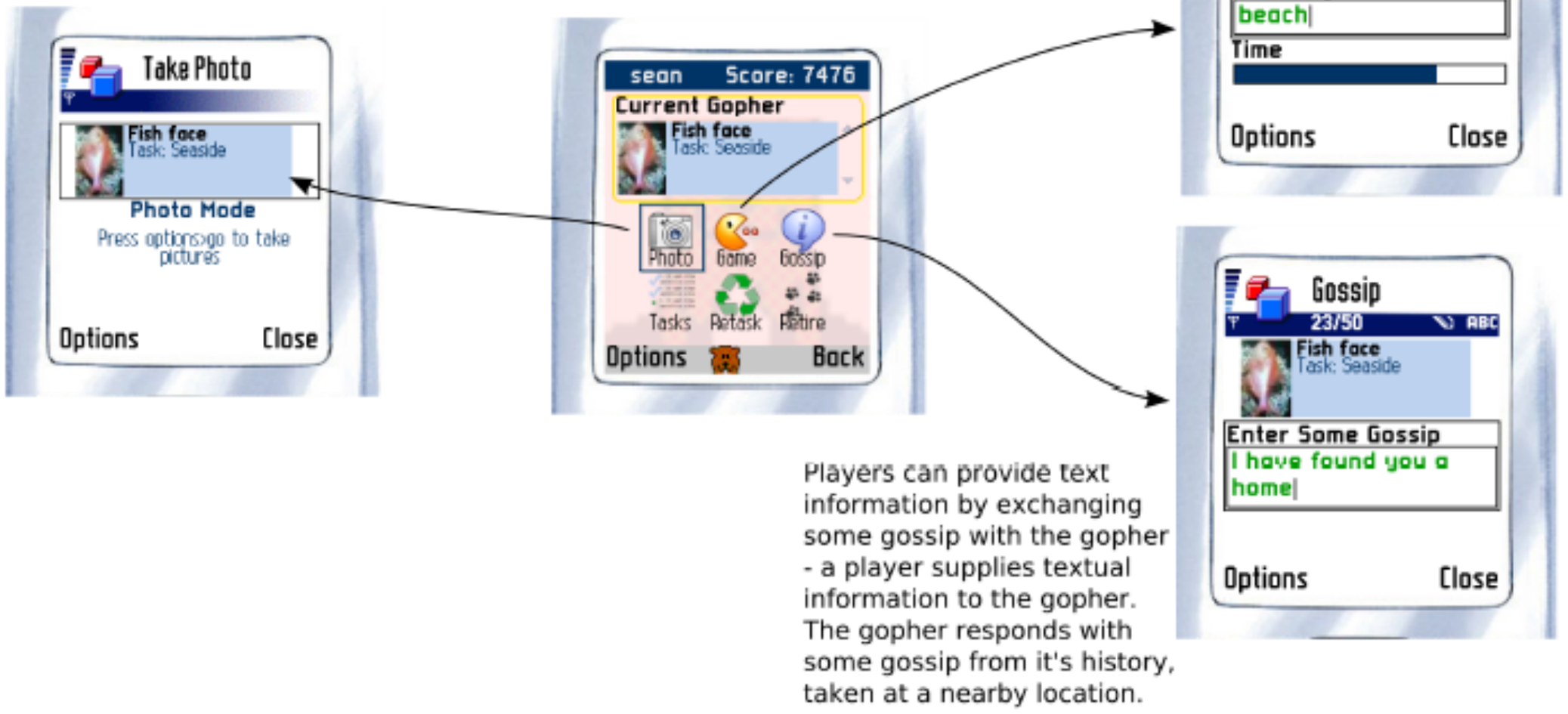


Figure 1. Real world experience, acquiring gophers

Visual feedback can be provided in the form of camera phone images - players photograph their current location and supply this to the gopher. The gopher responds with an image from its history, taken at a spatially nearby location.

Gophers can participate in a word guessing game, based on their real-world location. Players supply semantic descriptions relative to their current whereabouts. They are awarded points depending on the accuracy of their guesses.



Players can provide text information by exchanging some gossip with the gopher - a player supplies textual information to the gopher. The gopher responds with some gossip from it's history, taken at a nearby location.

Figure 2. Real world experience, interacting with gophers

Properties of Social Games

1. Type of information to be collected
2. Game Structure
    1. Output-agreement Game
    2. Input-agreement Game
    3. Inversion-problem Game
    4. Output-optimization Game
3. Verification Method
    1. Symmetric
    2. Asymmetric
4. Game Mechanism
    1. Collaborative
    2. Competitive
    3. Hybrid
5. Player Requirement

Categorization of Social Games

TABLE I
CATEGORIZATION OF SOCIAL GAMES

Game Structure	Verification Method	Game Mechanism
Output-agreement	Symmetric	Collaborative or Hybrid
Input-agreement	Symmetric	Collaborative or Hybrid
Inversion-problem	Asymmetric	Collaborative or Competitive or Hybrid
Output-optimization	Symmetric or Asymmetric	Collaborative or Competitive or Hybrid

Subjective vs. Objective Information

- For subjective information, the information presented for the same subject is affected by users because of different choices of vocabularies for the same subject.
  - lower probability on players' correct outputs being the same
- For objective information, the information presented for the same subject is NOT affected by users because of same choices of vocabularies for the same subject.
  - higher probability on players' correct outputs being the same

Game Structure (1)

- Game structure defines the key elements of a game including players' input, players' output, the relationship among the input and output of players, and the winning condition

- Four types of game structure
  1. Output-agreement Game
  2. Input-agreement Game
  3. Inversion-problem Game
  4. Output-optimization Game

Game Structure (2)

- Output-agreement Games: All players are given the same input and must produce outputs based on the common input
  - An output-agreement game should be used to collect objective information
- Input-agreement Games: All players are given inputs that are known by the game (but not by the players) to be the same or different. The players are instructed to produce outputs describing their input, so their partners are able to assess whether their inputs are the same or different. Players see only each other's outputs
  - An input-agreement game should be used to collect subjective information

Game Structure (3)

- Inversion-problem Games: The first player has access to the whole problem and gives hints to the second player to make a guess. If the second player is able to guess the secret, we assume that the hints given by the first player are correct.

- Output-optimization Games: All players are given the same input and their outputs are the hints of other players' outputs.

  - An output-optimization game should be used to collect subjective information, because the output pattern of players reflects outputs of players are strongly affected by others' outputs. It is subjective.

Verification Methods

- Verification method of a game defines the method to check the output accuracy of players by asking players to do the same task or different tasks
- Symmetric Verification Games: Either an output-agreement game or an input-agreement game is symmetric verification
- Asymmetric Verification Games: Players are assigned to one of the roles to do different tasks

Game Mechanism

- Game mechanism defines the relationship of all players in the game in order to achieve the winning condition
- Collaborative Games determine the winning condition of all players. The accuracy of output is guaranteed by collaboration of all players.
- Competitive Games determine the winning condition of a player. Output accuracy is guaranteed by information stored in a database. Players' enjoyment in the game can be increased in competition.
- Hybrid Game

Player Requirements (1)

- Player requirement defines the rules on accessing the game of all players.
- In Synchronous Games, players have to give real-time response to other players' action.
- In Asynchronous Games, players do not have to give real-time response to other players' action. The information collected from one player is stored in a database and will be used to determine the correctness of other players' output.

Player Requirements (2)

- Number of players define the following types:
- Single-player Games: It allows one player to play and the other's moves can be simulated from the prerecorded game. Only inversion-problem game can be a single-player game.
- Two-player Games: It allows two players to play together.
- Multi-player Games: It allows multiple players to play together. Only hybrid games can be a multi-player game.

Summary

TABLE II
CATEGORIZATION OF SOCIAL GAMES WITH EXAMPLES

Game Structure	Verification Method	Game Mechanism	Player Requirement		Examples
			Num of Player	Game Play	
Output-agreement	Symmetric	Collaborative	2	Synchronous	ESP, Matchi, Squigl, OntoGame
		Hybrid	Multi-players	Synchronous	Common Consensus, Social Heroes
		Hybrid	Multi-players	Asynchronous	Gopher Game
Input-agreement	Symmetric	Collaborative	2	Synchronous	TagATune
		Hybrid	N/A	N/A	N/A
Inversion-problem	Asymmetric	Collaborative	1 or 2	Synchronous	Peekaboom, Verbosity
		Competitive	2	Asynchronous	Dogear, CyPRESS, CARS
		Hybrid	1 or Multi-players	Synchronous	Phetch
Output-optimization	Symmetric	Collaborative	2	Synchronous	Restaurant Game
		Competitive	N/A	N/A	N/A
		Hybrid	Multi-players	Synchronous	Diplomacy

- Gentry et al. analyzed the security and reliability against malicious parties on distributed human computation systems.

- Su et al. performed comprehensive experiments using real datasets to study the impacts of user behavior on the quality of human-reviewed data.

- Snow et al. compared the quality of non-expert annotations and existing gold standard labels provided by expert labelers.

- Sheng et al. proposed an analysis to model the data quality using repeated labeling with a cost. They found that, with repeated labeling, it is possible to improve the data quality at low cost, especially when labels are noisy.

- A number of previous works analyzed the performance of the ESP game which was the first social game and was subsequently adopted as the Google Image Labeler.

Final Remarks

- Future Work
  - Models, theories, etc.
  - Tools, platforms, etc.
  - Performance metrics, e.g., accuracy, complexity, etc.
- To provide a better understanding about Human Computation Systems (HCS) systematically
- To facilitate future research activities in the field of HCS

Crowdsourcing

Sheng-Wei (Kuan-Ta) Chen, Institute of Information Science, Academia Sinica, Taipei, Taiwan

- Crowdsourcing = Crowd + Outsourcing
- Soliciting solutions via open calls to large-scale communities
  - INNOCENTIVE
  - oDesk
  - Amazon Mechanical Turk - Marketplace for work
  - Yahoo! Answers
  - Wikipedia

What Are Crowdsourceable?

- Software development - USD $25,000 per job
- Data entry - USD $4.4 per hour
- Image tagging - USD $0.04 per image
- General questions - points on Yahoo! Answers
- Image understanding - USD $0.01 to $0.02 per task
- Human action recognition - USD $0.01 per task
- Linguistic annotations (word similarity) - USD $0.2 per 30 word pairs

Multimedia QoE Assessment

- Quality of Experience (QoE) = User's subjective satisfaction about a service (multimedia content)
- To provide end-user experience, we measure the QoE of multimedia content, e.g, image, voice, video, etc.
  - Efficiency vs. Reliability
  - Objective evaluation approach
  - Subjection evaluation approach

Evaluation Approaches

- Objective Evaluation
  - Cannot capture all the QoE dimensions that may affect users' experiences
  - Cannot include external factors, e.g., quality of headsets, distance between the viewer and the display
- Subjective Evaluation
  - Opinions, e.g., 1=bad, 2=poor, 3=fair, 4=good, and 5=excellent
  - Difficult to define the ordinal scales concisely
  - Difficult to verify users' scoring results

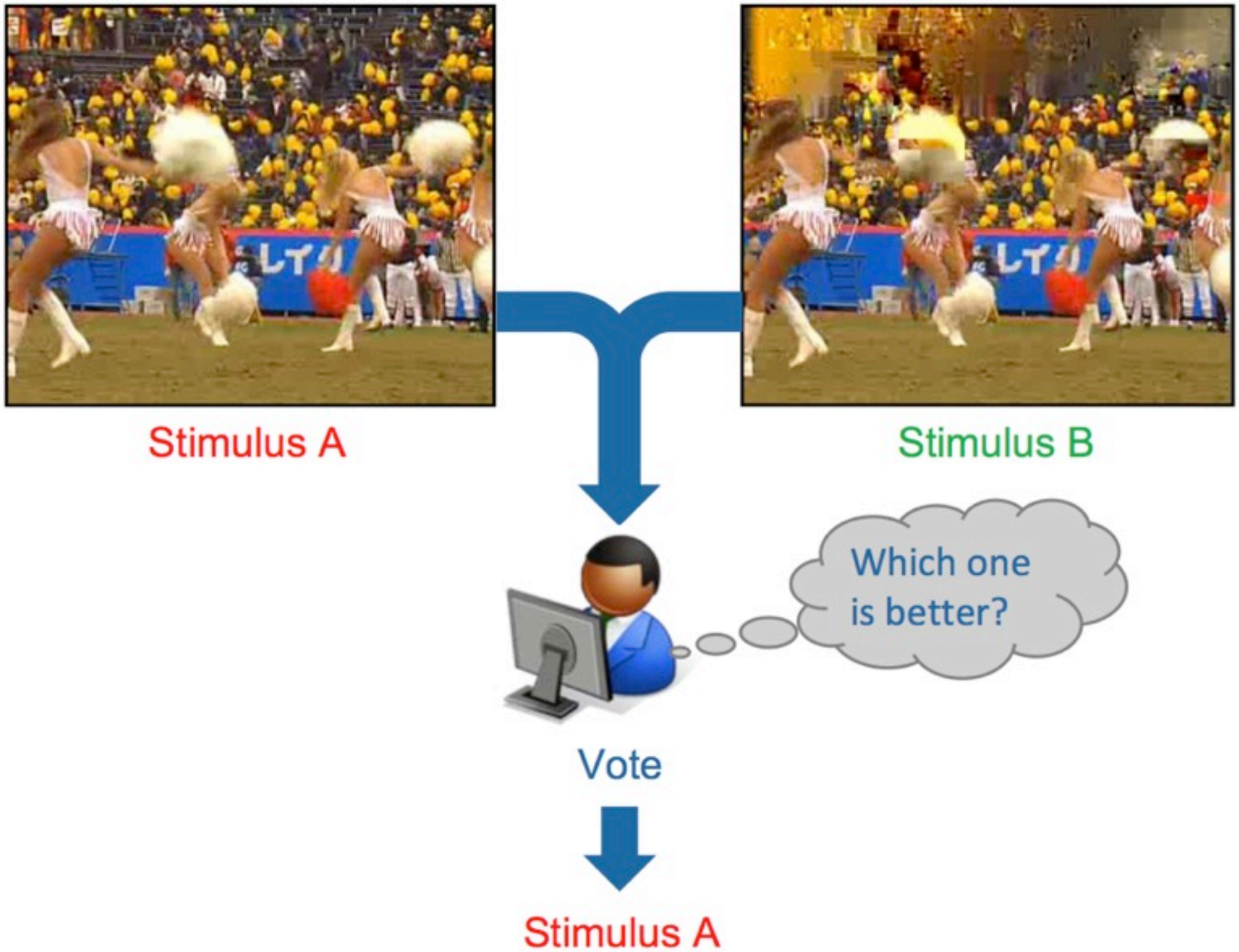
Drawbacks of Subjective Evaluation

- High economic cost
  - Participant payment
- High labor cost
  - Supervision labor
- Physical space/time requirements
  - Transportation cost
  - Laboratory space
  - Difficult to find motivated participants

Crowdsourcing Challenges

- Not every Internet user is trustworthy
  - Experiments without supervision so no quality assurance
  - Increased variance and bias
  - Need to find a way to detect problematic inputs!

Paired Comparison Test



Features of Paired Comparison

- Generalizable across a variety of multimedia applications
- Simple comparative judgement
- Interval scale QoE scores can be calculated
- Verifiable users' feedback

Verification of Users' Inputs

- Transitivity property
  - If A > B and B > C then A should be > C
- Transitivity Satisfaction Rate (TSR)

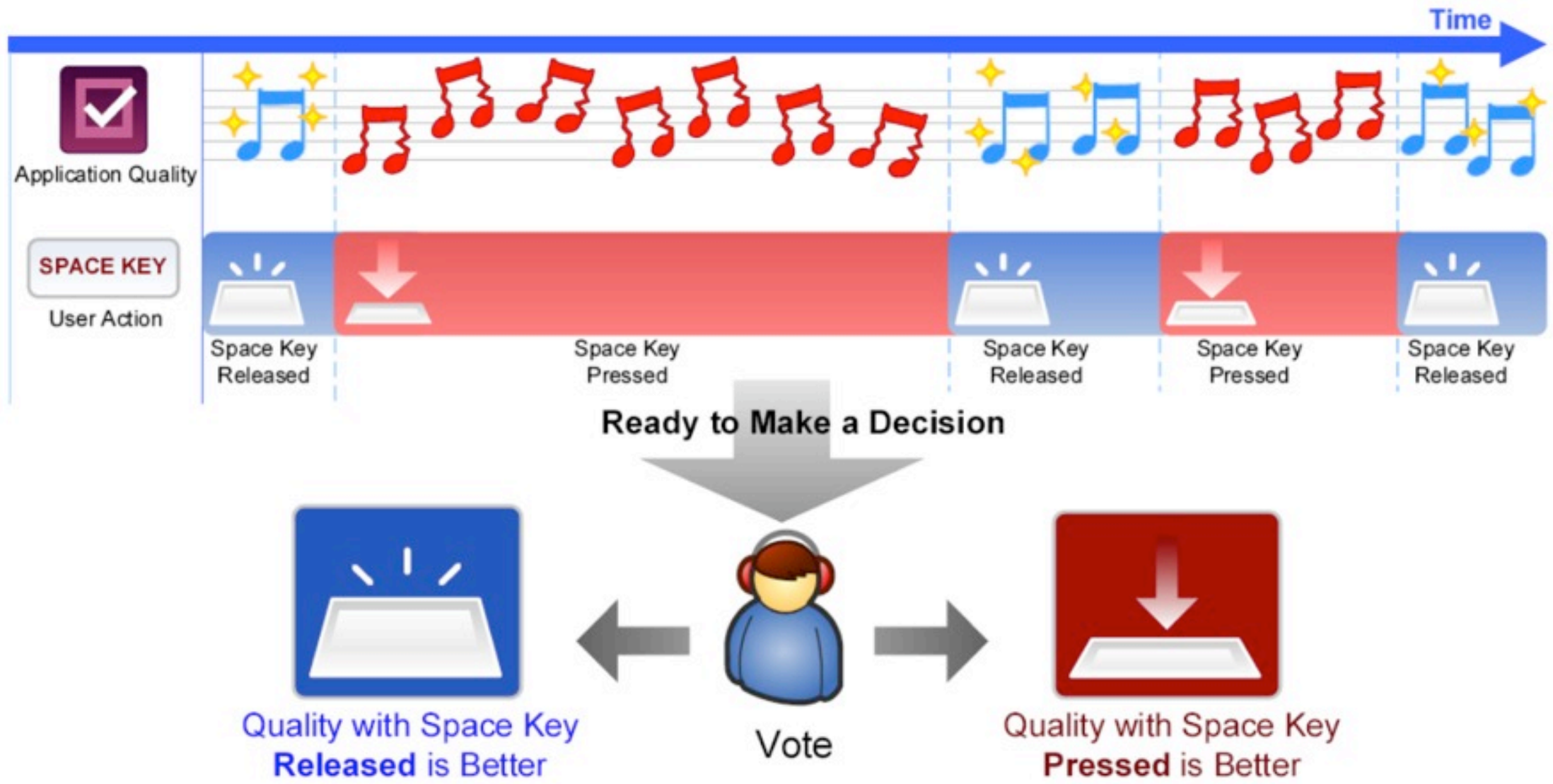
$$\frac{\#\text{ of triples satisfy the transitivity rule}}{\#\text{ of triples the transitivity rule may apply to}}$$

- Detect inconsistent judgements from problematic users
  - TSR = 1 => perfect consistency
  - TSR >= 0.8 => generally consistent
  - TSR < 0.8 => judgement are consistent

Experiment Design

- Suppose our task is to evaluate the effect of n audio processing algorithms, e.g., audio encoding
  - Select an audio clip (source clip) as the evaluation target
  - Apply the n algorithms to the source clip and generate *n* different versions of the clip (test clips)
  - Create an Adobe Flash-based system for users to evaluate the *n* test clips
  - A user need to perform 2 out of *n* paired comparison

Concept Flow of Acoustic QoE Evaluation



Which One is Better?

Participant Source

- Laboratory
  - Recruit part-time workers at an hourly rate of USD $8
- MTurk
  - Post experiments on the Mechanical Turk web site
    - Pay the participant USD $0.15 for each qualified experiment
- Community
  - Seek participants on the website of Internet community with 1.5 million members
  - Pay the participant an amount of virtual currency that was equivalent to USD $0.01 for each qualified experiment

Evaluation of the Framework

- Three participant sources
  - Laboratory
  - Amazon Mechanical Turk
  - Community
- Each with different cost structure
- Compare the cost required by each participant and the data quality produced

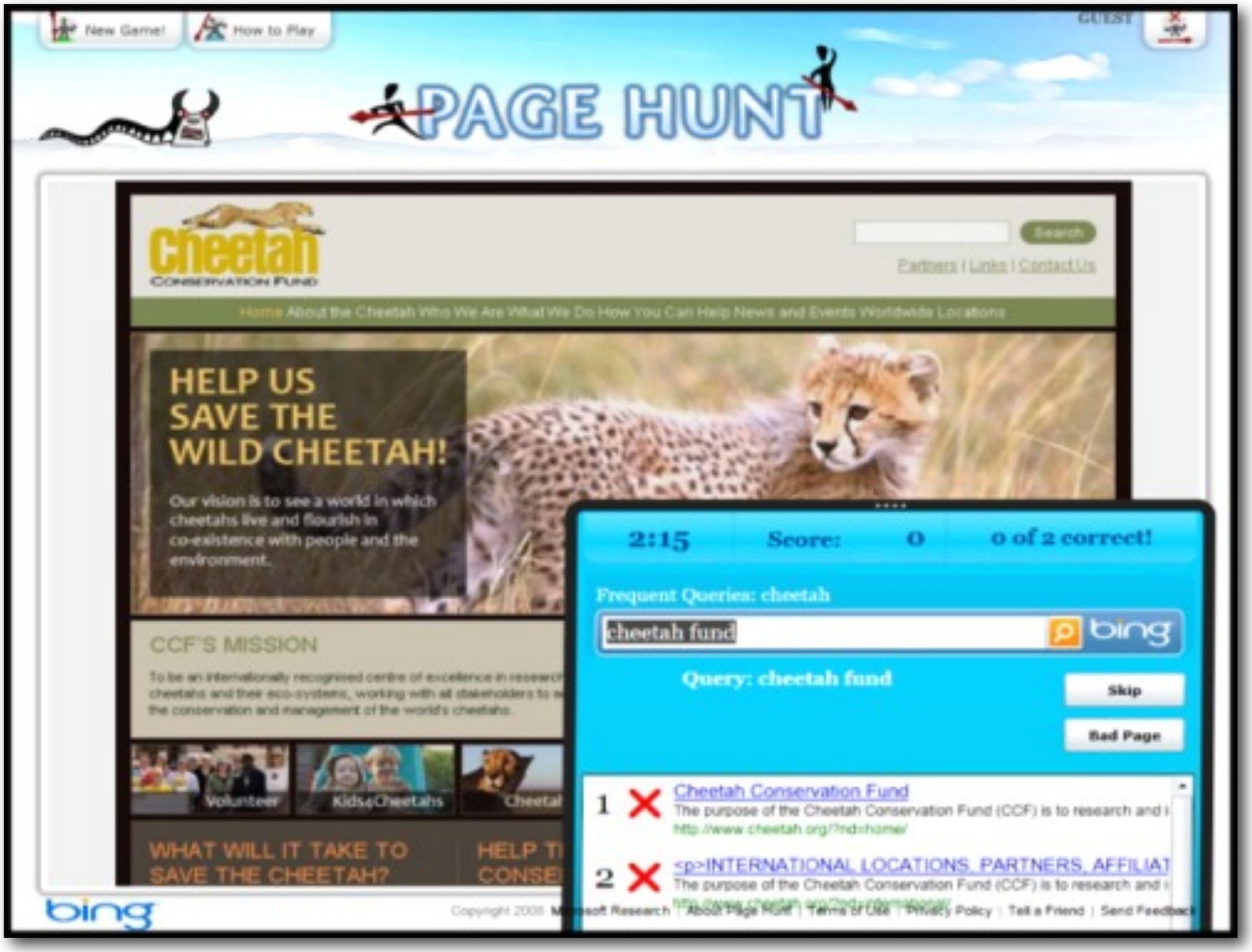
- The first **crowdsourcable** QoE evaluation framework
- Users' inputs can be **verified**
  - the transitivity property: A > B and B > C ➔ A > C
  - detect inconsistent judgements from problematic users
- Experiments can thus be outsourced to Internet crowd
  - **lower monetary cost**
  - **wider participant diversity**
  - **maintaining the evaluation results' quality**

Case Study	Experimenter Source	Total Cost (dollar)	# Rounds	# Person	Qualified Rate	Cost / Round (cent)	Time / Round (sec)	Avg. TSR
MP3 Bit Rate	Laboratory	50.97	1440	10	67%	3.54	16	0.96
	MTurk	7.50	750	24	47%	1.00	9	0.96
	Community	1.03	1,470	93	54%	0.07	25	0.96

Chen et al, "A Crowdsourceable QoE Evaluation Framework for Multimedia Content," *Proceedings of ACM Multimedia 2009*.

http://mmnet.iis.sinica.edu.tw/link/qoe

Improving Search Engines Using Human Computation Games [Hao Ma, et al., SIGIR'09, CIKM'09]



Improving Search

- Human Relevance judgments and implicit measures of relevance predicated on web pages being surfaced
- But what if some pages don't get surfaced?
- Page Hunt: a human computation game that's like search in reverse: going from web pages to queries

Motivation and Response

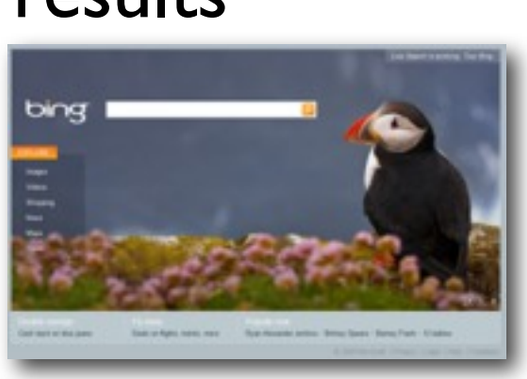
- Our Motivation
  - No games for tagging Web pages
    - No single player human computation games
      - People like competition, not just cooperation!
- Our response:
  - Page Hunt (one-player game)
  - Page Race (two-player competitive game)
  - Page Match (two-player cooperative game)

Implementation

- Implemented in C#
- Silverlight makes it zippy and fun
- Two player games created in collaboration with Max Chickering & Anton Mityagin (Microsoft) and Edith Law (CMU)

Rules

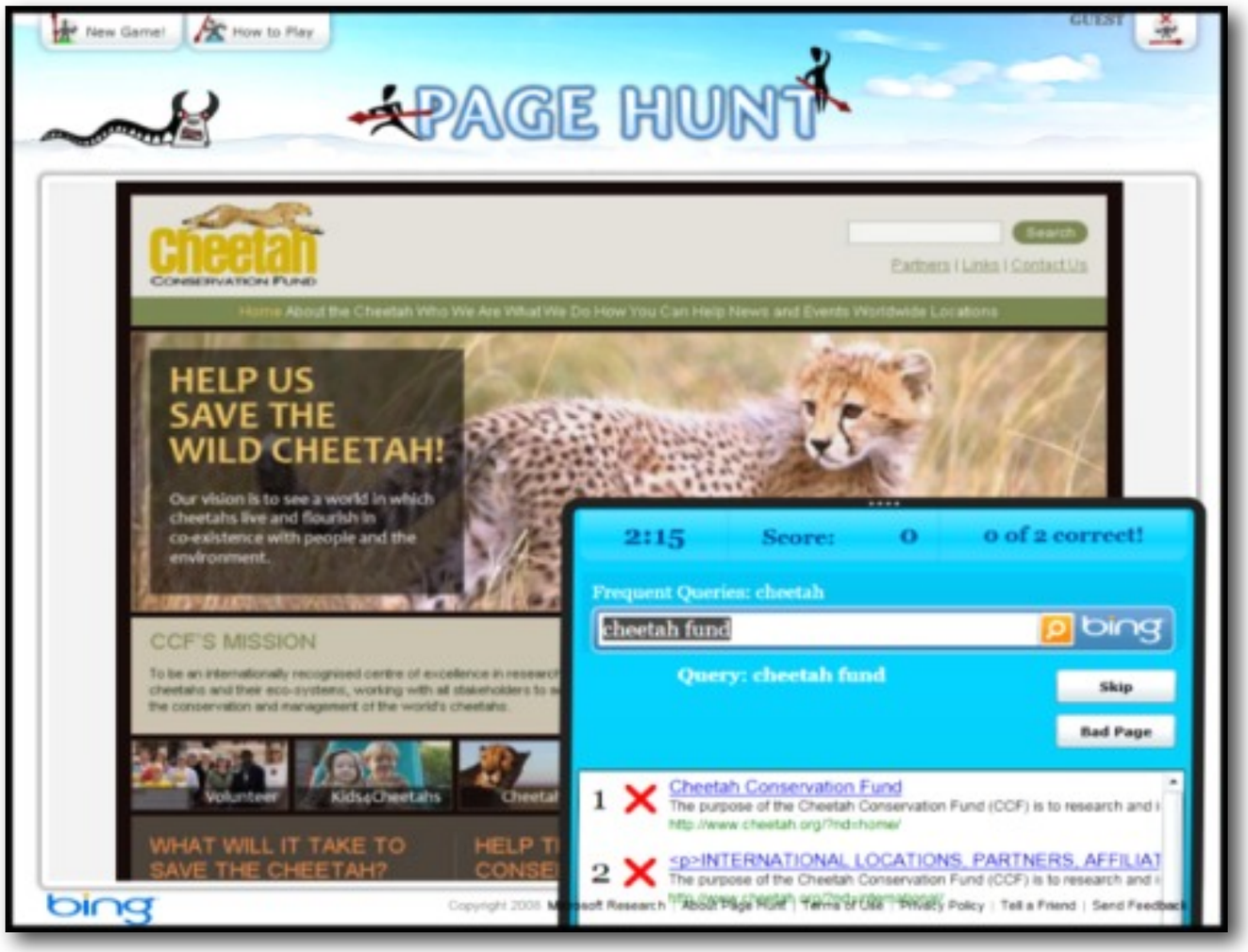
- Rules
  - Show a Web page
  - Player enters a set of terms about the page (the query)
  - Query is then executed on Bing, and the top 5 results shown (Bing is your implicit partner)



- Goal
  - Get the web page in the top 5 results for this query
- Taboo words (30% of the time); extra points for avoiding these Random 2x and 3x bonuses

Page Hunt on the Web

- Game on the web at http://PageHunt.msrlivelabs.com



Research Questions

- Can this be made a fun game?
  - Prerequisite to getting people to play the game and provide data
- How does Page Hunt data compare with other data?
  - Is the data realistic? Skewed in any way?
- How can we extract useful information from this data?
  - What can we do with this data

Making Page Hunt Fun

- Page Hunt uses
  - Timed Responses
  - Score keeping/Leader Board
  - Frequent Queries
    - Related to Taboo Queries
  - Randomness
    - Random pages
    - Random score boosts
- ‘Measuring’ Fun: Repeat games/Engagement

Some Statistics

- Data from a few days of play in July 2009
  - #players: 10,227
  - #Query-URL rows of data
    - Raw: 155,298
    - Non-null query: 123,734
    - #Wins: 45,637
  - #URLs being labeled: 698
  - Avg queries/url: 177
  - Avg wins/url: 65
  - Avg wins/player: 4.5

Nature of Queries Elicited (1)

- Issue: Are users queries realistic?
  - Or are they more like known-item search?

- Users have page in front of them
  - Cut and paste not allowed, but users may:
    - Paste long phrases from the page
    - Type in discriminative, unusual words/phrases

Nature of Queries Elicited (2)

- Small random subset classified into 3 categories:
  - Over-specified queries
    e.g. [start here medline plus hives] for
    http://www.nlm.nih.gov/medlineplus/hives.html
  - Under-specified queries
    - e.g. [hives]
  - Just-Right Queries
    - e.g. [medline plus urticaria]
    - [medline plus hives]

Nature of Queries Elicited (3)

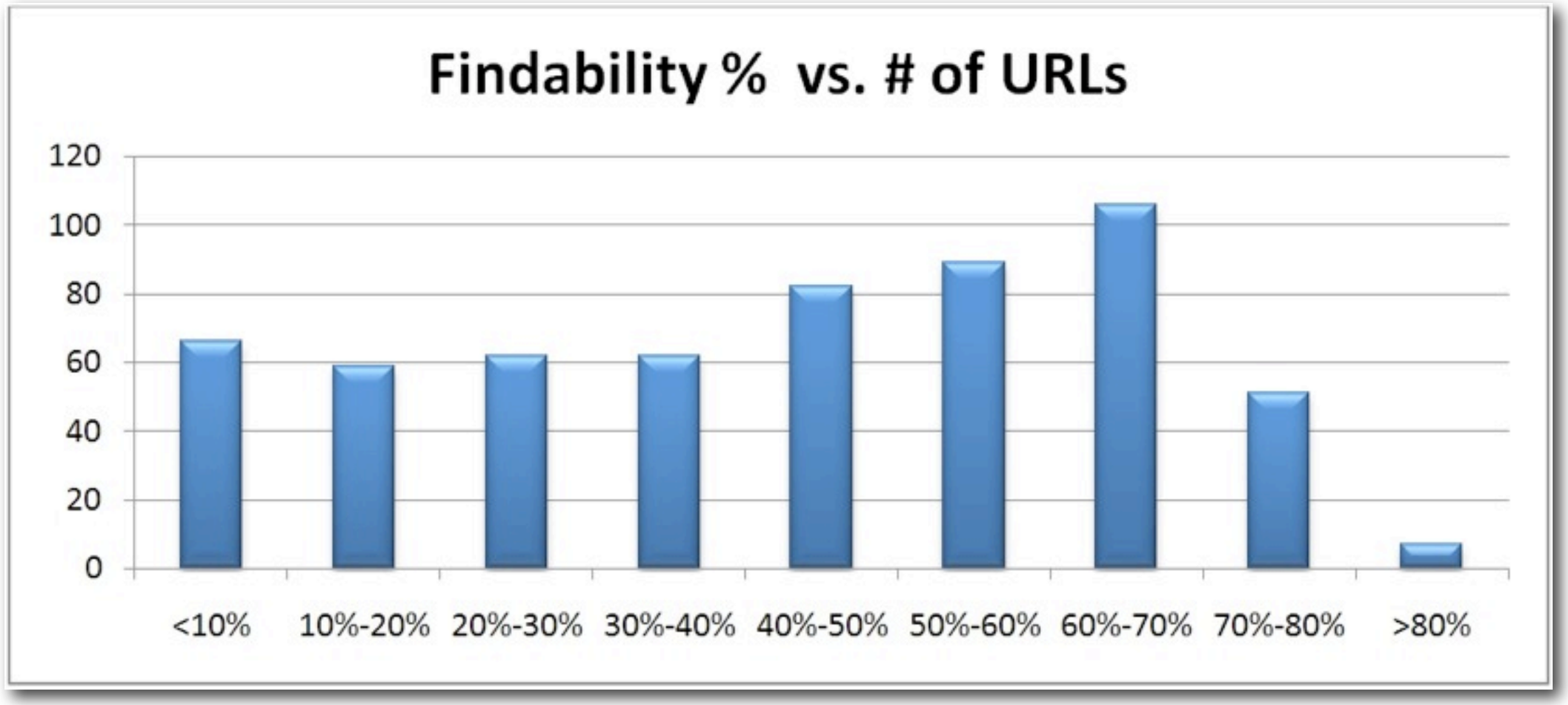
- Results:
  - Over-specified queries 15%
  - Under-specified queries 7%
  - Just-Right Queries 78%
- Time pressure seem to inhibit people from looking for highly discriminating/minimal queries.
- Many different correct queries possible for each URL; noise can be eliminated by thresholding.

How This Data Can Help Search Engines

- This game can gets us data to
  - Learn about query refinement
  - Diagnose problems in ranking
  - Obtain additional anchor text
  - Induce grammars from queries to trigger instant answers
    - about celebrities, weather etc.
  - Evaluate human relevance data
  - Compare search engines

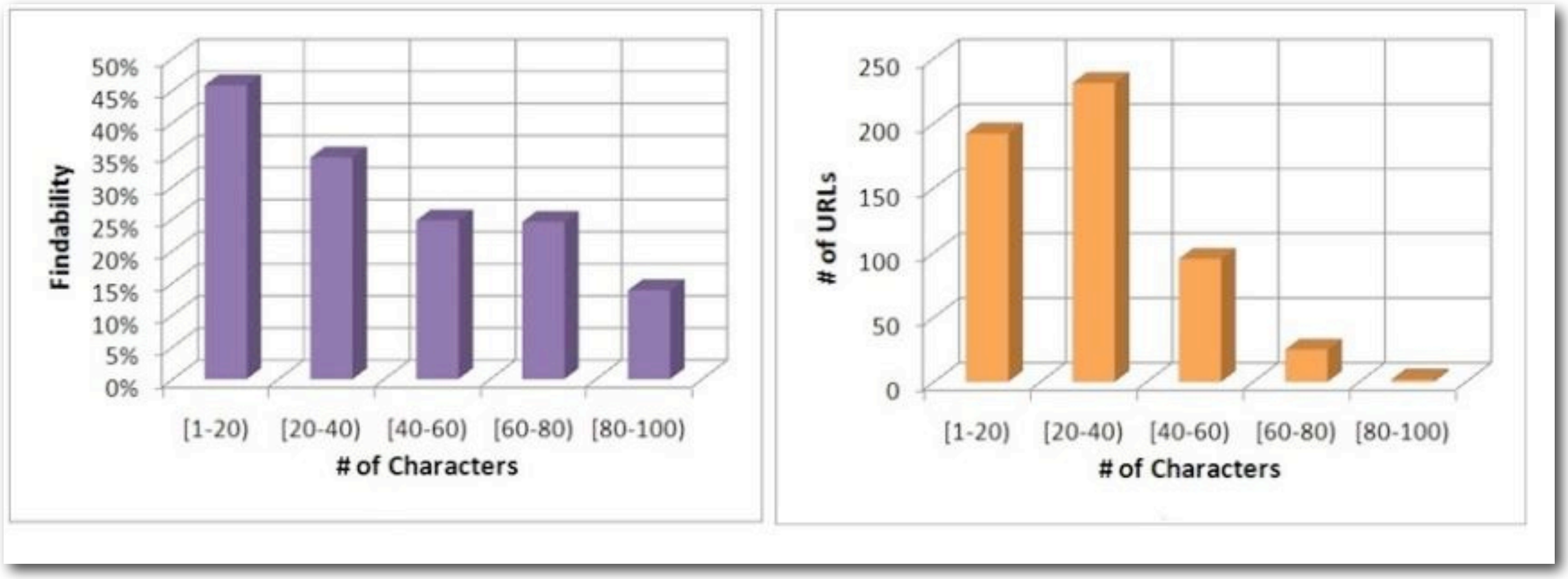
URL Findability Levels

#URLs 'hunted' down



About 10% of URLs @ < 10% findability,
About 11% of URLs @ > 70% findability

URL Length vs. Accuracy



Because of variants? Low page/static rank?

Learning Query Alterations

- Yes, we can! We can learn query alterations using all queries submitted for a given URL
- E.g., for http://www.labor.state.ny.us, we get:

new york state dept of labor
new york department labor
new york state department of labor
NY Department Labor
nysdol
new york state department of labor
New york state department of labor
new york state department of labor
ny labor

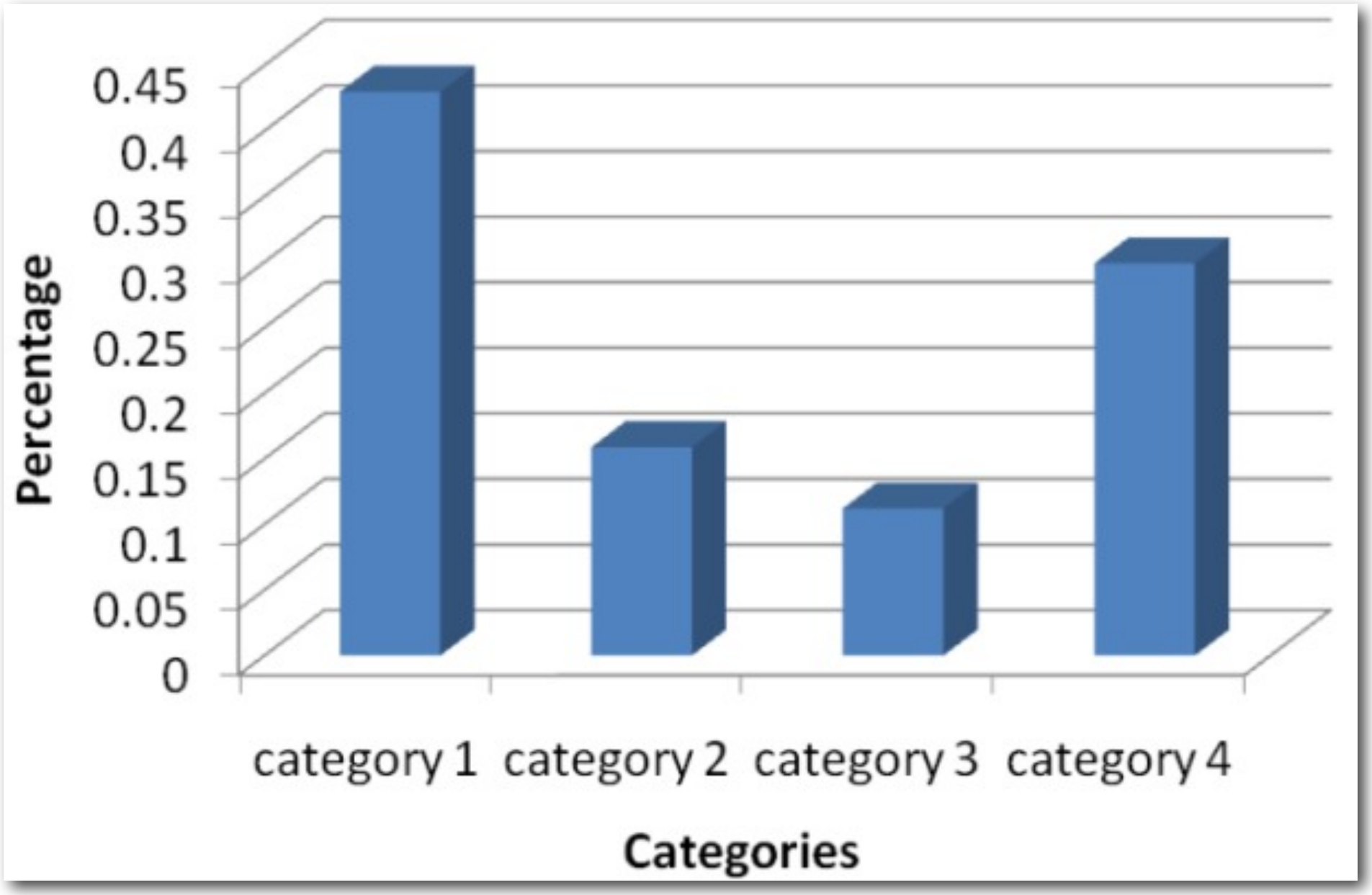
Alterations obtained

- Four classes of potential alterations got:
  - Punctuation/Spelling/Morphology variations
  - Site – name variations
  - Acronym – expansions
  - Conceptual equivalents

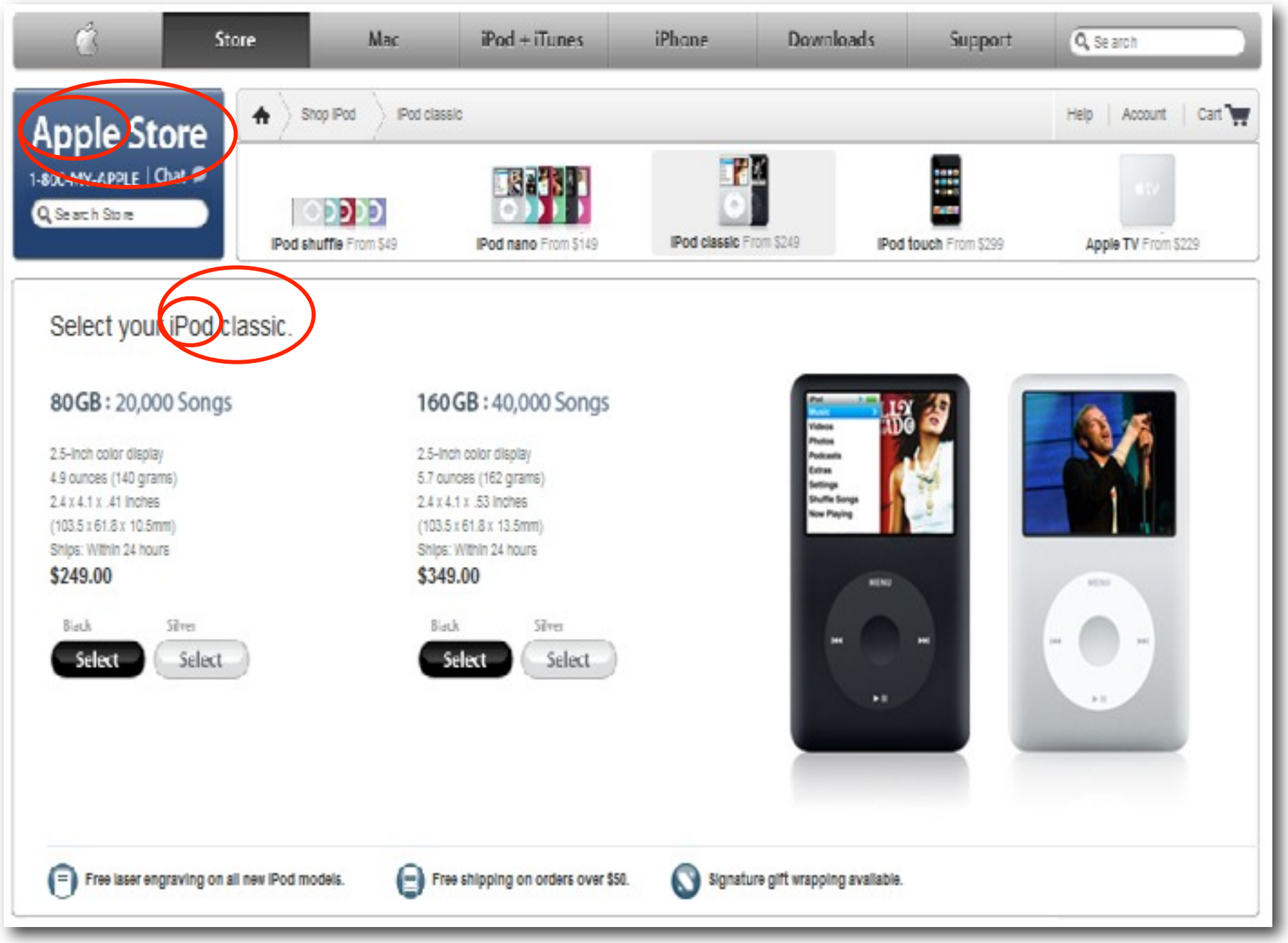
Query Alteration Examples

Original Query	Replacement	Category	Probability
jc penny	jc penney	Spelling	0.323
T mobile	T-mobile	Spelling	0.557
acid reflux	acidrefluxconnection.com	Sitename to site	0.26
Zune	Zune.com	Sitename to site	0.43
wayn	where are you now	Abbr./Initialism	0.72
iht	international herald tribune	Initialism	0.82
Jlo	Jennifer lopez	Conceptual	0.78
jay chou	Zhou Jielun	Conceptual	0.85
Capital city airport	Kentucky airport	Conceptual	0.57

Distribution of Categories



Eye-tracking/Page-Element Weighting



1. Apple
2. Apple Store
3. Apple ipod
4. Apple ipod classic

Summary

- Crowdsourcing provides a new paradigm and a new platform for scientific research
- New applications, new methodologies, and new businesses are emerging with the aid of crowdsourcing


I think you'll be
delightfully surprised
by the quality of my work
on this assignment.
I crowdsourced it.

Q & A