



# Haiti Earthquake Photo Tagging: Lessons on Crowdsourcing In- Depth Image Classifications

Zhi Zhai<sup>1</sup>, Tracy Kijewski-Correa<sup>2</sup>,  
David Hachen<sup>3</sup>, Greg Madey<sup>1</sup>

<sup>1</sup>Department of Computer Science and Engineering

<sup>2</sup>Department of Civil and Environmental Engineering

<sup>3</sup>Department of Sociology

University of Notre Dame

United States



Macau, China  
08/23/2012



# The Convenience of Modern Technologies



# Two Trends

- » “Cognitive Surplus”:

Advances of Modern Technologies

I. More Free Time

II. A Large Proportion Not Used Meaningfully

Watching TV, Talking on the Phone, Online Chatting, etc.

- » The Development of Information Technology.

- A better means to organize a large number of virtually connected people to effectively work together.

# Citizen Engineering Experiments/Prototypes

# Example 1 (Scientific Portal)

## - Galaxy Zoo

ENGLISH | POLSKI

# GALAXY ZOO.org

Welcome | Home | The Science | How to Take Part | Galaxy Analysis | Forum | Press | Blog | FAQ | Links | Contact Us

Login Register

**Dear Galaxy Zoo users,**

Thanks for making Galaxy Zoo such a success! THIS SITE IS NOW ARCHIVED; go [here](#) for the latest Galaxy Zoo project.

With your help, we collected millions of classifications, and have done more and better science faster than we ever believed possible. The first papers and follow-up observations are complete, and you can follow our progress on the [BLOG](#) and [FORUM](#).

This site is still alive for nostalgia's sake, but your classifications will not form part of the public data release. **But we need you now more than ever!** Galaxy Zoo 2 - which asks for more detailed classifications of roughly 250,000 of the brightest galaxies in our sample is now live. So go [here](#) to get classifying.

**Log In**

User Name:

Password:

Remember me next time.

[Register](#)  
[Forgot Password](#)

# Example 3 (Crowd Byproduct)

## - reCaptcha

The diagram illustrates the process of correcting OCR errors using a crowd. It shows a scanned line of text: "This aged portion of society were distinguished from". Below this, the OCR output is shown: "niis aged pntkm at society were distinguished frow." Red arrows point from the original text to the OCR output, highlighting the errors. A red arrow labeled "Scanned type" points to the original text, and another red arrow labeled "OCR reads as" points to the OCR output. The errors in the OCR output are underlined: "niis", "pntkm", "at", and "frow."

The screenshot shows a reCAPTCHA interface. At the top, there is a snippet of text: "steamboat train, from New this morning ran off the track. New-London. Four cars plunged". The word "morning" is highlighted with a red box. Below this, a red arrow points to a reCAPTCHA box. Inside the box, two words are shown in a distorted font: "morning" and "UPON". Below the words, there is a text input field with the prompt "Type the two words:". To the right of the input field are three buttons: a refresh button, a back button, and a reCAPTCHA logo with the text "stop spam. read books."

Our ongoing research:

Haiti Earthquake

Photo Tagging



## Haiti

Haiti: A country in the Caribbean Sea

Population: 9,719,932    Area: 10,714 mi<sup>2</sup>

2010 Haiti Earthquake:

Time: Jan. 12, 2010    Magnitude: 7.0

Casualties: 316,000 Deaths



# Sample Photo 1



# Sample Photo 2



# Project Sketches

- » Background: 2010 Haiti Earthquake
- » Thousands of Photos Taken on Site
- » 242 Citizen Engineers Participated
  - College Students as Surrogates
- » 9318 Photo Classifications on
  - 400 Sample Photos Over 17 days

# Concerns & Considerations

- » Goal:

  - Achieve highly trustworthy classifications from a large number of user inputs

- » Challenges:

  - Diversified user backgrounds

  - Potential malicious users

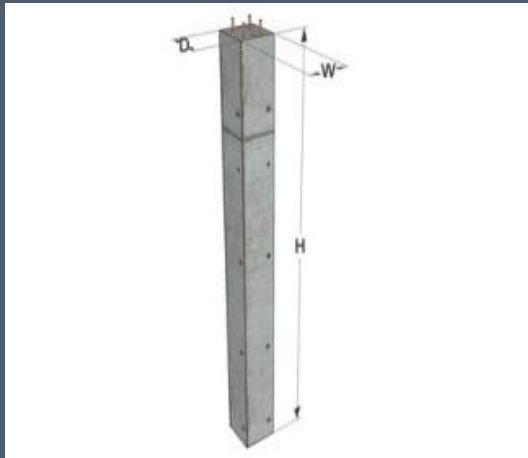
  - Large data size

# Civil Engineering

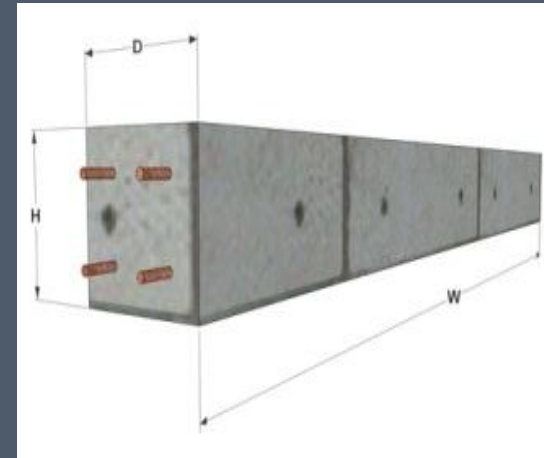
## Knowledge For Users

(Images are from the online tutorials,  
designed by structure engineering  
professors involved in the project)

# Damaged Parts



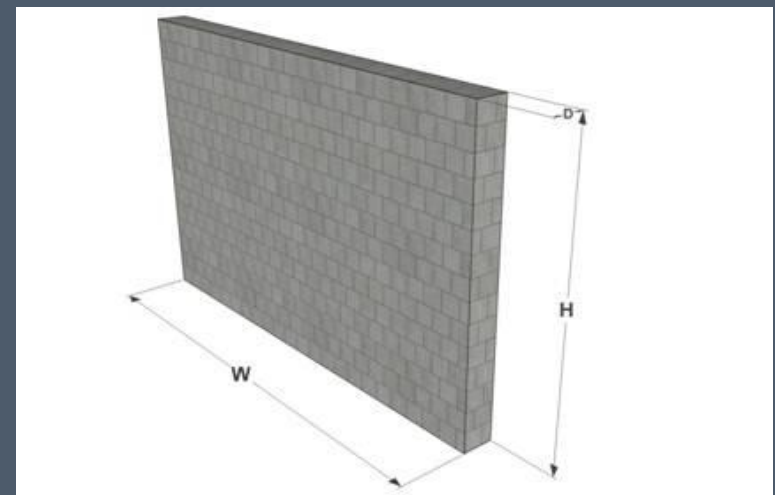
Column



Beam



Slab



Wall

# Damage Type - Column



# Damage Type - Beam





# Damage Type - Slab



# Damage Type - Wall



User Tasks:

Photo Tagging

# Sample Question

## Classify Photos

Number of Photos You Completed [?](#)

You've tagged 2 photos (1% of our database)

Your Standing in the Crowd [?](#)

You are in the 24th percentile (of all participants)



Is there damage in (any of) [the beam\(s\)](#)?

- Yes
- No

Is there damage in (any of) [the column\(s\)](#)?

- Yes
- No

Is there damage in (any of) [the slab\(s\)](#)?

- Yes
- No

Is there damage in (any of) [the wall\(s\)](#)?

- Yes
- No

[BACK](#)

[NEXT](#)

Clicking on links (underlined phrases in blue) will take you to the relevant page in the tutorial.

# Question Designs

Which **building** needs to be tagged?

- Tag entire photo
- Tag part of photo
- Cannot determine

BACK NEXT



What **primary elements** are visible and can be assessed in this photo?  
(Check all that apply)

- Beam
- Column
- Slab
- Wall
- Cannot determine

BACK NEXT



Is there damage in (any of) **the beam(s)**?

- Yes
- No

Is there damage in (any of) **the column(s)**?

- Yes
- No

Is there damage in (any of) **the slab(s)**?

- Yes
- No

Is there damage in (any of) **the wall(s)**?

- Yes
- No

BACK NEXT



What **damage patterns** are visible in (any of) the beam(s)?  
(Check all that apply)

- Flexure
- Shear
- Concrete Loss

What **damage patterns** are visible in (any of) the column(s)?  
(Check all that apply)

- Flexure
- Shear
- Concrete Loss

What **damage patterns** are visible in (any of) the slab(s)?  
(Check all that apply)

- Flexure
- Shear
- Concrete Loss

What **damage patterns** are visible in (any of) the wall(s)?  
(Check all that apply)

- Sliding Shear
- Diagonal Shear
- Out-of-Plane



What is the **severity** of the worst damage pattern in (any of) the beam(s)?

- Yellow
- Red

What is the **severity** of the worst damage pattern in (any of) the column(s)?

- Yellow
- Red

What is the **severity** of the worst damage pattern in (any of) the slab(s)?

- Yellow
- Red

What is the **severity** of the worst damage pattern in (any of) the wall(s)?

- Yellow
- Red

BACK NEXT

# Post Experiment Data Analysis

» Various Data Cleaning Approaches

We focused on two of them:

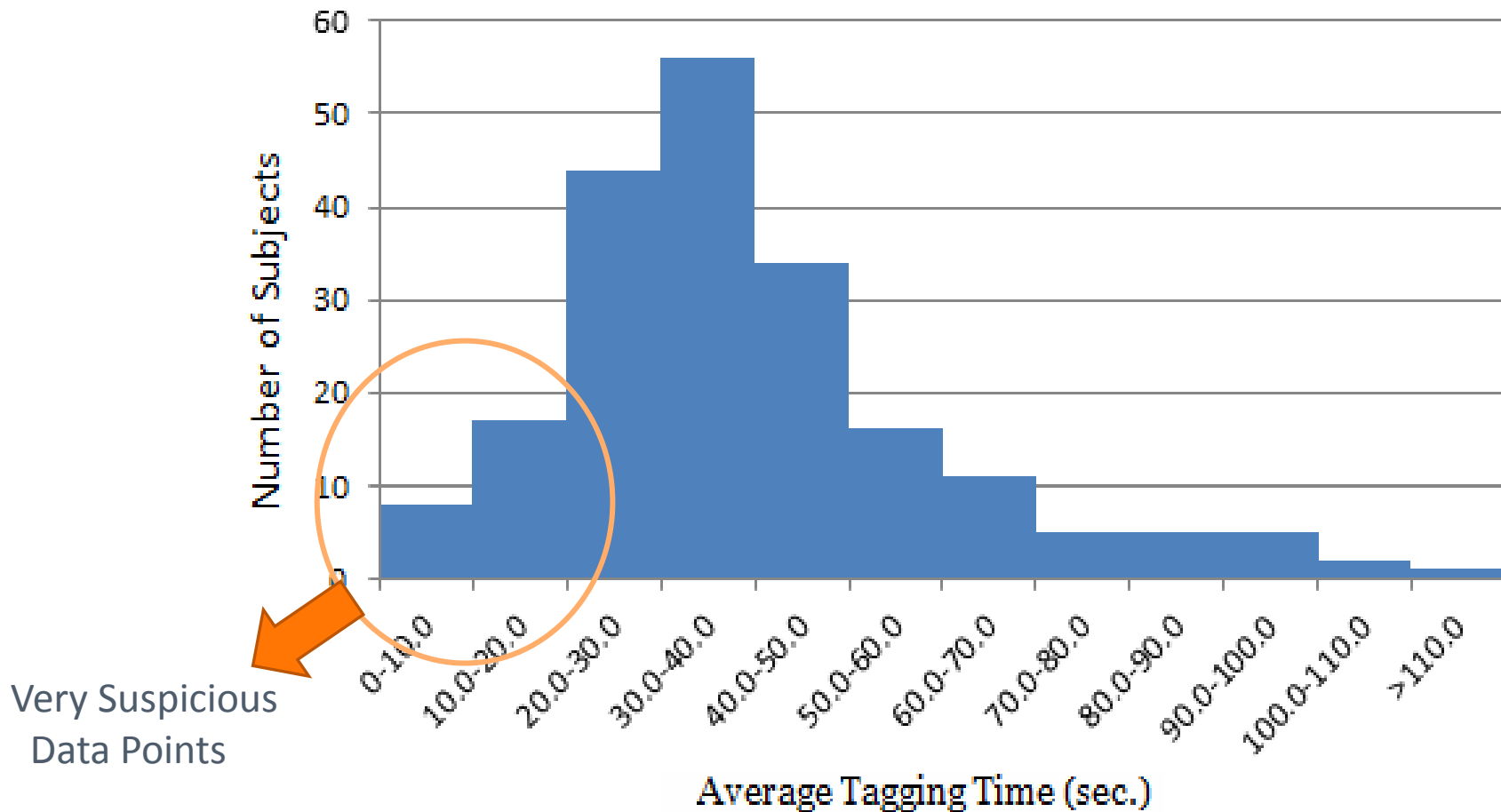
- I. Free Loader Trimming
- II. Long Invalid Sequence

# Data Cleaning:

## Method I – Trim Freeloaders

Average Time: 35-40 secs./photo

Freeloaders: less than 10 secs./photo



# Ground Truth

- » 3 PhD graduate students in civil engineering provided expert judgments on the 400 sample photos.

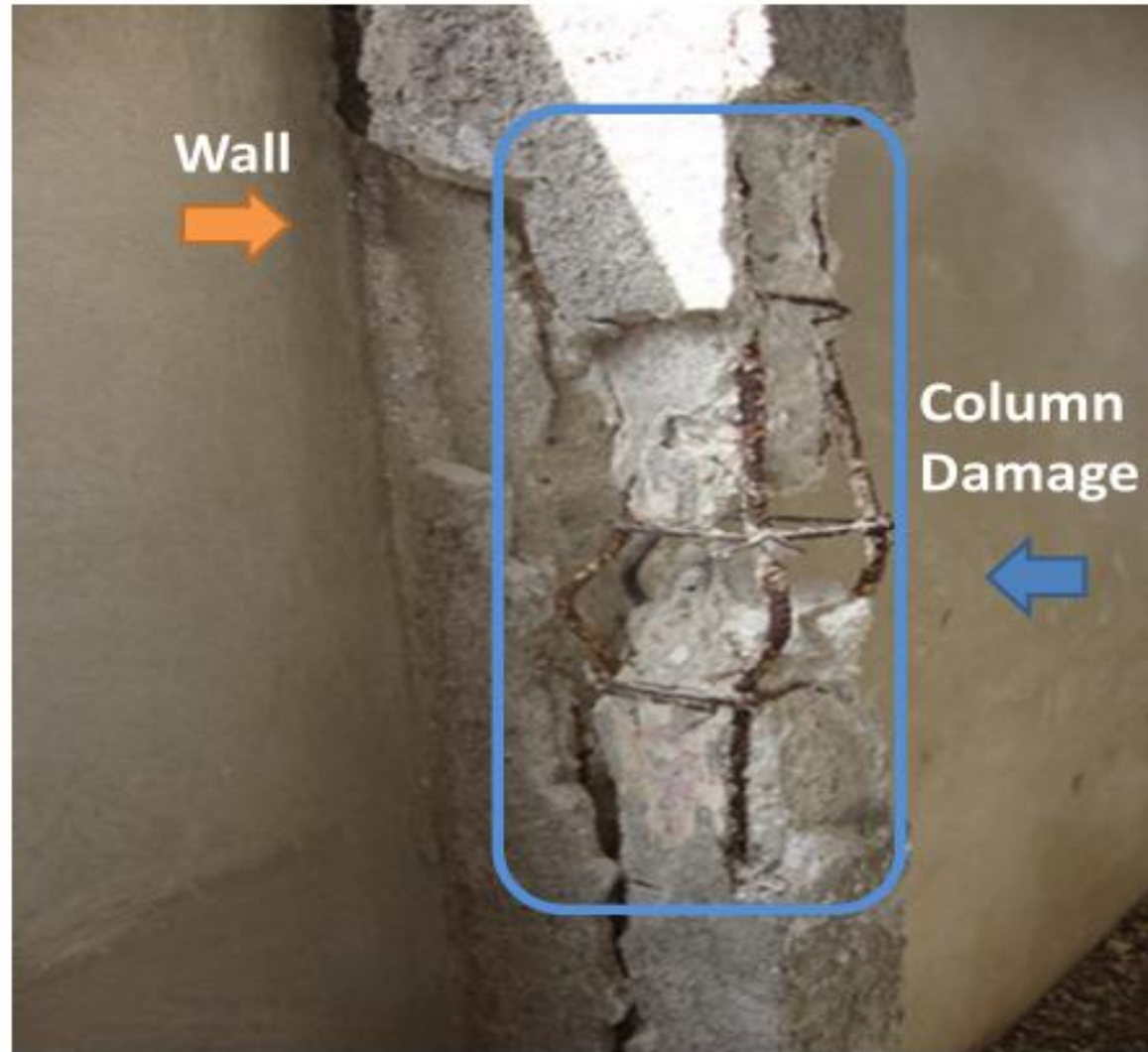




# Classification Results:

- » Accuracy significantly increased after data pruning.
- » Clear shot photos: the crowd generate highly trustworthy results
- » Difficult ones (ambiguous objects or multiple tagging targets): less stratifying results

# Clear Shot Photos: High-Quality Results



# Highly Composite Photos - Still Challenging



# Future Work (1):

## Prioritized System Designs

- » Which aspects have the highest priority?



- » Teamwork: currently, most current platforms, including our photo tagging system, do not support teamwork.

How can we team up users to achieve higher productivity? >

# Future Work (2): Stratified Citizen Engineering Systems

- » A whole spectrum of citizen engineers with variable backgrounds and expertise.
- » Our preliminary research focuses on low-end average citizen engineers and high-end expert engineers.
- » How about the ones in between?

How to foster junior engineers to be able to fulfill complicated tasks ?

# Thank you all!





## Acknowledgements:

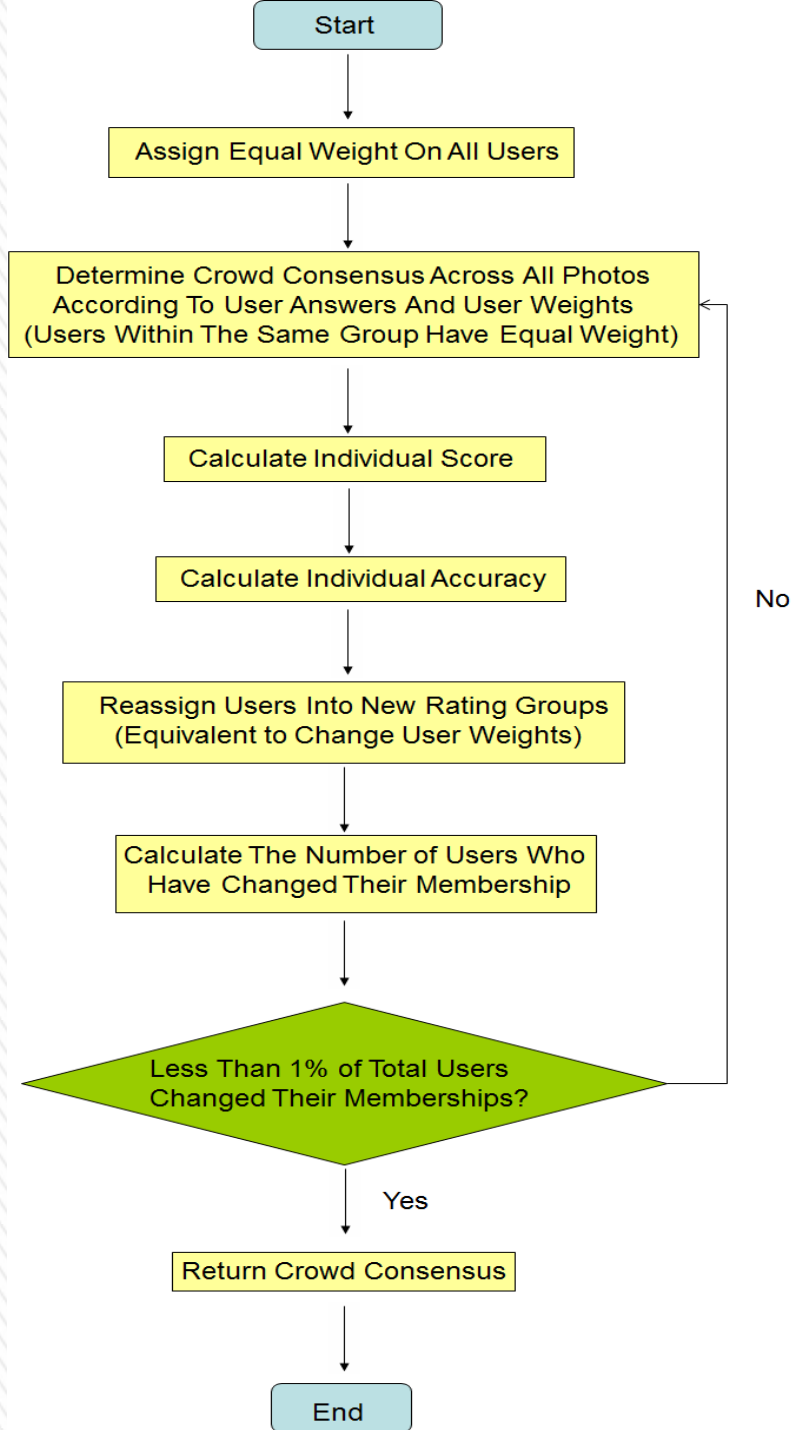
The research presented was supported in part by an award from the National Science Foundation, under Grant No. CBET-0941565.

We would also like to thank Drs. Jenny Vaydich and Zack Kertcher for valuable contributions.

# Backup Slides



# Workflow



# Example 2 (Commercial Platform)

## - Amazon Mechanical Turk

amazonmechanicalturk  
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.

**239,529 HITs** available. [View them now.](#)

### Make Money by working on HITs

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

As a Mechanical Turk Worker you:

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



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

### Get Results from Mechanical Turk Workers

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

As a Mechanical Turk Requester you:

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



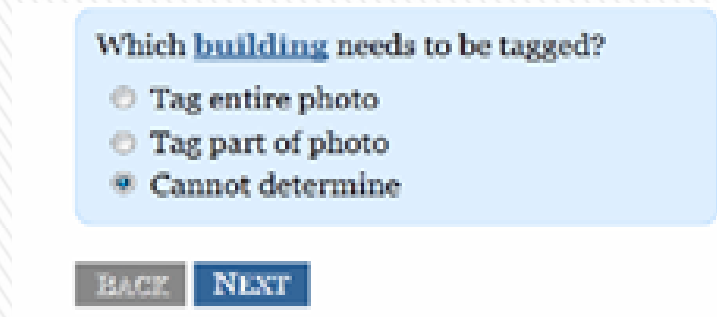
# Example 2 (Online Gaming)

## - GWAP (Gaming With A Purpose)

The screenshot shows the GWAP website interface. At the top, there is a navigation bar with tabs for 'ESP Game', 'Tag a Tune', 'Verbosity', 'Squigl', 'Matchin', 'Flipit', and 'PopVideo'. Below this is a login area with a 'Sign In' button, a 'remember me' checkbox, and a 'forgot password?' link. The main content area features a large banner with the text 'Play the Games, Change the Web.' and 'When you play a game at Gwap, you aren't just having fun.' To the right, there is a spotlight on the 'PopVideo' game, described as 'Race to pop as many words as you can in the video'. Below the banner are three circular buttons: 'Gender Guesser', 'Prizes', and 'Best Images'. At the bottom, there is a leaderboard titled 'Today's' with a 'Top 10' ranking. The top of the leaderboard shows the following data:

Rank	Username	Score
6	charliegund	99 K
7	Flipmstr2	94 K
8	welloff	94 K
9	guest203595	70 K
10	Sher	60 K

# Investigating deeper into the user data: A shortcut in Question 1



Which **building** needs to be tagged?

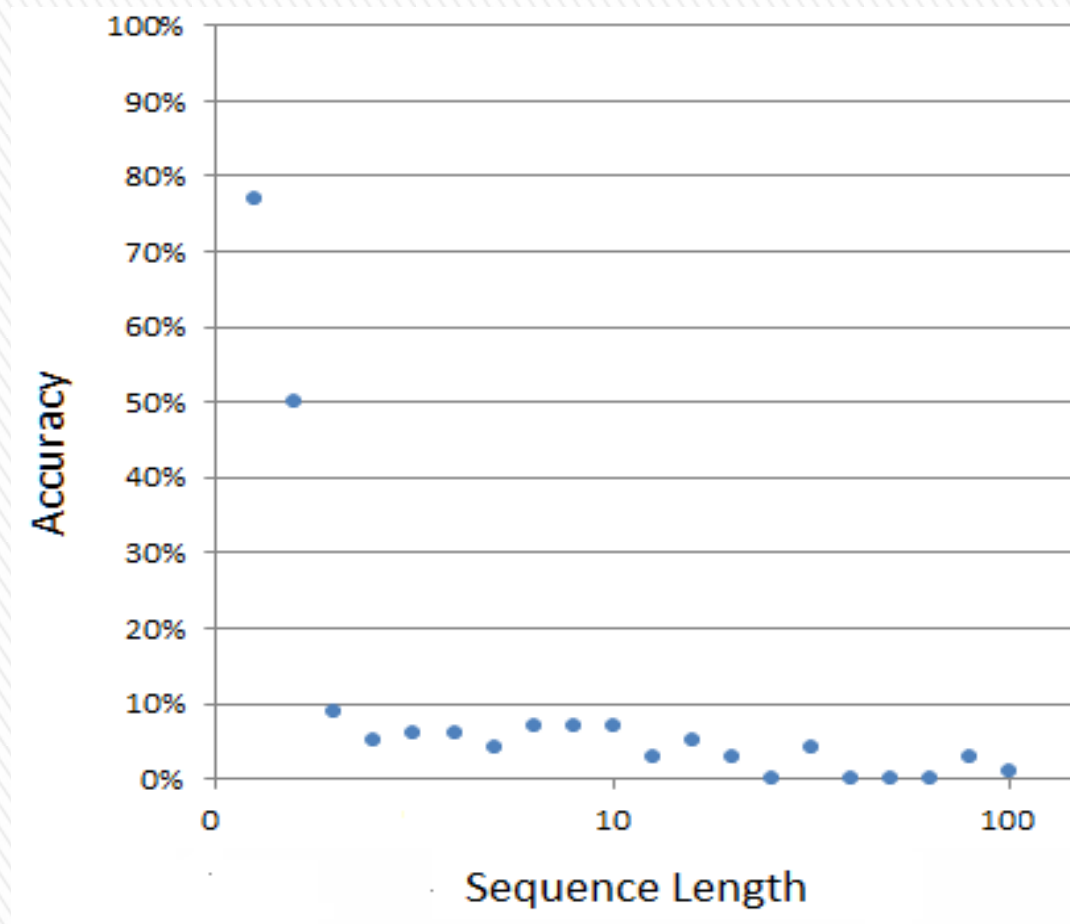
- Tag entire photo
- Tag part of photo
- Cannot determine

BACK NEXT

- » Some users selected “Cannot Determine” as a shortcut to artificially increase the number of photo classifications in order to win prizes.
- » These behaviors introduced unreliable data points to the data set.

# Data Cleaning:

## Method II – Detecting Invalid Sequence



» The length of suspicious sequences becomes an Indicator of Data Quality

# Comparisons Between the 2 Approaches

- » Intraclass Correlation Coefficient

ICC : A descriptive statistic to measure the similarity between data entries.

- » Crowd Consensus Score

Comparing crowd consensus with the ground truth

Ground Truth: Opinions of 3 Professionals from  
the Department of Civil Engineering

In both criteria, Method II shows better performance than Method I.

# Suggestions for Future Research

- » Blending objective questions with subjective ones.
- » Measuring users' confidence level
- » Providing users stronger motivation
  - I. Monetary Rewards: basic salary + extra bonus
  - II. Moral Encouragements:
    - honor list, social media recognitions
    - thank-you letters/notes from
    - the local community in Haiti.