Citizen Engineering: Methods for "Crowdsourcing" Highly Trustworthy Results

Zhi Zhai [1], David Hachen [3], Tracy Kijewski-Correa [2]
Feng Shen [1], Greg Madey [1]

[1] Computer Science & Engineering
[2] Civil Engineering & Geology Sciences
[3] Department of Sociology
University of Notre Dame

HICSS - 45
January 4-7, 2012
Grand Wailea, Maui, HI

Overview

» Background

» Challenges and Concerns

» Experiment: Haiti Earthquake Photo Tagging

» Experimental Results and Lessons Learned

» Summary and Future Work

Background

» "Cognitive Surplus" [Shirky 2010]

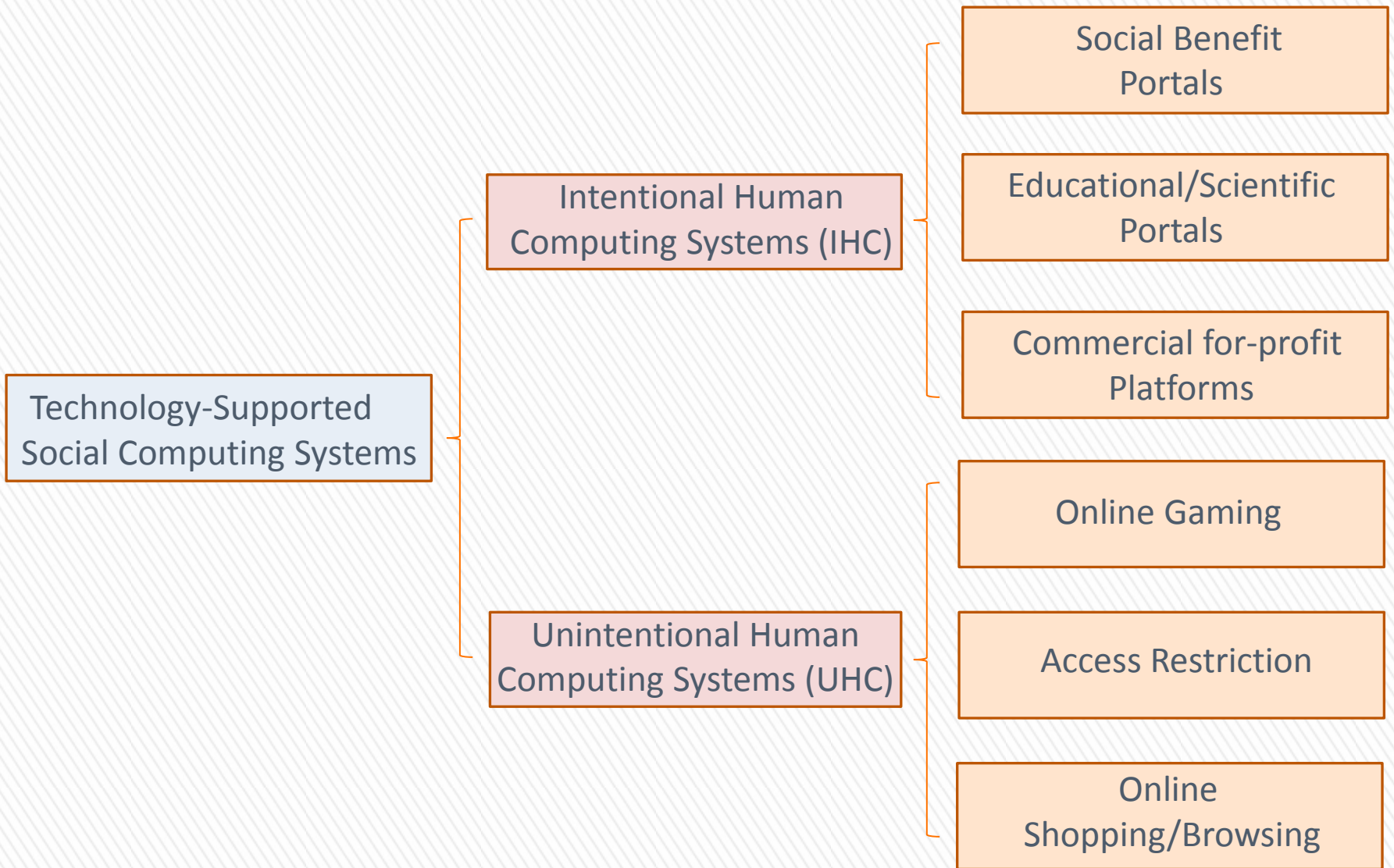Advances of Modern Technologies

=> More free time

=> A large proportion could be used more productively.

» The development of Information Technology.

» Exemplified by *Wikipedia*, a large number of loosely organized groups of people can effectively work together to achieve a common good.

Projects Based on Citizen Participation

Technology-Supported Social Computing Systems

- **Intentional Human Computing Systems (IHC)**
  - Social Benefit Portals
  - Educational/Scientific Portals
  - Commercial for-profit Platforms

- **Unintentional Human Computing Systems (UHC)**
  - Online Gaming
  - Access Restriction
  - Online Shopping/Browsing

Some questions remain unsolved yet....

Challenges

» Given the high diversity of online workers' background, how can we effectively extract reliable results from a myriad of crowd inputs of varying quality?

» To investigate, our prototype website:

  Haiti Earthquake Photo Tagging Experiment

  * 25 Questions Per Photo

  * 400 Photos

  * 204 Participants with at least 1 classification

  * 9318 Classifications

Previous Research We Drew Ideas From

» ImageCat [ImageCat 2012] :
Inspiration 1: Natural disaster photo analysis
for homeland rebuilding.

» Galaxy Zoo [Lintott 2008]:
Inspiration 2: Having fun while doing meaningful work.

» Open Source Software Movement
Inspiration 3: Virtual Organization/ Online Community

» The collective intelligence genome [Malone 2010]:
Inspiration 4: Human Collective Work
Organization and Evaluation.

Web Front Page

OSD-CI Experiment

Home **Login**

Welcome to the OSD-CI Experiment

You are invited to participate in an online experiment run by the OSD-CI. In this experiment, we hope to evaluate the effectiveness of online behavior in classifying (tagging) photographs of earthquake damage. If you decide to participate, we will ask you to login to a secure website, answer a few questions and take a brief tutorial. The tutorial will provide all the information you will need to tag photographs. We estimate that tagging each photograph will take less than two minutes. After completing the preliminary set of 25 photos you may continue tagging as many of the 400 photographs as you chose.

- **You will receive $10 if you choose to participate in this experiment.**

YOU ARE MAKING A DECISION WHETHER OR NOT TO PARTICIPATE. BY CLICKING THE "I HAVE READ ALL OF THE ABOVE TEXT, I AM AT LEAST 18 YEARS OLD AND WOULD LIKE TO PARTICIPATE" OPTION YOU INDICATE THAT YOU ARE A NOTRE DAME STUDENT, AT LEAST 18 YEARS OF AGE AND HAVE DECIDED TO PARTICIPATE HAVING READ THE INFORMATION PROVIDED ABOVE.

We have reached our quota of volunteers for the experiment. The experiment is now closed for new registrants. Only those that have registered to participate by 12-01-2010 may continue to participate.

* OSD-CI: Open Source Designing of Civil Infrastructure

Tutorial

Photo Classifying Tutorial

Jump To Step 2: Primary Elements

Step 2: Identify which primary elements of the building are visible and can be assessed

Question: What primary elements are visible and can be assessed in this photo? (Select all that apply)

Note: Not all elements will be visible in every photo, but photos can have multiple elements. Please identify all visible elements in the photo including both damaged and undamaged elements if they are visible.

| Response Options (select all that apply)	Definition of Elements	Example
COLUMN	Vertical element where the height (H) is considerably larger than the depth (D) and width (W).	
BEAM	Horizontal element where the width (W) is considerably larger than the depth (D) and height (H).	
SLAB	Horizontal element used as a roof or floor where the width (W) and depth (D) are considerably larger than the height (H).	
WALL	Vertical element used for partitioning where the height (H) and width (W) are considerably larger than the depth (D).	
Cannot determine	Photo does not appear to have any of the above elements.	

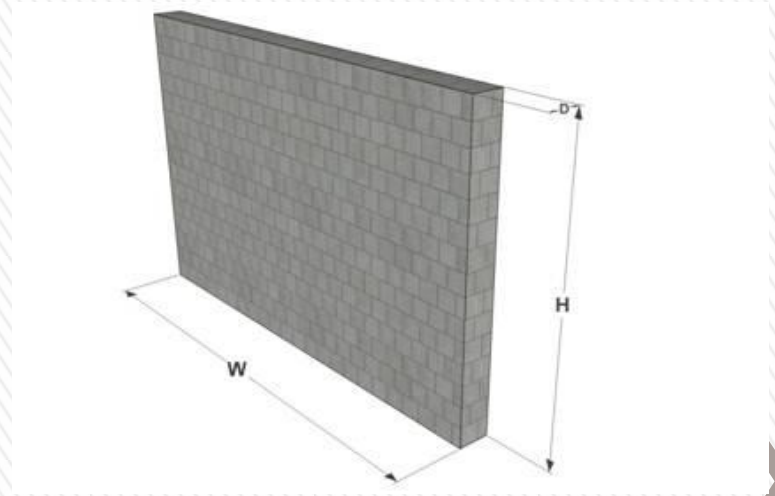# Potential Damaged Parts

Column
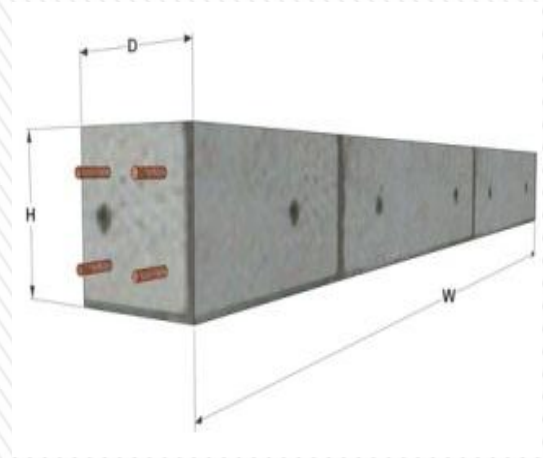
Beam
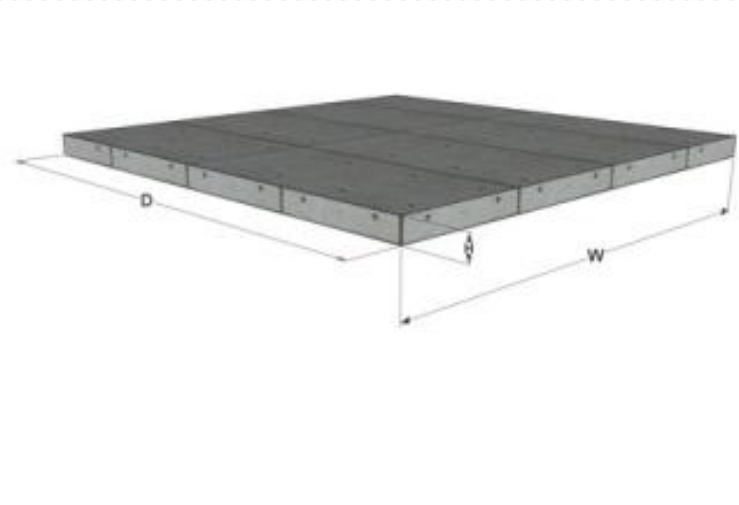
Slab

Wall

Damage to Column

Damage to Beam

Damage to Slab

Damage to Wall

User Tasks -

Photo Classification: What type(s) of damage does this photo have?

User Interface

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 Flow for a Typical Photo

Identifiable?

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

Damage Existence?

Damage Pattern?

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)?
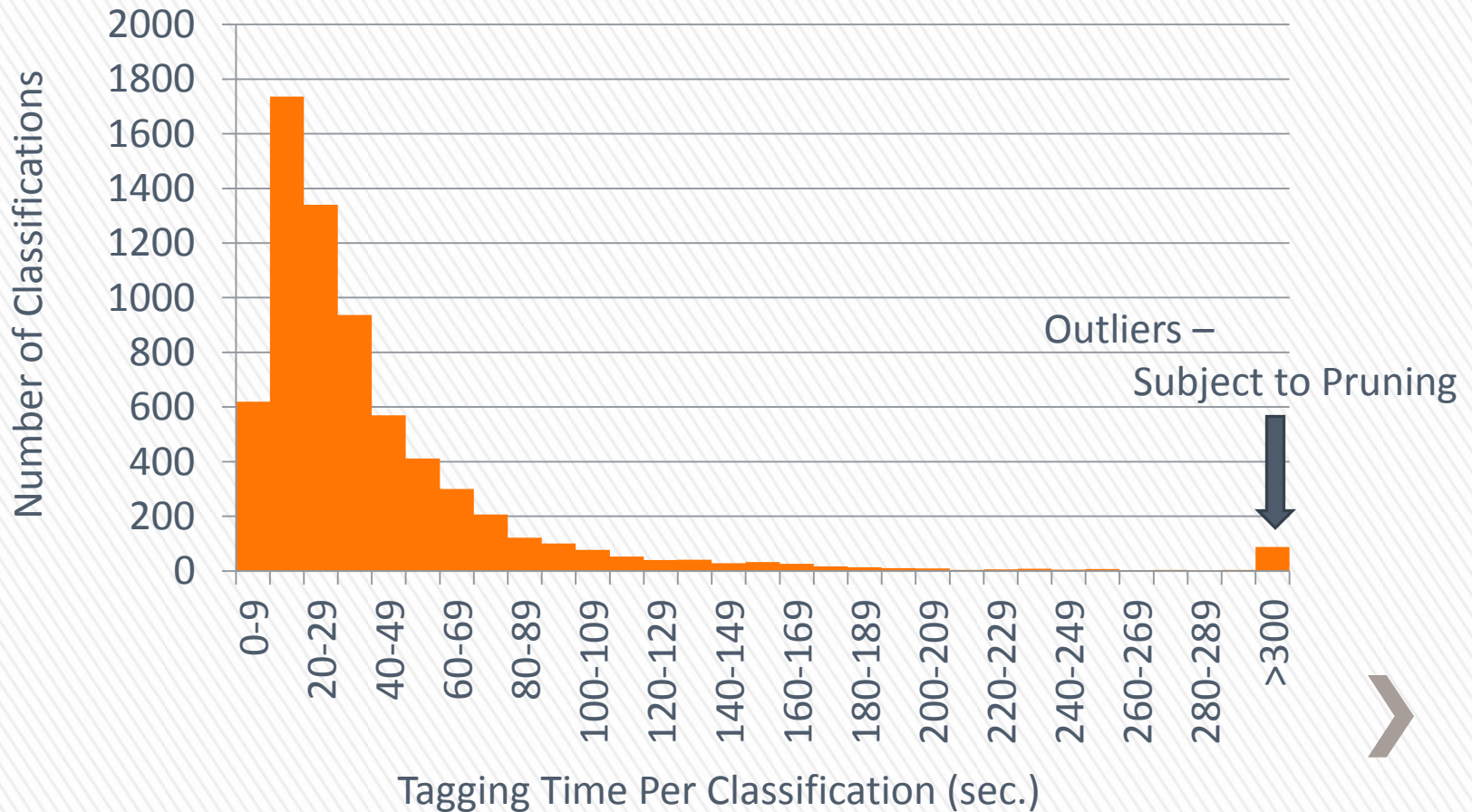- Yellow
- Red

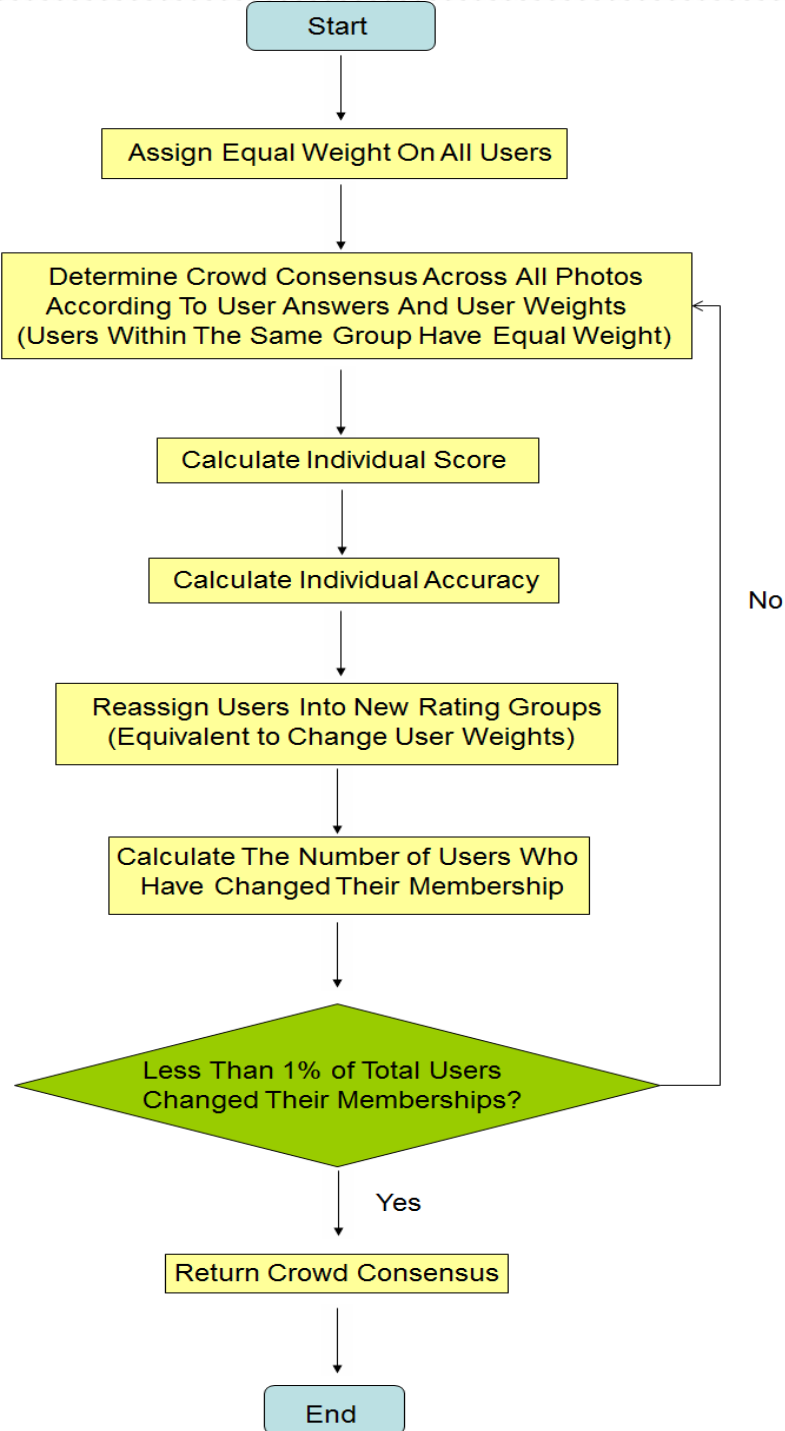BACK NEXT

Severity?

Post Experiment Data Analysis

Data Preparation - Cleaning

Sleeper: Tagging time more than 300 sec.

Data Preparation - Results

» Before Data Cleaning

> 9318 user-image pairs (Classifications)

» After Data Cleaning:

> 6186 user-image pairs (Classifications)

Algorithm

>> Goal: To retrieve good results from a large number of inputs, which came from a cohort of online users.

# Data Analysis...

Question 1:
Number of Classifications
correlated with User Accuracy ?

» More photo classifications

=> More Enthusiastic + More Serious

=> Higher Accuracy Rate?

Null Hypothesis: No statistically significant correlation between the Number of Photo Classifications and User Accuracy.

Alternative Hypothesis: There is a statistically significant correlation between the two statistics.

Number of Classifications vs. Accuracy
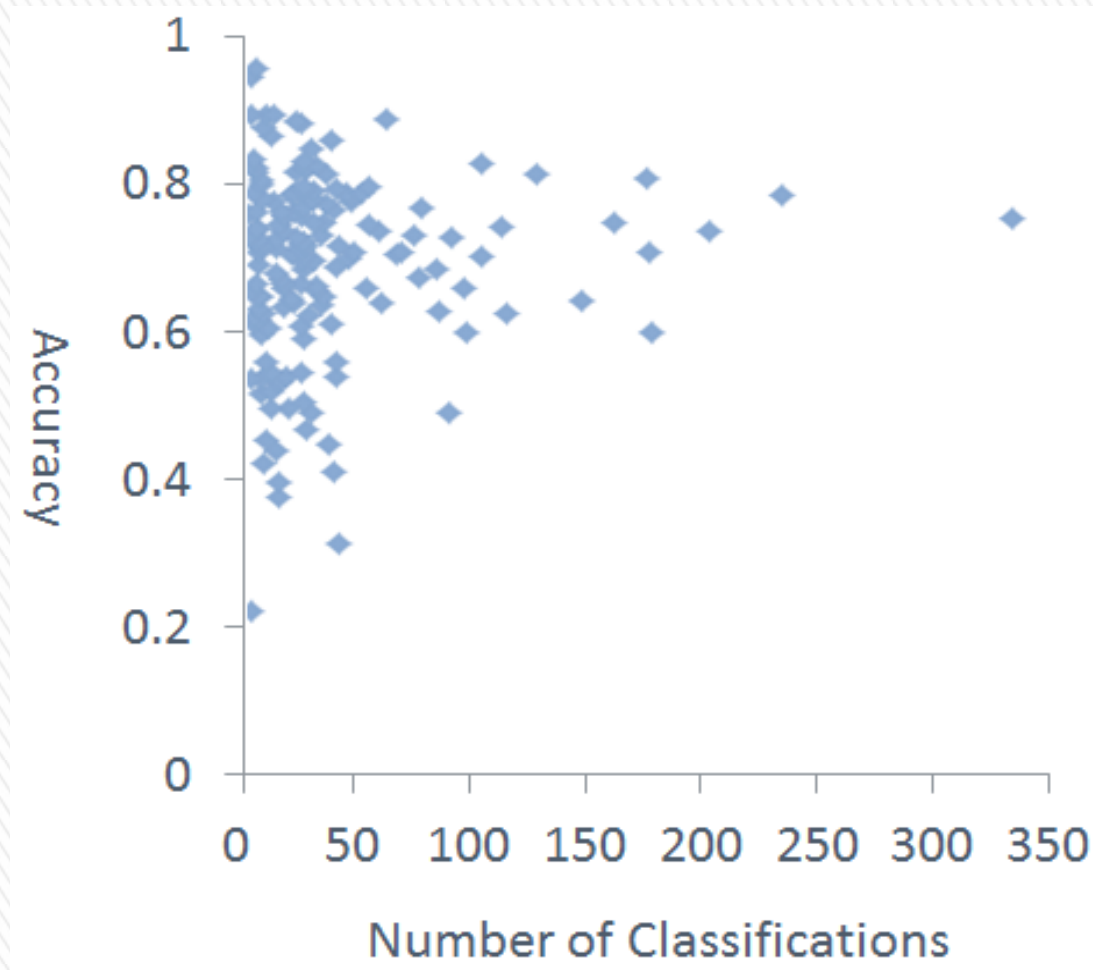
Number of Classifications correlated with User Accuracy ?

Method: Student's t-test for potential correlation.
Divide users into 2 accuracy groups, and then compare
the number of classifications associated with each group of users.

Answer: No. There is no Statistical Correlation.

Student's t-test: t-value = 0.95
                          z-value = 3.25
t < z, no statistically significant difference.
So, cannot reject the null hypothesis on 99% confidence
level.

Question 2:
Average Tagging Time is correlated with User Accuracy ?

» Longer Tagging Time
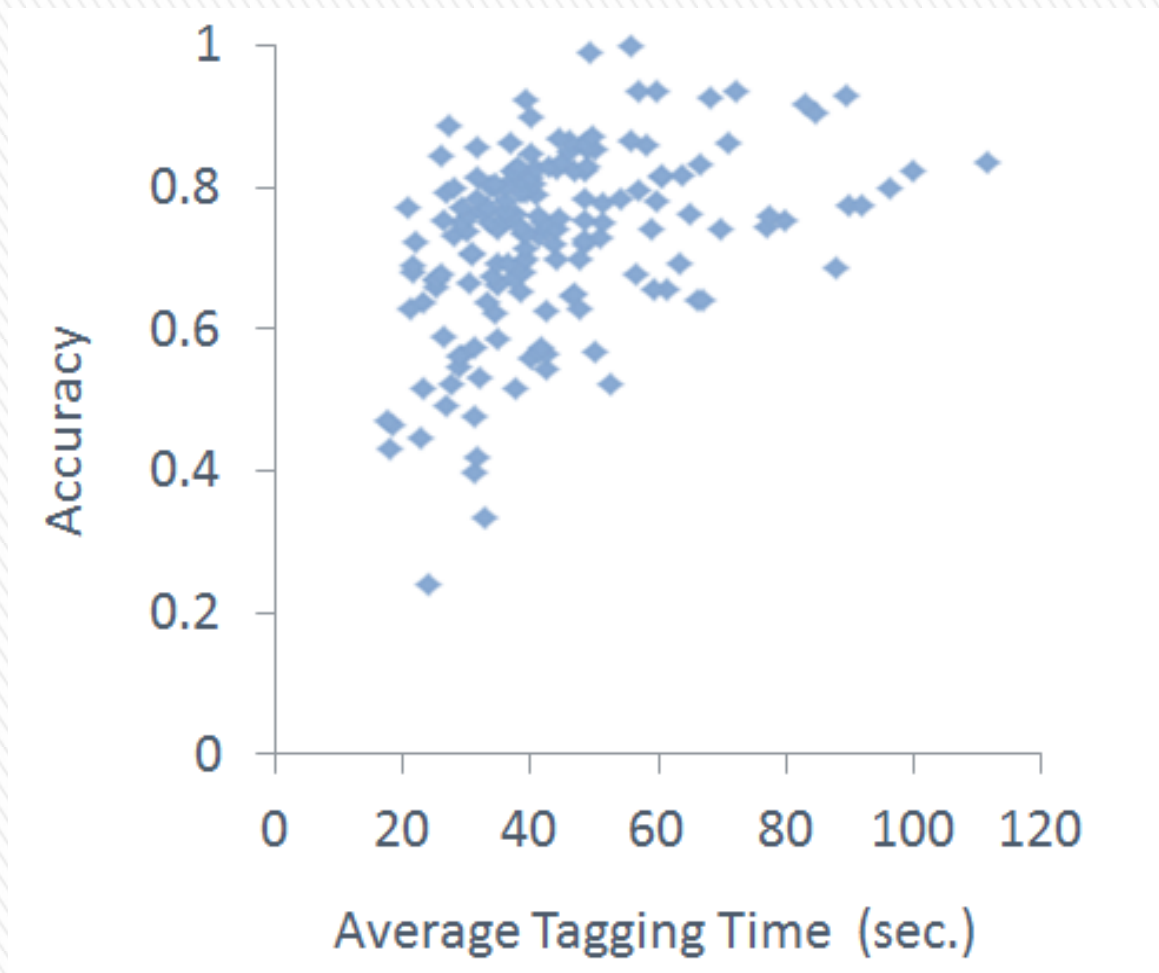
=> More Responsible + More Careful

=> Higher Accuracy?

Null Hypothesis: No statistically significant correlation between the Average Tagging Time and User Accuracy.

Alternative Hypothesis: There is a statistically significant correlation between the two statistics.

›

Average Tagging Time vs. Accuracy

Question :
   Average Tagging Time is
        correlated with User Accuracy ?

Again... Student's t-test

Answer: Yes. There is a statistical correlation.

        Student's t-test: t-value = 6.91
                              z-value = 3.25
t > z, there is a statistically significant difference.
So, reject the null hypothesis on 99% confidence level.

Summary

» Contributions

> Web portal architecture

> Invalid inputs detection

> Highly trustworthy result extraction

> Data Analysis

Future Work

Road Ahead ...
Redundancy vs. Efficiency:

» May not need to employ dozens of users

to classify simple photos.

» May need to add more users to classify

difficult photos to enhance the accuracy.

» What is the optimal number of workers per job?

Road Ahead …
Stratified Citizen Engineering Systems

» A whole spectrum of citizen engineers with variable backgrounds and expertise.

» Our preliminary research focuses on average citizen engineers.

» How about the high-end expert engineers? How can we shepherd junior engineers to fulfill complicated tasks ?

Acknowledgements

References

» [ImageCat 2012] http://www.imagecatinc.com/

» [Lintott 2008] C. J. Lintott, et. al. Galaxy Zoo: morphologies inspection of galaxies from the Survey. *Monthly Notices of the Royal Society*, 389:1179–1189, 2008.

» [Malone 2010] T. Malone, et. al. The collective intelligence genome. *MIT Sloan Management Review*, 51(3), 2010.

» [Shirky 2010] C. Shirky. Cognitive Surplus: Creativity and Generosity in a Connected Age. *Allen Lane*, 2010.

» Thanks very much.

» Questions.