Task Design I
Jérôme Waldispühl, McGill University

\[ a^2 + b^2 = c^2 \]
Designing HIT

Input

Task routing

Task design

Task aggregation

Output

Today
Know your audience

Money is not everything! A good design enable to:

- Facilitate entry of new users,
- Maintain a large community,
- Improve efficiency and quality of the results,
- Motivate workers to provide a good answer,
- Manage both competition & community aspects.
Amazon Mechanical Turk

• Small tasks
• Small reward
• Typically classification, translation & texts,
• New uses in sociology, psychology.

Find the meaning of the following hashtag

- For this hashtag below, enter its definition or its meaning. The definition should explain what that hashtag means.
- Include links if it refers to a website entering the full address, e.g. "Dubai" is an emirate in the United Arab Emirates #UAE. A city within this emirate is called Dubai also. # is a correct definition in case of #Dubai hashtag.
- If you do not know the definition just return the HIT so somebody else can enter the correct definition. If you fail to enter the right definition the HIT will be rejected and it will affect your stats.
- One Hashtag may have different meanings. For example #IMU stands for "I Miss You" and also for "Inbound Marketing University". Please enter as many meanings as you can.

Hashtag: #KFUPM

Enter your hashtag definition here:

(You must ACCEPT the HIT before you can submit the results.)
Demographics of Mech. Turk

Time spent on Mechanical Turk per week

- Less than 1 hour per week: 0.00%
- 1-2 hours per week: 5.00%
- 2-4 hours per week: 10.00%
- 4-8 hours per week: 20.00%
- 8-20 hours per week: 25.00%
- 20-40 hours per week: 15.00%
- More than 40 hours per week: 5.00%

Number of HITs completed per week

- Less than 1 HIT per week: 0.00%
- 1-5 HITs per week: 5.00%
- 5-10 HITs per week: 10.00%
- 10-20 HITs per week: 20.00%
- 20-50 HITs per week: 30.00%
- 50-100 HITs per week: 20.00%
- 100-200 HITs per week: 15.00%
- 200-500 HITs per week: 10.00%
- 500-1000 HITs per week: 5.00%
- 1000-5000 HITs per week: 2.00%
- More than 5000 HITs per week: 1.00%

Weekly Income from Mechanical Turk

- Less than $1 per week: 40.00%
- $1-$5 per week: 15.00%
- $5-$10 per week: 10.00%
- $10-$20 per week: 10.00%
- $20-$50 per week: 10.00%
- $50-$100 per week: 5.00%
- $100-$200 per week: 5.00%
- $200-$500 per week: 5.00%
- More than $500 per week: 5.00%

(Ipeirotis, 2010)
Demographics of Mech. Turk

Mechanical Turk is my primary source of income (paying bills, gas, groceries, etc)

- FALSE: 70.00%
- TRUE: 30.00%

Mechanical Turk is my secondary source of income, pocket change (for hobbies, gadgets, going out)

- FALSE: 50.00%
- TRUE: 50.00%

I participate on Mechanical Turk to kill time

- FALSE: 90.00%
- TRUE: 10.00%

Comparing India and United States

- India: 80.00%
- United States: 20.00%

(Ipeirotis, 2010)
Demographics of Mech. Turk

(Ross et al., 2010)
Demographics of Mech. Turk

Nationalities

Gender

(Ross et al., 2010)
reCAPTCHA

“Completely Automated Public Turing test to Tell Computers and Humans Apart”

• Prevent abuse from bots,
• Help to digitize books.

(von Ahn et al., 2008)
reCAPTCHAs

- OCR cannot recognize 30% of the words,
- 1 word digitizes difficult words,
- 1 word for control,
- 750,000,000 users (>10% of humanity).

(von Ahn et al., 2008)
Breaking CAPTCHAs

1. Hypothesize locations of letters in the image,
2. Extract strings that form candidate words,
3. Choose the most likely word(s).

Success rate ranging between 80% and 95%.

(Mori and Malik, 2003)
Usability/Difficulty trade-off

• Text-based CAPTCHAs can be difficult for foreigners.
• Color, font and string length can impact security and usability.

Fully secure CAPTCHAs are possible but one must consider the time required to solve it.

(Yan and El Ahmad, 2008)
Visual vs. Audio CAPTCHA

<table>
<thead>
<tr>
<th></th>
<th>Visual</th>
<th>Audio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Success rate of (original) CAPTCHA</td>
<td>~85%</td>
<td>~45%</td>
</tr>
</tbody>
</table>

But a simple re-design of the Audio CAPTCHA interface boosted the success rate to 65%!
200,000,000 of web users play online games once a week.

Casual games are the most popular and characterized by:
• Low level entry (design & setup)
• Simple mechanics,
• Non-punishing,
• Fast paced game,
• gender-neutral & non-violent.

(igda, 2008)
Game With a Purpose

(von Ahn and Dabbish, 2004)

(Kawrykow et al., 2012)
Demographics of casual games

- Motivations: Relax, pass time, socialize, achieve goals,
- Do not see themselves as “gamers”,
- Do not spent money on specialized hardware,
- Majority of 30-45 yo female playing puzzle, cards, words.

Demographics are essential to define potential & objectives.

GWAP are games, and a game must be fun!

(igda, 2008)
Citizen science

Workers:
• Highly motivated workers,
• Limited by lack of knowledge.

Scientists:
• Need large dataset,
• Must simplify the human task.
Workers gain knowledge

- Variant of the 20 question game used to train classifier.
- Task routing is heavily used to determine which question will be the most informative.

(Branson et al., 2010)
Citizen-science UI

Avoid bias in data collection due to:
- over-estimation of easy targets,
- over- or under-report of measurements.

(Sullivan et al., 2008)
Learning by doing

Learn a language and translate the web.
• Free!
• Use image tagging, sentence translation, etc.
• Applicable to any knowledge field.
Learning by doing

Long involvement and motivated users: The system has time to estimate the expertise of users and see evolution over time.

Challenges:
• How to select lessons?
• When does the system “teach” or “query”?
• How to estimate the expertise of users?
• How to design the UI to promote learning?
• Can we pair users? With which benefit?
Temporary market

In case of emergencies, crowdsourcing techniques could quickly engage a large population... for a short time. How to build such systems?
Supporting workers

• Search for the best task,
• Information to decide to do a task vs. another one,
• Monitor progresses and productivity,
• Communicate problems,
• Understand competition,
• Find support within the community,
• Express opinions.
Supporting requesters

• Automate and optimize task design,
• visualize workers’ expertise and progresses,
• Meet deadlines,
• Meet budget,
• Understand competition with other requesters,
• Express opinion about workers.
New trends

Mechanical turk will no longer be the primary resource for crowdsourcing applications?

• Query language for databases (Parameswaran & Polyzotis, 2011)
  $$\text{travel}(I) := \text{rJpeg}(I), \text{hClean}(I), \text{hBeach}(I), \text{aLarge}(I)$$

• Integration inside word processors (Bernstein et al., 2010)

![Diagram of The Human Macro and Mechanical Turk Worker Preview](image)
Reference

Human Computation
Edith Law, Luis von Ahn
Morgan & Claypool Publishers