# Reputation in mTurk

*Building trust between workers and requesters*

Ana Carmona
Bhavna Verma
Matt Lease

INF 384C - Organizing Infromation
School of Information | Univeristy of Texas at Austin

## PROBLEM

In the crowdsourcing community there has been a great deal of debate regarding ethics and justice in crowdsourced labor markets. Amazon Mechanical Turk, the prime example of this, currently neglects its users with a lack of proper metrics and features to build a sense of trust between workers and requesters, and proper methods to reduce the occurrence of misconduct by both parties, that ultimately affect the quality of the work produced with the service.

Where few research papers address the issue of workers being exploited, there are also numerous works that address the problem of workers delivering low quality work. Requesters are oftentimes accused of creating fraudulent HITs with exploitative purposes and even taking workers' work without pay.

Currently, mTurk doesn't provide built-in functionality that allows its users to rate and review each other, with the resulting consequence of both parties being blind about the intentions of the other party.

## RESEARCH

To get a thorough overview of the situation inside Mechanical Turk, we firstly studied the functionality of mTurk both through the worker and requester's perspectives. We performed heuristic evaluations of the tool to understand its strengths and weaknesses in the labor itself, and also in the usability of the service. Additionally, we analyzed turkOpticon, a third party browser plugin, that allows workers to review requesters, but doesn't aloow them to rate workers or respond to worker's comments.

We also performed a literature review of recent work regarding Mechanical Turk, reputation systems in labor markets and otherwise, people and service ratings in web services, and the future of crowdsourcing.

Finally, we reviewed other web services similar to Mechanical Turk and other well established web services that did incorporate rating and reputation systems to build trust in their communities. The purpose of this was to review best practices outside this context, to determine the best possible approach to implement a reputation system that would be comprehensive and appropriate to this service in particular.

**Literature Review**

ACM
ASIS&T
CHI
Others

- Crowdsourcing
- Labor Markets
- Reputation systems
- Amazon Mechanical Turk

**Benchmarking**

eBay
Amazon
Yelp
Other crowdsourcing

- Reputation
- Ratings
Comments
Feedback

**Heuristic Evaluation**

Amazon mechanical Turk
Turkopticon

- Requester role play
- Worker role play
Functionalities
Features

## ANALISYS

After a thorough analysis of the information obtained, we determined the best actions to solve the problem. Firstly, the need for an external reputation system was solidified. For the quality of the work produced in Mechanical Turk, the motivation and participation of its users, and the creation of a community that relies and trusts each other, a rating system must be implemented.

We also determined a new set of criteria to use for evaluating users and their services. These factors are listed below:

- Workers:
  Time to complete, quality of work produced, performance statistics, HITs completed, identity verified, willing to communicate.
- Requesters:
  Communication, generosity, fairness, promptness, adequate description of HIT, adequate provided time for completion, overall perception?, approval rate of worker work, rejection rate, amount of hits created, identity verified.

## RESOLUTION

To solve the problems discussed above, we have designed a third party browser extension that allows the users to review their interactions with each other, and provide feedback on the ratings already provided of them.

On the technical side, the ratings and feedback would be stored in an external database that can be accessed through the browser extension, which acts as an access point between the mTurk interface and the database.

When a worker is looking for HITs or a requester is managing her workers, both parties can access the other's ratings if already available in the site, as well as adding their own opinions.