# Enabling Trust in Crowd Labor Relations through Identity Sharing

**Jörn Klinger**
Department of Linguistics
University of Texas at Austin
klinger@utexas.edu

**Matthew Lease**
School of Information
University of Texas at Austin
ml@ischool.utexas.edu

## ABSTRACT

While online Crowdsourcing marketplaces provide a powerful avenue for facilitating new forms of information-driven micro-labor, their practical value is significantly reduced by worker “spam” and employer fraud. We hypothesize anonymity of parties is a major source of these problems, and we thus propose a human-centric solution: encourage employers and workers to voluntarily de-anonymize in order to reap a potential benefit of more productive and profitable labor interactions. To facilitate voluntary identity sharing, we have built a prototype identity management application allowing individuals to associate their crowdsourcing worker/employer identities to their public profiles on social network sites. By providing a vehicle for identity sharing, the prototype provides the foundation for a future user study of employers and workers engaged in known-identity crowd labor relationships.

## Keywords

Crowdsourcing, online labor, online identity, trust, spam

## INTRODUCTION

Crowdsourcing (Howe, 2006) has emerged as a powerful new mechanism for connecting employers and workers for mutually-beneficial micro-labor relationships. Many companies, individuals and scientists around the globe now use Crowdsourcing to solve a variety of problems where automation falls short, like analyzing language and images.

On large-scale online Crowdplatforms like Amazon's Mechanical Turk (AMT), advertised as "a marketplace for work", *requesters* typically post large volumes of micro-tasks, which are then accepted and completed by anonymous online workers for monetary compensation. Tasks that particularly lend themselves to such a model of labor include checking videos or pictures for sensitive content, transcribing audio files, or assessing the relevance of web-pages for a search result. Since individual micro-tasks involves very brief work (e.g. making a judgment), typical compensation is less than five cents per task. However, workers typically complete many such micro-tasks, so compensation is often discussed in aggregate terms. Overall, Crowdsourcing offers a fast, cost-efficient solution for employers, while workers have the freedom to work when they want, for whoever offers work on a given day, on whichever tasks they want to work on. Nonetheless, many challenges remain, such as determining fair and legal pay for work of varying difficulty and quality requirements in a global economy (Horton and Chilton, 2010).

Anywhere money changes hands becomes a natural target for spam and fraud, and Crowdsourcing has quickly become such a target, especially as the volume of work and money accelerates. While a considerable amount of effort has been directed toward automated approaches to combating such absuses (see Related Work section), we hypothesize that common countermeasures may address the symptoms of spam and fraud rather than underlying causes.

After a brief discussion of spam and fraud on Crowdsourcing platforms, as well as common countermeasures, we describe our approach of linking worker and requester IDs on Crowdsourcing platforms to their real world identities in order to promote trust between parties. We describe a prototype application for identity sharing and discuss future work using it to study employers and workers engaged in known-identity labor relationships.

## RELATED WORK

A variety of work has focused on fighting spam and fraud. *Spam* refers to work submitted by individuals which purports to satisfy the requested task but which may have been produced without care or attention to quality (e.g. in the most egregious case, random clicking on multiple choice answers). Ipeirotis, Provost and Wang (2010) suggest that spammers generate about 30% of the answers submitted to AMT. Kittur, Chi and Suh (2008), who defined spam simply as invalid answers, found it at a rate as high as 48.6%. While often produced by humans, there may also be significant spam generated by automated “bots” that accept Crowdsourcing tasks, pretending to be human workers, and submit poor work (McCreadie et al., 2010).

*ASIST 2011*, October 9–13, 2011, New Orleans, LA, USA.

*Fraud* occurs when employers do not pay workers for completed work, or trick workers to installing malicious software, etc. Many forms of unethical work also exist. Ipeirotis (2010) reports that over 40% of the tasks posted on AMT by requesters who joined AMT between September 2010 and October 2010 asked workers to produce a fake rating or comment, create a fake account of some sort, perform fake clicks or other dubious activities that can be counted as spam. He envisions a future scenario in which spamming workers performs task of spamming requesters.

A common method for reducing spam when dealing with large amounts of data is to rely on redundancy to identify correct answers (Ipeirotis, Provost and Wang, 2010). The idea here is to have multiple workers perform the same task. Combining redundancy with majority vote improved the quality of data in many cases - see Pameswaran and Polyzotis (2011) as well as Eagle (2009). While it improves the quality of the data, as Ipeirotis, Provost and Wang (2010) point out, redundancy is expensive.

Another popular method for fighting spam is the use of "gold" standard data. In a Crowdsourcing task that required workers to rate the quality of Wikipedia articles, Kittur Chi and Suh (2008) embedded a series of questions to which the answers were known and thus easily verifiable - gold standard data. While the use of gold standard data allows one to identify spammers and exclude their answers, its use is expensive. Its creation requires humans with enough expertise, such that for large-scale data, costs rise.

Thus, while methods for combating spam exist, they are expensive and therefore undermine the prime advantage of Crowdsourcing - its low cost. Also, while treating the surface effects of spam, the above-mentioned methods are costly and allow little punishment for cheating parties. Requesters and particularly workers can create new profiles without much effort (Ipeirotis 2010b).

## IDENTITY LINKING

Our approach aims to fight fraud by taking away some of the anonymity from Crowdsoucing to establish trust between workers and requesters. The crucial idea is to establish a link between a worker or requester ID and the respective person's social network profile on sites like Facebook, Twitter, or LinkedIn[1]. Since users upload their photos, write about their lives and handle a significant part of social and professional interactions on social network sites, these profiles constitute an online approximation of real world identities that are established over time. Consequently, building, deleting and re-creating a profile with an established reputation is a major investment. If, for example, a worker produced spam, he would be subject to public criticism and exclusion. An employer committing fraud would face similar repercussions. While one could certainly invent a new social network profile, establishing public trust is time and effort-intensive. Crowdsourcing is valuable in aggregate, whereas individual tasks and transactions do not typically provide sufficient reward to motivate spam/fraud if some consequences were in place.

To link worker or requester IDs to their social network profiles, we have developed the following process, implemented by our prototype. Crowdplatform users log into their preferred social network site and authorize an application that allows them to enter their respective worker or requester ID. The user's work-relevant profile data is thus stored in our backend database, and a message containing a confirmation code is sent to the user's ID on the Crowdsourcing platform. From there the user retrieves the confirmation code and enters it into the application on the social network site, which establishes the linking of identities. The profile information we record in the database includes: languages spoken, hometown and current location (information that is already openly available to anybody on social networks such as Facebook). Sensitive data is not recorded, and data obtained is not disclosed to third parties.

After the identity linking is complete, requesters: (1) log into the application, (2) specify particular qualifications (for example competence in a certain language) that their task requires, (3) search the database for workers that match these qualifications, and (4) issue tasks specifically to them. They can also save and manage groups of trusted workers and send tasks to whole groups. Special qualifications and activation codes are sent to the workers who can then access the jobs offered to them through the application.

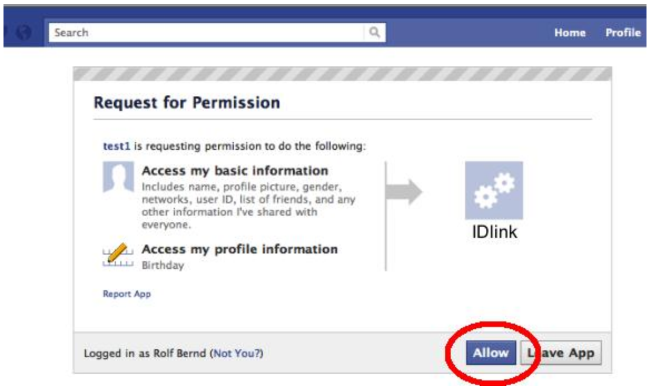
Our design also allows for allowing workers/employers to rate one another or provide other Web 2.0 feedback that has similarly transformed product shopping online. Besides de-anonymization in general, disclosed profile information also allows requesters to find workers that are better suited for a specific task. This is particularly relevant for research using Crowdsourcing participants (Mason & Suri, 2010), but it also improves the quality of work in general (Law, Bennett and Horwitz, 2011). Longer lasting worker-requester relationships are fostered, increasing trust.

## THE PROTOTYPE

Our current prototype links IDs between Facebook and AMT. We have built a front-end application for Facebook, and we store data in our own back-end SQL database. In the following description we focus on how a worker's ID is linked, but linking a requester ID is just as simple.[2] Assuming that a user is already signed up to both platforms, we first authorize the Facebook application, by simply navigating to its URL and clicking "Allow".

[1] URLs: www.facebook.com, www.twitter.com, www.linkedin.com or trusted non-social network ID providers like OpenID (openid.net).

[2] The procedure is very similar for requesters, with the only difference being that they click on "Link your Facebook profile to a requester ID" in the main menu.



Request for Permission

test1 is requesting permission to do the following:

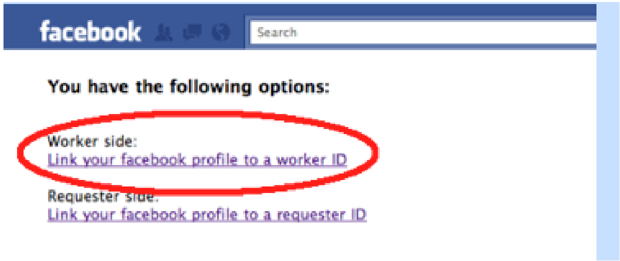
Access my basic information
Includes name, profile picture, gender, networks, user ID, list of friends, and any other information I've shared with everyone.

Access my profile information
Birthday

Logged in as Rolf Bernd (Not You?)

Figure 1. Authorizing the application.

After authorizing with the application, we see the main menu. Since no ID linking has taken place so far, the options available are to link the Facebook profile to a worker or a requester ID (it is possible to link a single profile to both a worker and a requester ID).



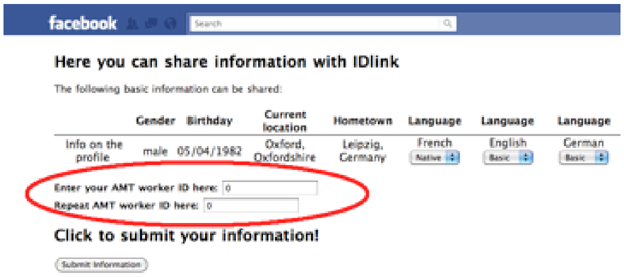
You have the following options:

Worker side:
Link your facebook profile to a worker ID

Requester side:
Link your facebook profile to a requester ID

Figure 2. Linking the Facebook profile to the worker ID.

The next step in this process is to click on "Link your Facebook profile to a worker ID." On the next screen, we see the information that the application reads from our Facebook profile. The current prototype accesses gender, birthday, current location, hometown and languages spoken. To give an idea of upcoming features, the prototype allows one to specify their skills in these languages.



Here you can share information with IDlink

The following basic information can be shared:

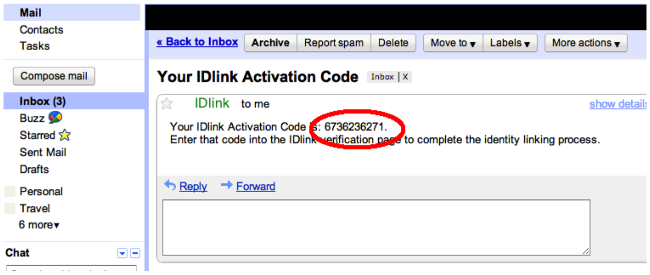
| | Gender | Birthday | Current location | Hometown | Language | Language | Language |
|---|---|---|---|---|---|---|---|
| Info on the profile | male | 05/04/1982 | Oxford, Oxfordshire | Leipzig, Germany | French<br>Native | English<br>Basic | German<br>Basic |

Enter your AMT worker ID here: 0
Repeat AMT worker ID here: 0

Click to submit your information!

Figure 3. Specifying language skills and entering worker ID to proceed with the linking of IDs.

We now enter our AMT worker ID into the respective field, confirm that ID and click "Submit Information". Once the information is submitted, it is temporarily stored in a centralized database. An activation code is sent to the email address associated with the previously specified worker ID.



[redacted]

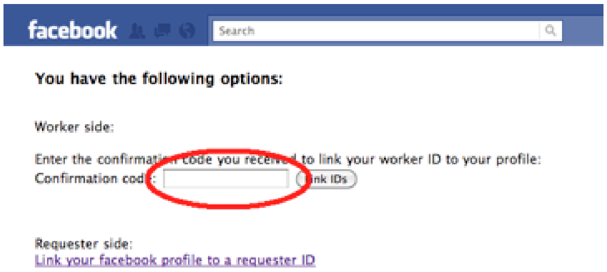
Your IDlink Activation Code

IDlink to me

Your IDlink Activation Code is: 6736236271.
Enter that code into the IDlink verification page to complete the identity linking process.

Figure 4. An activation code is send to the email associated with the users AMT worker account.

Then we retrieve confirmation code from that email and enter it into the application's main menu. Subsequently, the worker's information is permanently stored in the database.



You have the following options:

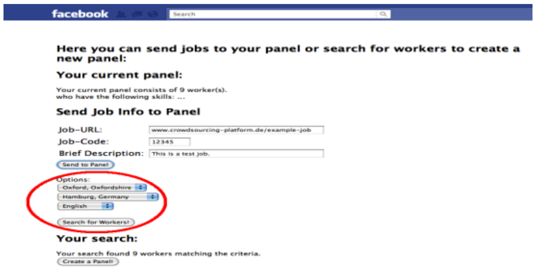
Worker side:

Enter the confirmation code you received to link your worker ID to your profile:
Confirmation code:

Requester side:
Link your facebook profile to a requester ID

Figure 5. The confirmation code is entered.

Once identities have been linked, workers can use the application to find jobs suited for them. Through the main menu, requesters manage and create groups of workers with specific skills or backgrounds and advertise jobs to them. The current prototype allows requesters to use hometown, current location and languages spoken as criteria when searching the centralized database for workers. Once a panel is established, the requester can issue jobs to the panel. To do so, they create the job on AMT, enter the link to the job into the "Job-URL" field, specify a code that the worker will have to enter as a qualification for the job, and add a brief description. Then they click "Send to Panel" in order to advertise the job to one of their worker panels.



Here you can send jobs to your panel or search for workers to create a new panel:

Your current panel:

Your current panel consists of 9 worker(s).
who have the following skills: ...

Send Job Info to Panel

Job-URL: www.crowdsourcing-platform.de/example-job
Job-Code: 12345
Brief Description: This is a test job.

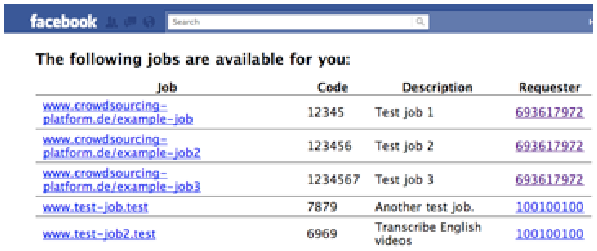
Options:
Oxford, Oxfordshire
Hamburg, Germany
English

Your search:

Your search found 9 workers matching the criteria.

Figure 6. Requesters can group workers into "panels" issue tasks to these panels as a group.

Analogously, workers see a list of jobs offered to them based on their skills and background.



The following jobs are available for you:

| Job | Code | Description | Requester |
|---|---|---|---|
| www.crowdsourcing-platform.de/example-job | 12345 | Test job 1 | 693617972 |
| www.crowdsourcing-platform.de/example-job2 | 123456 | Test job 2 | 693617972 |
| www.crowdsourcing-platform.de/example-job3 | 1234567 | Test job 3 | 693617972 |
| www.test-job.test | 7879 | Another test job. | 100100100 |
| www.test-job2.test | 6969 | Transcribe English videos | 100100100 |

**Figure 7. The Worker HQ allows workers to access and manage the jobs offered to them.**

For each job, workers can see the specified task link, qualification code, description, and a link to the requester's (corporate) Facebook profile, giving workers the opportunity to investigate the requester before taking a job.

## CONCLUSION AND FUTURE WORK

We have discussed how fraud and spam currently limit practical benefits of Crowdsourcing. We also reviewed existing solutions. We described our approach, which, instead of fighting the symptoms of spam, aims to prevent it by increasing trust relationships between workers and requesters. The applications allows users to link their Crowdsourcing IDs to social network profiles, which (without disclosing sensitive data) takes away some of the anonymity in the Crowdmarket leading to the possibility of holding users responsible for their fraud. While the current prototype is fairly basic and supports Facebook and AMT, future versions will extend to inter-operate with other public identity mechanisms and crowdsourcing platforms. The approach also lends itself to a history / reputation system, allowing employers to suggest fitting jobs to workers or workers to requesters based on their history. A potential concern is reduced anonymity leading to online slander. In the case of our application faulty behavior in online labor would then affect users' private online persona. The severity of this issue and whether intermediate solutions in which workers's avatars are not disposable, but their anonymity is preserved, are feasible will be within the scope of a more general, controlled user study.

In this study, Crowdworkers and requesters will use the prototype for a series of tasks varying in their structure and complexity. For example, some tasks will benefit from workers having experience in similar tasks - this allows testing potential advantages of a creating a stable panel of workers, while other tasks are simpler and do not require such prior knowledge. In addition to measuring time invested, money spent/earned as well as the amount of spam and fraud, participants will evaluate their satisfaction with the application and rate how trustworthy they found the workers/requesters they interacted with. Results will be checked against a control group issuing/working on a similar set of task, but using bare AMT sans the application. We hypothesize that identity sharing will significantly increase trust and consequently decrease spam and fraud, simultaneously improving cost-efficiency through better matching workers to tasks.

## REFERENCES

Eagle, N. (2009). txteagle: Mobile crowdsourcing. Internationalization. *Design and Global Development*, 447-456.

Horton, J.J. and Chilton, L.B. (2010). The labor economics of paid crowdsourcing. Proceedings of the 11th ACM conference on Electronic commerce, 209-218.

Howe, J. (2006) The Rise of Crowdsourcing, *Wired*, 14(6), URL (accessed 12 May; 2011): http://www.wired.com/wired/archive/14.06/crowds.

Ipeirotis, P. (2010). Mechanical Turk: Now with 40.92% spam. Accessed 12 May, 2011. http://behind-the-enemy-lines.blogspot.com/2010/12/mechanical-turk-now-with-4092-spam.html

Ipeirotis, P. (2010a). Be a Top Mechanical Turk Worker: You Need $5 and 5 Minutes. Accessed 12 May, 2011. htttp://behind-the-enemy-lines.blogspot.com/2010/10/be-top-mechanical-turk-worker-you-need.html

Ipeirotis 2010b: Mechanical Turk, Low Wages, and the Market for Lemons. Accessed 12 May, 2011. http://behind-the-enemy-lines.blogspot.com/2010/07/mechanical-turk-low-wages-and-market.html.

Ipeirotis P., Provost, P. & Wang, J. (2010). Quality management on amazon mechanical turk. In *HCOMP '10*, New York, NY, USA.

Kittur, A., Chi, E., Suh, B. (2007). Crwodsourcing User Studies With Mechanical Turk. *CHI 2007: ACM Conference on Human-factors in Computing Systems.*

Law, E., Bennett, P. & Horvitz, E. (2011). The Effects of Choice in Routing Relevance Judgments . To appear as a short-paper in *Proceedings of the 34th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2011)*.

Mason, W. & Suri, S. (2010). Conducting behavioral research on Amazon's Mechanical Turk. Available at SSRN: http://ssrn.com/abstract=1691163.

McCreadie, R., Macdonald, C. & Ounis, I. (2010. Crowdsourcing a news query classification dataset. *In the SIGIR 2010 Workshop on Crowdsourcing for search evaluation (CSE 2010), 31-38.*

Parameswaran, A. & Polyzotis, N. (2011). Answering Queries using Humans, Algorithms and Databases. *5th Biennial Conference on Innovative Data Systems Research (CIDR'11).*