

Crowdsourcing community heuristic evaluations

Simon à Campo¹ & Vassilis Javed Khan²

*Faculty Industrial Design, Eindhoven University of Technology, Eindhoven, NL
(Email: ¹ mail@simonacampo.nl; ² v.j.khan@tue.nl)*

Abstract. Crowdsourcing is growing in both industry and academia, providing new ways to conduct work. However, these online working environments show similarities with the industrial revolution, where workers have few to no rights. Although crowdsourcing is a new phenomenon, online communities have quite some history. We find a resource in the literature on how to build online communities and applied them to crowdsourcing platforms. We have gathered and adjusted community heuristics to evaluate crowdsourcing platforms. To support the evaluation task, we have developed a system which we present as a demo: <http://tinyurl.com/ecscwdemo>

Introduction

Crowdsourcing allows everyone with internet access to work by contributing to open calls (tasks) in return for an often monetary compensation. Amazon's Mechanical Turk¹ (MTurk) provides hundreds of thousands of microtasks for workers to fulfil, but fail to support their workers adequately. Most crowdsourcing platforms do not fulfil their workers' technical and social needs (Gray, Suri, Ali, & Kulkarni, 2016). Researchers do a call to recognize the sociality of work and the shared identities produced through paid collaboration (Kittur, Nickerson, & Bernstein, 2013). Some researchers even claim that crowdsourcing platforms are ideally similar to open-source communities (Stewart, Huerta, & Sader, 2009). The social element seems to be an important intrinsic motivation for the continued use of crowdsourcing platforms (Brabham, 2010; Soliman & Tuunainen, 2015; Zheng, Li, & Hou, 2011). Without the platforms providing these tools, workers have shown to create their own tools for collaboration (Gray et al., 2016). To our knowledge, faircrowdwork.org² is the only platform that attempts to list and rank crowdsourcing, currently listing 11 platforms. From a past version of crowdsourcing.org, we know that there are almost three thousand crowdsourcing platforms (Crowdsourcing.org, 2016). Therefore, this demo explores an approach

¹ www.mturk.com

² www.faircrowd.work

to speed up the ranking, by crowdsourcing the evaluation of crowdsourcing platforms using community heuristics.

Related Work

Our system draws inspiration from Crowdcrit, a system designed to allow crowdworkers to give high quality critique on visual designs (Heer & Bostock, 2010). The system allows non-experts (crowdworkers) to give expert feedback on designers by choosing critique from a list of the most common critiques. Our demo uses crowdsourcing to evaluate the social aspect of crowdsourcing platform, thus using non-experts to perform a complex evaluation.

Our previous work has gathered and adjusted online community heuristics to evaluate crowdsourcing platforms. By having a vibrant, active community, crowdsourcing platforms can potential benefit in many ways: loyalty to and continued use of the platform, more collaboration and increased trust. The evaluation of the platforms is a labour intensive task, taking more than an hour per platform. Therefore, we split up these evaluations into smaller tasks which we crowdsource using our demo.

Demo

To test if the crowdsourcing of complex evaluations is possible, we build a simple survey-like platform using Python and Django. We recruited crowdworkers using Microworkers.com³ which is a more user and worker friendly platform than the more popular MTurk.

The crowdworkers are asked to visit a crowdsourcing platform and fill in questions about the design and functionality of the platform. They are asked to create a profile to get access to the entire platform. The crowdworkers have to check a box, declaring they have visited the platform for around 5 minutes. The crowdworkers provide simple demographics; date of birth and gender. Then they can evaluate the community heuristic (Figure 1). The platform states the name of the heuristic followed by a statement about the heuristic e.g. “the purpose of the platform is clear”. Under the statement is extra information to guide the evaluator e.g. “does the platform clearly describe how it fulfils the member’s and platform owner’s goals”. The crowdworkers give a score either of “strongly disagree, disagree, neutral, agree or strongly agree” with the given statement, followed by an explanation of their choice. Last, we ask the crowdworkers to provide a

³ www.microworkers.com

Evaluate Jovoto

Purpose

Clarity

The purpose of the platform is clear

Does the platform clearly describes how it fulfils the members' needs and the platform owner's goals.

Score*

Strongly Disagree Disagree Neutral Agree Strongly Agree Unable to Answer

Explanation*

Explain your choice by providing evidence, where did you find the information. If you didn't find it, where would you expect it:

Screen shot*

Support your explanation with a screenshot link using [Snag.gy](#). Copy the generated url into the field

Back

Page 3 of 5

Continue

Figure 1. Screenshot of the demo asking crowdworkers to evaluate the community heuristic “clarity” of the crowdsourcing platform Jovoto. The crowdworker gives a score to show in what way they agree with the platform following the community heuristic. They are required to give a short explanation among with a screenshot of the platform supporting their explanation.

screenshot supporting their explanation. These help us to understand where on the platform, the crowdworkers look for evidence that adhere to the heuristic.

At the end of the evaluation, the crowdworkers are given an unique code that they can fill in on Microworkers.com. We manually match the data from our platform and the task on Microworkers.com to pay the crowdworkers.

In a pilot study, we asked 10 English speaking crowdworkers to evaluate the visibility and clarity heuristic of the platform Jovoto. The crowdworkers received \$1 for successfully completing the evaluation which would take a maximum of 15 minutes. A second pilot study with again 10 English speaking crowdworkers evaluated only the “life cycle” heuristic and were offered \$0,50. One can access the survey that the Microworkers used to evaluate the platform Jovoto at: <http://tinyurl.com/ecscwdemo>

Preliminary data

Most of the crowdworkers (60%) were female for the two studies. The age ranged between 19 and 35 years. The majority (70%) provided their email

address to perform evaluations in the future. All of the evaluators provided evidence for their explanation of which two used a direct link to the platform and the rest a screenshot link. For the heuristic “clarity”, only one person gave the score “Agree” and the rest (9 crowdworkers) chose “Neutral”. The scoring for the heuristic “visibility” were more diverse: 1x “Disagree”, 3x “Neutral” and 6x “Agree”. The average amount of words used for explanation of the heuristic “clarity” is 32,5 and 20 words for “visibility”. One of the better evaluations of the “clarity” heuristic noted “*The website gives pretty concise clear introductions on the https://www.jovoto.com/creatives/ page regarding how the site will benefit the creative side of the platform. On the https://www.jovoto.com/how-it-works/ it gives detailed information for the brand side of the platform. There is information on both pages as to what the owner's goals are but they aren't as clear and laid out as the other sections.*” And an example of an extensive visibility heuristic explanation: “*The purpose is made known to the visitor with the first sentence seen on the page 'Transform your products...' This gives a general impression of what the platform does. Scrolling down a little and the visitor can get a clearer idea of what it does with short phrases explaining the services provided.*” Two evaluators were inspired to investigate the evaluated platform further for their own use: “*I will spend more time on the site working on my profile, etc. after I have completed this survey.*”

The lesser paid evaluation of the “life cycle” heuristic, which contained only one question, had a more mixed quality of evaluations. ”. We declined two of the evaluations, since they didn't provide any sign of having performed an evaluation “*I think it is a great idea and I have seen it in online*”. One of the evaluators provided a relatively long 140 word explanation containing detailed information of how the platform could be improved “*...but a few examples of potential earnings on the How It Works page would be very useful*”. The scoring for the life cycle was: 6x “Neutral” and 4x “Agree”. The average words used for the explanation is 39 words (without the extensive explanation, 27 words).

Future work

The preliminary data looks promising and we will continue to study if the crowdsourcing of a complete evaluation is possible. The crowdsourcing of evaluations would allow us to create a listing and ranking of crowdsourcing platforms. Using this comparing platform for crowdsourcing platforms, workers and requesters could choose the best platform, fitting their needs. At the same time, platforms would like to increase their scoring to generate more traffic to their platform. Although we now evaluate the social aspect of platforms, more elements of the platform should be evaluated for the listing. The platform itself could become a community of crowdsourcing workers, helping to improve the online working environment.

References

- Brabham, D. C. (2010). Moving the Crowd at Threadless: Motivations for participation in a crowdsourcing application. *Information, Communication & Society*, 13(8), 1122–1145. <https://doi.org/10.1080/13691181003624090>
- Crowdsourcing.org:80/directory. (n.d.). Retrieved June 11, 2016, from <https://web.archive.org/web/20160611023436/http://www.crowdsourcing.org:80/directory>
- Gray, M. L., Suri, S., Ali, S. S., & Kulkarni, D. (2016). The Crowd is a Collaborative Network. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16* (pp. 134–147). New York, New York, USA: ACM Press. <https://doi.org/10.1145/2818048.2819942>
- Heer, J., & Bostock, M. (2010). Crowdsourcing graphical perception. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10* (p. 203). New York, New York, USA: ACM Press. <https://doi.org/10.1145/1753326.1753357>
- Kittur, A., Nickerson, J., & Bernstein, M. (2013). The Future of Crowd Work. *Proc. CSCW '13*, 1–17. <https://doi.org/10.1145/2441776.2441923>
- Soliman, W., & Tuunainen, V. K. (2015). Understanding continued use of crowdsourcing systems: An interpretive study. *Journal of Theoretical and Applied Electronic Commerce Research*, 10(1), 1–18. <https://doi.org/10.4067/S0718-18762015000100002>
- Stewart, O., Huerta, J. M., & Sader, M. (2009). Designing crowdsourcing community for the enterprise. *Proceedings of the ACM SIGKDD Workshop on Human Computation - HCOMP '09*, 50. <https://doi.org/10.1145/1600150.1600168>
- Zheng, H., Li, D., & Hou, W. (2011). Task Design, Motivation, and Participation in Crowdsourcing Contests. *International Journal of Electronic CommerceOnline) Journal International Journal of Electronic Commerce*, 15(4), 1086–4415. <https://doi.org/10.2753/JEC1086-4415150402>