Toward Effective Tasks Navigation in Crowdsourcing

Pavel Kucherbaev  
University of Trento  
Via Sommarive 5, Povo (TN)  
38123, Italy  
pavel.kucherbaev@unitn.it

Florian Daniel  
University of Trento  
Via Sommarive 5, Povo (TN)  
38123, Italy  
florian.daniel@unitn.it

Maurizio Marchese  
University of Trento  
Via Sommarive 5, Povo (TN)  
38123, Italy  
maurizio.marchese@unitn.it

Fabio Casati  
University of Trento  
Via Sommarive 5, Povo (TN)  
38123, Italy  
fabio.casati@unitn.it

Brian Reavey  
CrowdFlower  
2111 Mission Street, San Francisco (CA) 94110, USA  
brian.reavey@crowdflower.com

ABSTRACT
Crowdsourcing platforms are changing the way people can work and earn money. The population of workers on crowdsourcing platforms already counts millions and keeps growing. Workers on these platforms face several usability challenges, which we identify in this work by running two surveys on the CrowdFlower platform. Our surveys show that the majority of workers spend more than 25% of their time on searching tasks to work on. Limitations in the current user interface of the task listing page prevent workers from focusing more on the execution. In this work we present an attempt to design and implement a specific user interface for task listing aimed to help workers spend less time searching for tasks and thus navigate among them more easily.

Categories and Subject Descriptors
H.5.3 [Group and Organization Interfaces]: Web-based interaction

General Terms
Human Factors, Design, Performance

Keywords
Crowdsourcing, User Interfaces, Search

1. INTRODUCTION
Crowdsourcing is the practice of outsourcing work to an unknown group of people via the Internet, instead of assigning it to internal employees [9]. Crowdsourcing has been so far very successful in performing tasks which are still hard to automate using algorithms, while they can be relatively easily solved by humans. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

AVI’ 14, May 27 - 29 2014, Como, Italy  
Copyright 2014 ACM 978-1-4503-2775-6/14/05 $15.00  
http://dx.doi.org/10.1145/2598153.2602249

such as image object recognition, annotations, feedback collection, and similar.

Requests are the people who want to crowdsource their work. They publish tasks on crowdsourcing platforms where requestors meet potential workers - people who solve tasks for monetary reward, curiosity or other motivations. Some examples of crowdsourcing platforms are Amazon Mechanical Turk (MTurk), CrowdCloud, MicroWorkers, Mobileworks, CrowdFlower. There are several millions of workers currently involved in crowdsourcing, and there are thousands of tasks available to work on. These tasks are typically not well structured and descriptive, so workers spend a significant amount of time on searching tasks to perform [1, 8]. There are many short tasks which require about 2 minutes to work on, and spending another 2 minutes to find it is obviously not efficient.

In this work we aim to analyze the way workers search for tasks and we make an attempt to design an effective user interface for a task listing page that helps workers to find and navigate among available tasks easier and faster.

2. STATE OF THE ART
Chilton et al. [1] ran a survey on MTurk with 250 workers about the way they search for tasks. The researchers found that on a large scale workers sort tasks to see the recently posted ones or to see those with the most instances available. Workers look generally at the first page of the most recently posted tasks and the first two pages of the tasks with the most available instances. Chilton et al. discovered, that a favorable position in the search results (whether a task is on the first two pages) does matter: a task with favorable positioning was completed 30 times faster and for less money than when its position was unfavorable.

In [6], Man-Ching Yuen et al. conducted a survey on MTurk with 100 workers involved, where they found that 65% of workers prefer to select tasks similar to the ones they have done before. Also 67% of workers do not prefer to select tasks similar to those for which their work was rejected (rejection functionality is available on MTurk; if people have their work rejected, they do not get paid). The authors found that one of the main selection criteria are reward amount and task nature. Some workers prefer to perform research tasks where they need to find required information and write text, while others prefer to perform decision making tasks, such as to approve a picture content or to pick the best description for a given
Ipeirotis [4] proposed to improve the workers’ experience via (i) building a browsing system that allows workers to split all the tasks in categories and let workers navigate among them; (ii) improving the search engine to include all the task fields into the indexing algorithm; and (iii) using a recommender system to propose tasks to workers.

Current research does not explain how workers search for tasks on other platforms than MTurk, how different task listing user interfaces affect workers’ productivity and workers’ overall satisfaction with the crowdsourcing platform.

3. HOW WORKERS NAVIGATE TASKS

We first tried to understand how much time workers spend searching for tasks to work on, if there is any difference in tasks searching behavior between workers, and whether searching is a critical problem for workers. We decided to conduct two surveys in order to answer these questions.

3.1 Survey I

With the first survey we had the goal to identify how much time workers spend searching for tasks to work on and to collect some feedback from workers to get an understanding of the problem. This survey included 6 multiple choice questions:

1. How long have you been on this microtask platform?
2. How much time (approximately on average) do you usually spend on searching the next task to work on?
3. How much time (approximately on average) do you usually spend on executing one task?
4. What is your approximate ratio (searching time/ execution time)?
5. Are you fine with your ratio (searching time/ execution time)?
6. Would you like to focus more on execution and less on searching?

Some of these questions aim to validate the answers on other questions. For instance, question 4 validates the answers to questions 2 and 3. Question 5 makes a worker think about the answer to the previous question. Questions 6 aims to validate the answer to question 5 and to get a worker’s opinion. This question is biased and therefore stays in the very end of the questionnaire.

This survey was conducted as a task on CrowdFlower [1] with a reward of $0.05, and published to all the channels suggested by the CrowdFlower platform by default (CrowdFlower is a meta-platform which publishes tasks on other crowdsourcing platforms, such as MTurk and other rewards websites, such as Neobux). We requested and collected 500 responses for this survey.

Results.

We collected all the results in 2 days. All the responses were from the Unites States. The demographics information about workers that participated in the survey, such as sex, age, education, is not available, as it is not provided by CrowdFlower. The dataset with responses is publicly available for further analysis [2].

These results show that 38% of all workers spend less than 12.5% of all their time on the platform for searching tasks, while 42% of workers spend more than 25% (Figure 1a). From the collected data we calculated that on average workers spend about 27% of their time searching for tasks to work on. About 33% of all workers spend 1 - 2 minutes searching a new task to execute, while 24% of workers spend more than 5 minutes (Figure 1b). Figure 2 shows that workers with different levels of experience tend to spend different amounts of time on searching tasks. People who have been a member of a platform for less than one month (29%) spend about half of their time searching for tasks to execute, while workers with 2 - 5 months of experience spend 25% of their time. On crowdsourcing platforms a task can have many instances. About a third of the most experienced workers (more than 3 years on a platform) spend almost all the time on execution, most likely because they focus on tasks with many available instances to minimize the amount of time they spend on searching.

Figure 1: a) Part of the time workers spend on searching tasks, b) Time workers spend on searching a new task to execute

Workers spend a similar amount of time searching for tasks regardless of their preferred platform. In Figure 3 we show a distribution of searching time between workers from MTurk, FusionCash and Instagc (only platforms with more than 25 responses for the survey are presented).

Figure 2: Distribution of the searching time by experience

http://crowdflower.com
http://bit.ly/1hDqL1K
CrowdFlower. This survey consisted of three parts: a second survey. This survey was conducted in collaboration with
are the possible solutions to this problem, we decided to conduct
searching for tasks is a significant problem and to understand what
for some workers. Moreover, workers would specifically like the ability to
filter specific tasks to perform, blocking tasks for which they are
not eligible and tasks they have already completed.

3.2 Survey II

From the first survey we learned that searching is a problem for some workers. In order to understand for how many workers
searching tasks to work on is a problem for about 2/3 of all workers,
Discussion.

Workers spend a significant amount of time (25% - 50%) on
searching tasks to execute. From the feedback collected from the
survey we identified that searching tasks is a problem for some
workers. Moreover, workers would specifically like the ability to
filter specific tasks to perform, blocking tasks for which they are
not eligible and tasks they have already completed.

We assumed that workers from different countries might answer
differently on this survey, so we decided to conduct this survey in
three different regions: 1) USA, 2) Western Europe, 3) Southeast
Asia. In each region we requested 250 responses, rewarding $0.10
for each response.

Results.

After collecting the results (Table 1), we filtered responses pro-
vided by bots or by people who clearly spammed the survey. All
results were manually analyzed, and responses containing inappro-
priate data were removed, such as those answering with “5000” the
question “How many hours per week do you spend on crowdsourc-
ing platforms”. These responses (14% from the USA, 18% from
Europe and 3.6% from Asia) were removed from the analysis.

The main problems, which workers face are: 1) searching for
tasks, 2) working on poorly designed tasks, 3) rejection of provided
work, 4) tasks with few available instances, 5) slow or inadequate
responses to worker support inquiries. In Table we can also see
the distribution of worker preferences for proposed solutions. In
column “All” the percentage out of all workers who voted for a so-
lution is presented, in column “Critical” we report the percentage
of workers for whom searching is a critical problem. The major-
ity of workers prefers “Ranked keyword search”, “Categories” and
“Subscription box”.

Discussion.

From the data collected in the second survey we identified that
searching tasks to work on is a problem for about 2/3 of all workers,
confirming the findings of the first survey.

CrowdFlower has a community of workers, called CrowdLab,
with which CrowdFlower employees interact in an online chat room.
After analyzing the survey responses and discussing with Crowd-
Lab, we have identified that workers want to have more control over
the task selection process. Unfortunately often there are not many
tasks available (especially for workers outside the USA), so work-
ers want to have a task list as big as possible with different filtering
and sorting tools. Because of this, workers are skeptical about the
efficiency of an online-radio option. Workers have an issue with se-
lecting a task to work on even when there are only around 30 tasks
available because they feel the information displayed about each task usually is not descriptive enough, and to understand whether a task is worth working on they must actually try to complete it.

4. TASKS LISTING PAGE REDESIGN

We started to work on the redesign of the task listing page (Fig-ure 4). From the interaction with the CrowdLab community we identified that it is hard for workers to figure out whether a task is worth working on or not. TurkOpticon [5] shows ratings of requestors on the Mturk task listing page, which helps workers to make a decision about a task. Workers on CrowdFlower have an option to complete an exit survey after they are done with a task. We decided to show the average overall satisfaction level (on a scale from 1 to 5), as we think that this can give workers some feeling about the quality of a task before they actually work on it. For instance, if a task is evaluated with a score above 4.0, it is a good indication that the task is well designed and provides a fair reward.

Workers have problems with filtering tasks on CrowdFlower, and thus we decided to show to which category each task belongs. These categories are selected by requestors when they design tasks. Unfortunately it is an optional field and about 60% of tasks are not categorized, so we used the TF-IDF algorithm [2] to calculate a similarity level with a training set of 20,000 task titles.

The new design of the user interface has a set of additional features: new tasks appear on the listing page in realtime without reloading the page; workers can sort, filter and search tasks using different criteria; the average satisfaction level, taken from the exit surveys is displayed. The task listing page user interface for CrowdFlower has already passed several iterations (presented here [http://codesign.io/ubcswp](http://codesign.io/ubcswp)) with the CrowdLab community in order to satisfy their needs. The main concern was about making the user interface clean enough to make it comfortable for the eyes. Workers were positive about having the average satisfaction level, as it gave a meaning to exit surveys they complete in the end of tasks. They were also positive about an ability to filter tasks by requestor type (“admin” - CrowdFlower employees, “pro” - enterprise accounts, and “basic” - all other accounts), as the fact, that a task is published by an “admin” requestor, tells that this task has high quality in terms of user interface and reward amount.

5. CONCLUSION

We have analyzed the way workers interact with the task listing page on the CrowdFlower platform by conducting two surveys. We identified a set of issues that workers face: spending a significant amount of time (25% - 50%) searching for tasks to work on, insufficient information about tasks, lack of task categorization, and lack of sort/filter/search functionality. Having in mind this set of issues, we have designed and implemented a first prototype of a new user interface for the task listing page for CrowdFlower. At present the new interface passed through a set of iterations with feedback from the closed CrowdLab community of workers. This is an ongoing project and we plan to conduct a usability study with a broader set of workers in order to evaluate the current user interface along with testing some other techniques, such as suggesting tasks to workers using recommender systems and various ways of describing these recommendations.

6. REFERENCES


<table>
<thead>
<tr>
<th>Not trustful responses, %</th>
<th>USA</th>
<th>Europe</th>
<th>Asia</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>14 (35/250)</td>
<td>18 (45/250)</td>
<td>3.6 (9/250)</td>
</tr>
<tr>
<td>Main source of income, %</td>
<td>47</td>
<td>37</td>
<td>61</td>
</tr>
<tr>
<td>Searching is a critical problem, %</td>
<td>35</td>
<td>31</td>
<td>31</td>
</tr>
<tr>
<td>Searching is a problem, but not critical, %</td>
<td>29</td>
<td>42</td>
<td>38</td>
</tr>
<tr>
<td>Searching is not a problem, %</td>
<td>36</td>
<td>27</td>
<td>31</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ranked Keyword Search, %</th>
<th>Critical</th>
<th>All</th>
<th>Critical</th>
<th>All</th>
<th>Critical</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30</td>
<td>24.6</td>
<td>27</td>
<td>19</td>
<td>24</td>
<td>28</td>
</tr>
<tr>
<td>Suggestion Box, %</td>
<td>19</td>
<td>20</td>
<td>25</td>
<td>23.4</td>
<td>24</td>
<td>23.6</td>
</tr>
<tr>
<td>Categories, %</td>
<td>18</td>
<td>21.8</td>
<td>20</td>
<td>31.7</td>
<td>21</td>
<td>24</td>
</tr>
<tr>
<td>Subscription, %</td>
<td>16.7</td>
<td>11.6</td>
<td>11.1</td>
<td>10.2</td>
<td>16</td>
<td>10.4</td>
</tr>
<tr>
<td>Radio, %</td>
<td>8.3</td>
<td>13</td>
<td>6.3</td>
<td>7.3</td>
<td>6.7</td>
<td>4.5</td>
</tr>
<tr>
<td>None, %</td>
<td>6.9</td>
<td>8.8</td>
<td>9.5</td>
<td>8.3</td>
<td>8</td>
<td>8.3</td>
</tr>
</tbody>
</table>

Table 1: Distribution of workers preferences among proposed solutions