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 [3]. 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,

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

Requestors 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), Crowd-Cloud, 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, 6]. 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

product.

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 availably for further analysis[2].

These results show that 38% of all the workers spend less than 12.5% of all their time on the platform for searching tasks, while
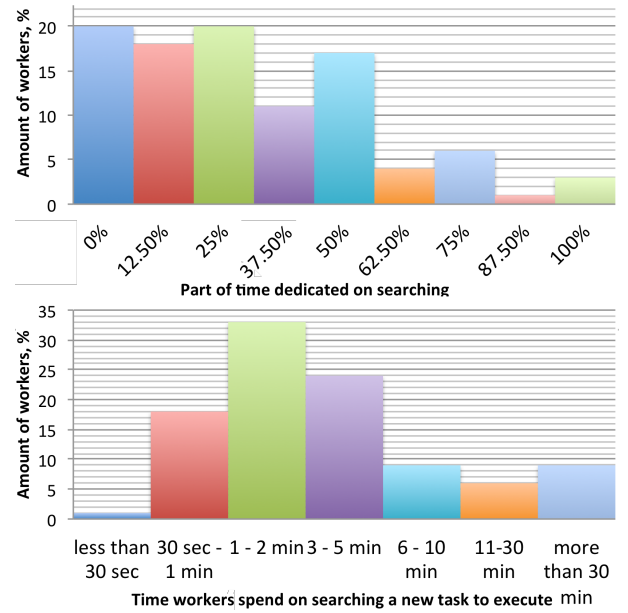
Figure 1: a) Part of the time workers spend on searching tasks, b) Time workers spend on searching a new task to execute

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.
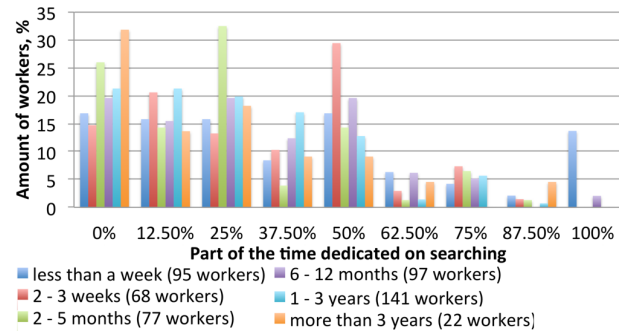

Figure 2: Distribution of the searching time by experience

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).
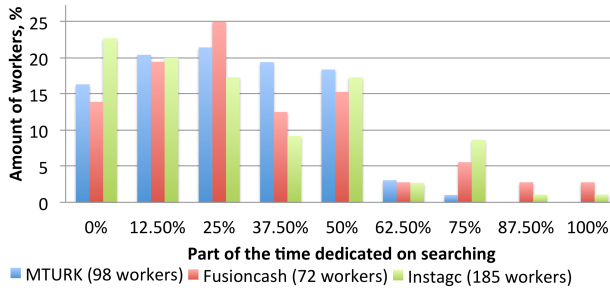
Figure 3: Distribution of the searching time by crowdsourcing platform

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.

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 for tasks is a significant problem and to understand what are the possible solutions to this problem, we decided to conduct a second survey. This survey was conducted in collaboration with CrowdFlower. This survey consisted of three parts:

1. *Questions to understand whether searching is a problem*: Is searching appropriate tasks to perform a problem for you? (if yes) Is this problem critical (a primary) to you? (if no) What is the biggest problem/issue that you face on crowdsourcing platforms/reward sites?

2. *Questions to understand which solutions from a proposed list are most appropriate*: Please pick a solution you think is the best (or none); Please describe in words: How do you understand the solution (described below) you have selected? What are the positive consequences of applying such a solution? What are the negative consequences of applying such a solution?

3. *Questions to get more information about a worker answering this survey*: Is performing tasks on crowdsourcing platforms/reward sites a main source of income for you? How many hours per week do you spend on crowdsourcing platforms/reward sites? Do you use several platforms/sites for performing tasks? (if yes) On which platforms/sites do you perform tasks? (if yes) If the solution you selected is implemented will you be more likely to stay only on the platform/site where it is implemented? If you have any additional feedback please leave it here.

The list of possible solutions to decrease the searching time was defined together with the research department of CrowdFlower, and included the following options:

- *Ranked Keyword Search* - a regular search, which generates a list of tasks, ranked according to the relevance to the worker's query. The search considers task title, task description, task keywords and requestor name;

- *Suggestion box* - similar to what Amazon.com provides for suggesting relevant products. After workers complete a task, they see a suggestion of several tasks relevant to the previous one, following the logic: "people who performed this task also performed these tasks";

- *Radio* - here we want to make an experience of working on crowdsourcing platforms similar to listening to online radio stations, such as Pandora or Spotify. Listeners select a song they want to listen to and the system creates a playlist of relevant songs. Workers pick a task to work on and every time they finish a task the system proposes a similar one, allowing the worker to flag each task with "like" or "dislike", allowing them to skip those they "dislike";

- *Subscription* - a worker subscribes to favorite requestors or task keywords. With a defined period (hour/day/week) this worker receives a notification message (email, tweet or sms), or if in push mode, then the worker receives a message every time a new task appears, which satisfies their defined parameters;

- *Categories* - all tasks are classified into a predefined set of categories, such as decision making tasks and content generation tasks.

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 provided by bots or by people who clearly spammed the survey. All results were manually analyzed, and responses containing inappropriate data were removed, such as those answering with "5000" the question "How many hours per week do you spend on crowdsourcing platforms". These responses (14% from the USA, 18% from Europe and 3.6% from Asia) were removed form 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 1 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 solution is presented, in column "Critical" we report the percentage of workers for whom searching is a critical problem. The majority 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 workers 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 selecting a task to work on even when there are only around 30 tasks

	USA		Europe		Asia	
Not trustful responses, %	14 (35/250)		18 (45/250)		3.6 (9/250)	
Main source of income, %	47		37		61	
Searching is a critical problem, %	35		31		31	
Searching is a problem, but not critical, %	29		42		38	
Searching is not a problem, %	36		27		31	
	Critical	All	Critical	All	Critical	All
Ranked Keyword Search, %	30	24.6	27	19	24	28
Suggestion Box, %	19	20	25	23.4	24	23.6
Categories, %	18	21.8	20	31.7	21	24
Subscription, %	16.7	11.6	11.1	10.2	16	10.4
Radio, %	8.3	13	6.3	7.3	6.7	4.5
None, %	6.9	8.8	9.5	8.3	8	8.3

Table 1: Distribution of workers preferences among proposed solutions

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 (Figure 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.
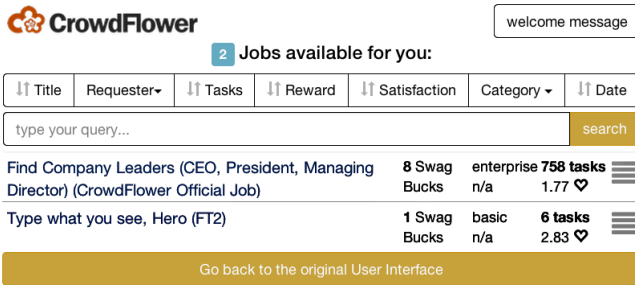

Figure 4: User Interface developed at CrowdFlower

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) with the CrowdLab commu-

nity, 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

[1] L. B. Chilton, J. J. Horton, R. C. Miller, and S. Azenkot. Task search in a human computation market. HCOMP '10, pages 1–9, New York, NY, USA, 2010. ACM.

[2] D. Hiemstra. A probabilistic justification for using tf-idf term weighting in information retrieval. *International Journal on Digital Libraries*, 3(2):131–139, 2000.

[3] J. Howe. The rise of crowdsourcing. *Wired*, 14(14):1–7, October 2006.

[4] P. G. Ipeirotis. A plea to amazon: Fix mechanical turk! http://bit.ly/GRvrAn.

[5] L. C. Irani and M. S. Silberman. Turkopticon: Interrupting worker invisibility in amazon mechanical turk. CHI '13, pages 611–620, New York, NY, USA, 2013. ACM.

[6] M.-C. Yuen, I. King, and K.-S. Leung. Task recommendation in crowdsourcing systems. CrowdKDD '12, pages 22–26, New York, NY, USA, 2012. ACM.