

## Crowd4U: An Initiative for Constructing an Open Academic Crowdsourcing Network

**Atsuyuki Morishima**  
University of Tsukuba  
Ibaraki, Japan

**Sihem Amer-Yahia**  
CNRS, LIG  
Grenoble, France

**Senjuti Basu Roy**  
University of Washington Tacoma  
Tacoma, USA

### Abstract

We describe the Crowd4U initiative, which aims at constructing an all-academic open and generic platform for microvolunteering and crowdsourcing worldwide. Crowd4U provides a microtask-based platform in which most workers are volunteers at universities and other research institutions. Crowd4U is open in the sense that the platform can interact with other platforms, researchers can register their tasks, and the underlying code is not a black box. It is generic as it allows to register virtually any task. Crowd4U has already been used by several projects for public and academic purposes.

### Introduction

Crowdsourcing proved to be a promising approach for involving humans in completing tasks that are difficult to a computer. However, how to unleash its full potential is still in the phase of exploration by different research communities. Currently, many research projects and academic crowdsourcing projects use commercial platforms, such as Amazon Mechanical Turk and CrowdFlower, as black boxes by using externally provided functionalities. This paper describes the Crowd4U initiative, which aims at constructing an all-academic open and generic platform for microvolunteering and crowdsourcing worldwide. The initiative has been building a microtask-based crowdsourcing platform that is similar to AMT with many advanced features. Most workers are university members including students, and staff and faculty members. Since they perform tasks voluntarily, we call them *contributors*. Figure 1 shows some statistics. Crowd4U plays three major roles:

**A support role as a platform for crowdsourcing projects with academic and public purposes.** We intend Crowd4U to be used for academic and public purposes. Crowd4U already hosts several non-commercial crowdsourcing projects. For example, L-Crowd is a project in the context of libraries. Currently, L-Crowd provides microtasks to identify different books that have the same ISBN in an effort to clean the bibliography database of the National Diet Library (Morishima et al. 2013). Another example is the CrowdHealth project which aims at revealing nutrition and

#Registered tasks	34, 275
#Registered contributors	462
#Countries of contributors	12
#Universities of collaborators	24
#Projects	7
Average throughput per day	1015 (fourth week of May, 2014)

Figure 1: Some statistics on Crowd4U as of July 24, 2014. Crowd4U does not require registration to perform tasks. The estimated number of anonymous contributors exceeds 2000.

health information hidden in tweets to improve quality of life. CrowdHealth uses the crowd to label tweets that are related to nutrition and health and then learn a classifier for automatic tweet annotation.

**A research role as a platform for advancing the science of crowdsourcing.** Crowd4U allows us to conduct research that requires implementation at the platform level. For example, we have implemented a declarative framework that dynamically generates tasks and assigns them to appropriate contributors during the execution (Morishima et al. 2012). We also plan to test task-to-contributor assignment algorithms (Roy et al. 2013). Yet another project uses the history of performed tasks to learn rich contributor profiles.

**A social role as an infrastructure for addressing emergencies.** One of the important purposes of Crowd4U is to construct a worldwide infrastructure that can be used to help each other when necessary. For example, the Microvolunteering in Natural Disaster (MIND) project aims to develop efficient crowdsourcing algorithms to handle natural disasters such as tornados and tsunamis. An example of microtask in MIND is the identification of damaged buildings that will serve to train image recognition algorithms. Such a task helps build a model to generate survey damages at scale in the aftermath of a disaster.

### Technical Features

Several features make Crowd4U unique and attractive.

**Complex crowdsourcing.** Crowd4U supports a high-level abstraction for implementing different but inter-related microtasks to implement complex logics as a whole. The language, named CyLog, is based on Datalog and can be used to express complex dataflows with predicates that are evaluated by humans. This feature allows developers to im-



Figure 2: The task-on-the-floor system, a Crowd4U tablet in a library, and the screenlock application for smartphones

plement interesting tasks easily. For example, five kinds of microtasks were combined to infer the path of a tornado in a disaster-related project. Another example is *dynamic task-translation tasks* that ask contributors to translate task descriptions that appear in another set of tasks. The translation results are immediately and automatically passed to the tasks and directed to contributors who speak different languages. Yet another example is *collaborative microtasks* that show to contributors the latest results by other contributors and ask them to perform tasks considering the results so far.

**Incentive schemes.** Crowd4U can serve as a testbed for various incentive schemes. Several incentive schemes are already available. For example, some contributors are university students who perform microtasks (voluntarily!) when they try to download PDF files used in courses. Some of them are driven by gamification mechanisms (e.g., ranking of their contributions). Other incentives include platforms on which Crowd4U is accessible (Figure 2). It is possible to perform microtasks by using “Crowd4U terminals” located on campus of seven universities in Japan, France and Belgium. It is also possible to use a *task-on-the-floor* system placed at three universities in Japan. Using that system, contributors perform tasks projected on the floor while walking. Finally, tasks are available as smartphone screenlocks and contributors can perform a task when they unlock the screen of their smartphones.

**Openness.** First, Crowd4U is open to other platforms in the sense that it has an http-based API to exchange tokens and receive data from other platforms. This way, Crowd4U can be used both for recruiting contributors and for implementing complex crowdsourcing applications. Second, Crowd4U has a form-based RAD (rapid application development) tool for researchers to easily create microtasks. This helps implement frequently-used patterns of crowdsourcing, such as majority voting and translation. And finally, the engine is not a black box: although the code is not an open-source at present for the security reason, the code is given to researchers if they use it for their research.

**Contributor selection and dynamic assignment.** Crowd4U allows requesters to select contributors to perform a particular set of tasks based on their attributes (e.g., spoken language), and the results of microtasks so far, and dynamically distribute tasks to eligible contributors. In addition, requesters can specify the *order* in which tasks are performed, so that dynamically generated tasks are automatically sorted for task assignment.

## Challenges

In addition to general challenges in this area, such as data quality, acquisition of human computation resources and the

effective design of microtasks, Crowd4U has its own challenges as the purpose is to develop an open and generic platform for researchers worldwide.

**Task assignment.** Many crowdsourcing platforms solve the task assignment problem by providing the market mechanism where each task has a reward and workers self-appoint themselves to perform tasks. In contrast, Crowd4U supports various task assignment strategies and incentive schemes. While contributors can choose their favorite tasks from a list of available tasks, Crowd4U also supports a push-style assignment such as the task-on-the-floor system. One of our challenges is to choose the right task assignment by considering *contributor-centric or requester-centric constraints*.

The CyLog language can be used to express complex conditions that make use of a plethora of information from previously performed tasks. For example, results of tasks contributors performed in the past could be used to create and maintain contributor profiles that could be used in clever task assignment. A challenge is to develop a clever contributor-selection mechanism that exploits such data.

**Task adaptation.** Different task distribution schemes, such as the task-on-the-floor systems, smartphone unlocking applications, and dedicated tablet terminals, have different characteristics. For example, tasks performed on PC screens are appropriate for relatively complex and time-consuming tasks, while tasks performed in the task-on-the-floor systems must be easy enough to be performed in a second. An interesting challenge is to adapt execution plans to the available resources in the different schemes.

**International collaboration.** Crowd4U contributors are from a variety of countries who are diverse in languages, culture and time zones. How to design a framework to adapt and distribute tasks to those contributors is an important challenge. Currently, a solution taken by some projects is to use dynamic task-translation tasks. In some tasks that require subjective answers, we may need to consider cultural differences. Gamification for players with different time zones is also a challenge.

**Optimization in complex crowdsourcing.** Some of the functions such as task-translation tasks are implemented using the ability of Crowd4U’s complex crowdsourcing, in which results of tasks are used to generate other sets of tasks. It is not straightforward to optimize such complex crowdsourcing applications that require complex data and control flows involving human activities.

## Acknowledgments

We are grateful to contributors partly listed at crowd4u.org. This work is supported by PRESTO from JST and the Grant-in-Aid for Scientific Research (#25240012) from MEXT.

## References

- Morishima, A., et al. 2012. CyLog/Crowd4U: A Declarative Platform for Complex Data-centric Crowdsourcing. PVLDB 5(12): 1918-1921.
- Roy, S., et al. 2014. Optimization in Knowledge-Intensive Crowdsourcing. CoRR abs/1401.1302.
- Morishima, A., et al. 2013. L-Crowd: A Library Crowdsourcing Project by LIS and CS Researchers in Japan. ICDL, 40-47.