Curio: A Platform for Supporting Mixed-Expertise Crowdsourcing

Edith Law, Conner Dalton, Nick Merrill, Albert Young, Krzysztof Z. Gajos
School of Engineering and Applied Sciences
Harvard University
Cambridge, MA

Abstract

Curio is a platform that enables researchers in the sciences and humanities, who are domain experts but not necessarily technically savvy, to create, monitor and control complex crowdsourcing projects with minimal effort. With Curio, we aim to contribute to the practice of citizen science and to make fundamental contributions to the study of human computation by developing new interfaces and algorithms for supporting mixed-expertise crowdsourcing, task decomposition, incentive design and quality control.

Science is data intensive. In order to test hypotheses about our natural environment, e.g., about climate patterns, species distribution and trajectories of stars, often we need to collect and analyze data over large geographical regions and many time periods. When carried out by a few scientists, this process is tedious, time-consuming and sometimes impossible. The idea of citizen science is to engage non-scientists in the collection and interpretation of data in order to answer scientific questions. Several large-scale citizen science projects exist today; Zooinverse (`http://zooniverse.org`), for example, hosts several carefully crafted projects engaging people to process data from a variety of domains, including medical images, museum records, oceanography data and archaeological artifacts.

Despite the success of these citizen science platforms, the use of crowdsourcing as a tool for science is not widespread – based on a set of 18 interviews conducted with researchers in the natural, social and medical sciences, we found that the vast majority of the researchers were unfamiliar with or reluctant to use crowdsourcing technology in their research. To date, the barrier to entry has been substantial: researchers must either engage in a non-trivial proposal and review process to launch a project with Zooniverse, build their own website and user base from scratch, or use paid crowdsourcing platforms such as Mechanical Turk, which requires a fair degree of technical sophistication to use effectively.

We introduce Curio (`http://crowdcurio.com`), a platform for crowdsourcing research tasks in the sciences and humanities. Curio is designed to allow researchers, who are domain experts but not necessarily technically savvy or

familiar with crowdsourcing, to create and launch a new crowdsourcing project within minutes and to monitor and control aspects of the crowdsourcing process with minimal effort. Curio will support an ecosystem of crowdsourcing projects, exposing participants to and drawing on their curiosity to help answer research questions in a wide variety of domains. There are a few existing examples of this new genre of DIY (do-it-yourself) citizen science marketplace, e.g., CrowdCrafting.org which enables "researchers, civic hackers, and developers" to create their own projects. Those existing platforms tend to be limited to simple annotation and classification tasks powered by simple workflows.

With Curio, in contrast, we are exploring a nascent model of citizen science, developing new interfaces and algorithms for supporting mixed-expertise crowdsourcing, and investigating a variety of human computation questions related to task decomposition, incentive design and quality control. By *mixed-expertise crowdsourcing*, we mean a crowdsourcing paradigm involving both amateurs and expert workers or requesters (e.g., scientists) whose expertise typically takes years of formal training to build. This paradigm raises interesting new research questions in human computation: How can we enable experts to create robust workflows for complex tasks that non-experts can contribute to? How can we maximize the little time that experts have to drastically improve overall crowdsourcing outcomes? How do we enable experts who are unfamiliar with crowdsourcing to understand and debug crowdsourcing results? How do we close the social gap between researchers and participants, allowing experts to trust novice contributions, and participants to perceive scientists as personable and collaborative?

In this paper, we discuss some of the key features of Curio for supporting mixed-expertise crowdsourcing. During the conference, we will demonstrate Curio by having the attendees play the roles of both workers and requesters, performing tasks for in a variety of Curio projects and creating their own crowdsourcing project on the platform.

Introducing Curio

Every project on Curio is driven by a research question: the platform draws on the natural curiosity of the citizen scientists to help with research tasks to answer those questions, and it allows citizen scientists to visualize the partial results whenever possible. This approach is distinguished from

those of other platforms, through which participants collect and annotate data in an open-ended way, not knowing exactly how and when the data will be used. Our question-driven approach is in spirit similar to a recently launched Zooniverse project called "Galaxy Zoo Quench" (`http://quench.galaxyzoo.org`), where the project has a limited time scale and is driven by a research purpose.

Designing Tasks and Workflows

Many researchers are end-users, as opposed to creators, of web technology. As such, the process of designing a task – i.e., reasoning about and creating the flow of activities and the corresponding user interface designs – can be challenging. On Curio, researchers create tasks through a set of guided questions, focusing their attention on specifying what they want participants to accomplish, as opposed to designing user interface.

Certain tasks are complex, requiring decomposition into simpler subtasks and chaining of these subtasks into a workflow. For example, in order to classify medical images or time series data, experts use a set of complex rules that take years to internalize. In these cases, instead of requiring the expert to explicitly decompose the classification task or to specify an algorithm, Curio asks experts to express their knowledge by demonstrating their reasoning on concrete examples, and automatically induces the underlying workflow.

Monitoring and Refining

Curio supports an iterative approach to crowdsourcing, where requesters (i.e., researchers) monitor the solution as it is being assembled and intervene to improve the computational outcomes, e.g., by clarifying the task instructions, providing additional examples, or communicating with the workers. Iterations are needed especially because experts often over-estimate or under-estimate the capabilities of non-experts, and may unknowingly use jargon in the task instructions that is inaccessible to novices. In addition, research is a non-linear process: the interests and needs of the researcher can change over time. Certain tasks may become higher or lower on the priority list, as the researcher refines his question and approach. Curio provides several algorithms for inferring quality in the absence of ground truth, data visualizations to support researchers in identifying problematic stages in the workflow and the flexibility to modify various aspects of the projects, without major disruption from the participants' points of view.

Communicating with Participants

In previous studies of citizen science platforms (Rotman et al. 2012), it was found that there is a disconnect between citizen scientists and scientists. Citizen scientists may be initially motivated to help, but they often find scientists "unfriendly" and lose motivation when they lose track of why their work matters. On the Curio platform, we facilitate researcher-participant communication in multiple ways. First, as participants are completing tasks, our platform intermittently shows them visualizations that emphasize their unique contribution as well as partial results to the research question. Second, the platform provides a voting system that allows researchers and participants to debate ambiguous cases in a more targeted way than open-ended forum discussions. Finally, researchers are encouraged to specify prizes that will appeal to the specific interests of the contributors, and participants earn prizes by completing tasks in an accurate and efficient way.

Sharing Data

One challenge in question-driven citizen science is that both the research intent and data is publicly exposed before the publication stage. Depending on the particular research discipline and its existing data sharing practices, researchers may or may not be comfortable making their work externally visible too early in the research process. To address this, Curio provides an option for researchers to keep their project private and by invitation only, and allows researchers to schedule the release of their data, thus encouraging sharing of resources without compromising the need for privacy.

Forming Hierarchical Teams

Beyond sharing data, researchers are often reluctant to assign major responsibilities to novices, whom they do not yet trust to have enough expertise and experience. To address this, Curio mimics the existing hierarchical culture and quality control practices in academia, where more senior members supervise those who are less experienced, and members of a research team meet informally to discuss and decipher ambiguous data. On Curio, researchers can specify a project to be accessible to participants by invitation only. This essentially allows each researcher to create his own crowd and select participants based on his unique set of expertise requirements and other difficult-to-articulate criteria, e.g., trust, accountability and personality. Researchers can also can invite *collaborators* with substantial domain knowledge (e.g., graduate students, postdocs, colleagues and advanced citizen scientists) to form a management team, to collaboratively monitor and debug the crowdsourcing process.

Conclusion

We expect the Curio platform to launch in the fall and host several projects, e.g., where participants annotate museum records to evaluate the effects of climate change on flowering times, and design new organic molecules for powering solar cells. As a human computation system, Curio raises exciting new research opportunities and challenges.

References

Rotman, D.; Preece, J.; Hammock, J.; Procita, K.; Hansen, D.; Parr, C.; Lewis, D.; Jacobs, D.; and Biswas, A. 2012. Dynamic changes in motivation in collaborative citizen science projects. In *CSCW*.

Acknowledgments