

Quality Hackathon: Evaluating the Products of Online Co-Production Systems

Andrea Wiggins
University of Maryland,
College Park
4121G Hornbake Bldg
College Park, MD 20742
wiggins@umd.edu

David Gurzick
Hood College
401 Rosemont Avenue
Frederick, MD 21701
gurzick@hood.edu

Sean Goggins
University of Missouri
221B Townsend Hall
Columbia, MO 65201
goggins@missouri.edu

Brian Butler
University of Maryland,
College Park
4120K Hornbake Bldg
College Park, MD 20742
bsbutler@umd.edu

ABSTRACT

This full-day workshop focuses on building Big Social Data research competencies for scholars interested in issues of contribution quality and contributor performance in online co-production systems that generate value through contributions by volunteers. The workshop is designed to engage discussion and promote co-working through a hackathon format to stimulate productive conversation and learning, using shared data sets to provide a common focus for participants to engage questions of contribution quality and contributor performance with multiple disciplinary, theoretical, and analytical backgrounds.

Categories and Subject Descriptors

K.4.3 [Computers and Society]: Organizational impacts – computer-supported cooperative work.

General Terms

Human Factors, Measurement, Performance

Keywords

Contribution quality; performance; user-generated content; online communities; peer production; data quality; hackathon

1. INTRODUCTION

As the contexts of online co-production continue to expand, from software in open source systems, to consumer tastes in recommender systems, to encoding knowledge in Wikipedia, to research data in citizen science, questions continue to resurface around contributor performance, contribution quality, and collective productivity. How do we evaluate productivity, characterize performance, or rate the quality of contributions from specific individuals? How and to what extent are concepts, measures, or frameworks identifiable and reusable across contexts?

The quality of co-production is an important factor in the broader adoption and use of these systems, and also for evaluation of overall success in goal-oriented communities [7]. In all cases, the products of these communities are considered suspect until accepted evaluative measures can be applied to verify the value of contributions [2], both at the level of the individual and collectively.

Permission to make digital or hard copies of part or all 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 third-party components of this work must be honored. For all other uses, contact the Owner/Author.

Copyright is held by the owner/author(s).

GROUP '14, Nov 09-12 2014, Sanibel Island, FL, USA

ACM 978-1-4503-3043-5/14/11.

<http://dx.doi.org/10.1145/2660398.2660429>

Yet naive measures of contribution volume and quality are rarely transferable across contexts and the approaches used to evaluate conventional products do not always translate to co-production environments with large-scale contributor bases and atomized or uneven individual contributions. As a result, both research and adoption of online co-production platforms is slowed by the need to redevelop contribution assessment for each new context.

Simplistic heuristics, such as “who contributes the most” and user-generated rankings have typically proven unsatisfactory in isolation [9]. Free/libre and open source software development was one of the earliest contexts for understanding the nuances of co-production, with complex, multifaceted frameworks developed to evaluate software quality outside of proprietary environments [10,15,13]. Substantial progress has been made in Wikipedia research, which has taken several different approaches to evaluate editors, their productivity, and the quality of their contributions [1,6,14].

In question-answering communities, numerous factors impact the evaluation of responses, including whether there is a bounty offered, the diversity of the community, and personal characteristics of the asker [3,5,11]. In other domains, such as citizen science, quality control strategies for water quality projects (e.g., [12]) are inapplicable to projects that involve contributors in gathering species occurrence data [4], and projects focused on image classification take altogether different strategies [8]. Due to wide variability in the nature of the data that can be extracted to evaluate contributions and performance, the degree to which concepts and findings from one context might apply to another remains unclear, and there are few examples of cross-context studies employing similar measures in different communities.

2. PURPOSE & GOALS

In short, evaluating contribution quality and contributor performance in co-production communities remains a critical question, but discussions of how to approach this challenge are limited by the context of participation and specifics of the available data. Transcending these constraints and sharing knowledge around the conceptual and practical issues of conceptualizing and evaluating contribution quality and contributor performance will help stimulate a higher level of research discourse across multiple contexts of participation and academic disciplines.

The primary goals of the proposed workshop are to facilitate productive discussions among individuals whose work focuses on different contexts of participation, foster the development of a network of researchers with shared interests in evaluating contribution quality and contributor performance, and provide

opportunities for developing skill and collaborative research relationships through hands-on engagement.

3. WORKSHOP

3.1 Focus and Audience

The conceptual focus of the workshop will be on identifying concepts, measures, tools, and analyses that can advance our understanding of contribution quality and contributor performance in large online communities. The audience for the workshop will be researchers and practitioners with interests related to these themes, and those seeking opportunity to develop collaborative relationships with others in the Group community. Although participants will have experience with diverse research contexts and approaches, applying such knowledge and skills to a new context provides a good opportunity for learning, synthesizing, and generating ideas.

Participants will apply with a short (~1 page) document describing their interest in studying contribution quality and contributor performance with a short summary of their applicable skills, theories, tools, and ideas. These documents will be summarized, aggregated, and distributed to all participants in advance to seed the day's activities. We expect that the participants will bring a broad mix of theoretical, methodological, technical, and analytical assets to the workshop, and providing background on participants' intellectual diversity helps set expectations for flexibility around specific hackathon activities as well as team composition. Despite its heterogeneity, the advanced technical knowledge and experience of the Group community should generate an adequate distribution of skills to allow each team to have a productive experience.

3.2 Activities

Workshop participants will be provided advance access to the data; we anticipate using one or two data sets. Examples of candidate data sets, pending evaluation of Terms of Use, include the Yelp Dataset Challenge data, Reddit images and comments, Stack Exchange data, and FLOSSmole data. An agenda with workshop plans and expectations will also be distributed in advance. Participants will be encouraged to examine the data, consider which other participants with whom they may wish to work, and prepare questions or ideas to work on during the workshop.

The workshop will open with a brief overview of the format, goals, and plans for the day (10 minutes). Each participant will provide a brief introduction to their work, their applicable skills, and specific interest in the workshop themes (30 minutes). We will then engage a brief brainstorming process, eliciting the questions and ideas attendees have considered in advance and iterating on these ideas (30 minutes). The participants will then be able to self-select into small groups of 3-4 individuals to pursue data-driven hacking oriented toward developing and implementing one or more measures of content quality or contributor performance in the shared dataset. The workshop organizers will join in with groups as needed.

Before lunch, we will break briefly to give two-minute status updates for each group, both to acknowledge the work completed to that point and to seed meal time conversation. Following lunch, the organizers will lead a short activity for the full group intended to bring awareness to the spectrum of emerging challenges and perhaps foster a new set of connection points for participants to work together. Participants will be encouraged to change groups or to form larger clusters as appropriate. We will reserve the last 90 minutes of the session for full-group discussion during which each team will present a short debrief on their progress over the course of

the day and share any datasets, visualizations, and analyses they have produced.

While the primary goal of the hackathon is to produce a functional analysis of contribution quality and contributor performance, we will endeavor to create a low-pressure environment for exploratory learning and data play. The hackathon workshop model, while loosely structured and lightly managed, was highly successful at CSCW 2014 and we anticipate it will work effectively for Group 2014 as well.

3.3 Potential Outcomes

There are a variety of potential outcomes from a hackathon-style workshop focused on data analysis. By creating the opportunity for participants to work together with tools and theories focused on shared data sets, we expect the workshop will create an environment suitable for professional development gains for participants: introduction to new research skills and theoretical perspectives, refinement of ideas and questions for research, and developing new collaborative relationships. New research projects and publications could also emerge from the starting point provided by co-working at the workshop. In addition, participants will gain experience with the hackathon model of peer production in a research-oriented context, which was a popular feature of prior related workshops.

3.4 Logistics

The workshop will benefit from having the following equipment available: one projector, 1-2 flip charts with markers, and snacks. We can accommodate up to 20 participants.

3.5 Organizers

Our co-organizers are experienced in facilitating and participating in workshops and hackathons, familiar with theories and tools suitable for addressing the problem of evaluating contribution quality in online production systems, and play well with others.

Andrea Wiggins co-organized the successful CSCW 2014 Online Communities Data (OCData) Hackathon and has participated in several community infrastructure development hackathons. She has expertise in the topics of human computation, open source systems, and large-scale collaboration; her methodological skills include operationalizing theory using digital trace data, measure development, and social network analysis.

David Gurzick participated in the CSCW 2014 OCData Hackathon and has experience working with a variety of related data sets. His expertise includes research in the design, evaluation, and operation of online communities and the development and configuration of open-participation systems. His methodological background includes the application of mixed-method analysis to sociotechnical data and the use of design science techniques. He has strong software development skills and extensive professional experience in developing online systems and applications with user-centered methodologies.

Sean Goggins co-organized the CSCW 2014 OCData Hackathon and led a similar 50-person event at the 2013 iConference; he has also organized several smaller production-driven hackathons to develop community resources. These events were all considered very successful despite differences in focus and widely varying audiences and needs. Goggins' subject matter expertise includes computer-supported cooperative learning and learning analytics, with methodological skills in social network analysis and computational linguistic analysis.

Brian Butler co-organized the CSCW 2014 OCData Hackathon, directed the Digital Societies and Social Technologies 2014 Summer Institute, and has led various doctoral/junior faculty workshops. In both cases, the events focused on the goals of building relationships within an interdisciplinary community, developing the capabilities of individual researchers, and catalyzing theory-oriented, data-enabled collaborative projects. He has worked with general online communities, Wikipedia, Q&A forums, and other co-production platforms. His prior work has focused on using empirical and simulation models to develop and test generalizable models of contribution and participation dynamics in online communities.

4. Acknowledgments

This work was supported in part by U.S. National Science Foundation Grant IIS 1449209.

5. REFERENCES

- [1] Blumenstock, J. E. (2008). Size matters: word count as a measure of quality on Wikipedia. In *Proceedings of the 17th international conference on World Wide Web*. ACM, New York, NY, 1095-1096.
- [2] Duguid, P. (2006). Limits of self-organization: Peer production and “laws of quality”. *First Monday*, 11(10).
- [3] Harper, F. M., Raban, D., Rafaei, S., & Konstan, J. A. (2008). Predictors of answer quality in online Q&A sites. In *Proceedings of CHI 2008*. ACM, New York, NY, 865-874.
- [4] Hochachka, W. M., Fink, D., Hutchinson, R. A., Sheldon, D., Wong, W. K., & Kelling, S. (2012). Data-intensive science applied to broad-scale citizen science. *Trends in Ecology & Evolution*, 27(2), 130-137.
- [5] Jurczyk, P., & Agichtein, E. (2007). Discovering authorities in question answer communities by using link analysis. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*. ACM, New York, NY, 919-922.
- [6] Kittur, A., & Kraut, R. E. (2008). Harnessing the wisdom of crowds in Wikipedia: quality through coordination. In *Proceedings of CSCW 2008*. ACM, New York, NY, 37-46.
- [7] Preece, J. (2001). Sociability and usability in online communities: Determining and measuring success. *Behaviour & Information Technology*, 20(5), 347-356.
- [8] Prestopnik, N. R., Crowston K., & Wang J. (2014). Exploring data quality in games with a purpose. In *Proceedings of iConference 2014*.
- [9] Riedl, C., Blohm, I., Leimeister, J. M., & Krcmar, H. (2010). Rating scales for Collective Intelligence in Innovation Communities: Why quick and Easy Decision Making Does not Get IT Right. In *Proceedings of ICIS 2010*.
- [10] Samoladas, I., Gousios, G., Spinellis, D., & Stamelos, I. (2008). The SQO-OSS quality model: measurement based open source software evaluation. In *Open source development, communities and quality*. Springer US, 237-248.
- [11] Shah, C., & Pomerantz, J. (2010). Evaluating and predicting answer quality in community QA. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*. ACM, New York, NY, 411-418.
- [12] Sheppard, S. A., & Terveen, L. (2011). Quality is a verb: The operationalization of data quality in a citizen science community. In *Proceedings of WikiSym 2011*. ACM, New York, NY, 29-38.
- [13] Spinellis, D., Gousios, G., Karakoidas, V., Louridas, P., Adams, P. J., Samoladas, I., & Stamelos, I. (2009). Evaluating the quality of open source software. *Electronic Notes in Theoretical Computer Science*, 233, 5-28.
- [14] Suzuki, Y., & Yoshikawa, M. (2012). QualityRank: assessing quality of Wikipedia articles by mutually evaluating editors and texts. In *Proceedings of the 23rd ACM conference on Hypertext and social media*. ACM, New York, NY, 307-308.
- [15] Taibi, D., Lavazza, L., & Morasca, S. (2007). OpenBQR: a framework for the assessment of OSS. In *Open Source Development, Adoption and Innovation*. Springer US, 173-186.