Understanding crowdsourcing projects: A systematic review of tendencies, workflow, and quality management

Fábio R. Assis Neto*, Celso A.S. Santos

Federal University of Espirito Santo, Av. Fernando Ferrari, 514 - Goiabeiras, Vitória, ES, 29075-910, Brazil

ABSTRACT

Crowdsourcing is a powerful method that leverages human intelligence to solve problems in the medical, linguistics, machine learning, and in a considerable number of other fields of study. This systematic review aims to understand how crowdsourcing projects are designed and executed in the state-of-the-art, considering the following dimensions: Task execution, quality management, and platform usage. Our results summarized trends of the important aspects of a crowdsourcing project, such as crowd and task types, crowdsourcing platforms, and activities used to manage the quality; we also addressed functions and limitations in traditional crowdsourcing platforms, the definition of a crowdsourcing workflow, and the lack of standardization when developing a crowdsourcing project. As future works, we wish to design a conceptual model that represents the important elements of a crowdsourcing project and their relationships, also, we wish to do further researchers focusing on how the quality mechanisms listed in this work should be used in crowdsourcing projects to get reliable results.

1. Introduction

Crowdsourcing, in short, is a general-purpose problem-solving method, which uses a group of participants willing to help solve a proposed problem (Doan, Ramakrishnan, & Halevy, 2011). This method has proven useful in medicine (Dumitrache, Aroyo, Welty, Sips, & Levas, 2013; Foncubierta Rodríguez & Müller, 2012; Irshad et al., 2017; Maier-Hein et al., 2014; Vernez et al., 2017; Zhai et al., 2013), linguistics (Chowdhury et al., 2014; Enochson & Culbertson, 2015; Lasecki et al., 2013), machine learning (Aigrain et al., 2016; Noronha, Hysen, Zhang, & Gajos, 2011; Roemmele, Archer-McClellan, & Gordon, 2014), software development and testing (Dietl et al., 2012; Dwarakanath et al., 2015; Fava, Shapiro, Osborn, Schäef, & Whitehead Jr., 2016; Leicht, Knop, Blohm, Müller-Bloch, & Leimeister, 2016; Weidema, López, Nayebaziz, Spanghero, & van der Hoek, 2016), and in a considerable number of other fields of study.

As stated by Yin, Gray, Suri, and Vaughan (2016), the crowdsourcing platforms hide personal attributes and social characteristics of participants. In these platforms, the communication usually is only in one direction from the person asking for a solution (the requester or problem owner) to each participant individually (the worker). Hence, participants commonly are abstracted as merely a black box method to accomplish microtasks associated with the problem solution. However, other authors consider the communication in crowdsourcing as a two-way process, such as stated by Oleson et al. (2011), which asked for feedback from the workers to generate ground truth data (gold units) in their crowdsourcing project. Furthermore, complex tasks require increased communication with the workers. As tasks become more complex, the relationship between the crowdsourcer and workers become more akin to outsourcing than crowdsourcing (Staffelbach et al., 2015).

---

* Corresponding author.
  E-mail address: fabio.ribeiro.neto@gmail.com (F.R. Assis Neto).

A crowdsourcing process can be viewed as both collaborative or cooperative. Roschelle and Teasley (1995) defined two statements that try to demonstrate the difference between collaborative and cooperative work:

Collaboration is a coordinated, synchronous activity that is the result of a continued attempt to construct and maintain a shared conception of a problem. (p. 70)

Cooperation is accomplished by the division of labor among participants as an activity where each person is responsible for solving a portion of the problem. (p. 70)

In a crowdsourcing collaborative process, collaborators usually have a broad view of the entire process and frequently rely on the communication and contributions of other participants to realize their work, e.g., Idea Jam[1] and Ushahidi[2] approaches. In another hand, a crowdsourcing cooperative process usually is controlled by a human computation algorithm, in which workers contribute individually, and applied to solve a problem proposed by a crowdsourcer and hidden from the workers, e.g., Mturk.

By following a workflow, each member of the crowd works on a portion or on the entire problem, performing a human computation task which produces useful results towards the solution to the crowdsourcer problem. Planning, coordination, and control of process are defined by a production manager, usually, the crowdsourcer, although other authors argue that the responsibility for workflow design can be shared between the crowd and the crowdsourcer (Kulkarni, Can, & Hartmann, 2012).

Kittur et al. (2013) defined that crowdsourcing workflows are needed once they facilitate decomposing tasks into subtasks, managing the dependencies between subtasks, and assembling the results. However, our view of this concept is broader. Crowdsourcing workflow is context-oriented and should be composed of at least one or more tasks sets and one or more quality control activities to guarantee reliable results, which can be executed in parallel, iterative, and sequential ways. Considering our conceptions, it is possible to set quality activities before, during, or after the task execution stages, thus allowing the crowdsourcer to better understand how the quality will be managed during his project. Considering that a workflow can contain multiple task sets, each set is composed of tasks with the same objective, which are executed following a workflow. Thus, our conceptions conform the ideas of Kittur et al. (2013), hence facilitating decomposing a complex task into subtask sets, managing their dependencies, and assembling their results.

Hence, in our conception, a crowdsourcing workflow is a unique entity related to each project. It should represent the planning and coordination done by the crowdsourcer in the entire project, thus, representing a broader view of who will participate, what will be produced, and how it will be done at every stage of the workflow.

Popular studies in crowdsourcing, as in Estellés-Arolas and González-Ladrón-de Guevara (2012), Brabham (2012), and Zhao and Zhu (2014) present the development of crowdsourcing field of study by evaluating known systems and general-purpose platforms, as the Amazon Mechanical Turk (Mturk),[3] CrowdFlower,[4] Innocentive,[5] and Threadless[6].

Classic definitions of the crowdsourcing term usually neglected concepts that we consider crucial for every project, in particular the definition of a crowdsourcing workflow and methods to control the quality of the entire project. Hence, this issues motivated us to review the state-of-the-art of crowdsourcing and try to understand how crowdsourcing projects were executed, not only how systems and platforms works. Thus aiming to validate our assumptions and disseminate new conceptions.

In this review, we extend the work presented in previous crowdsourcing surveys and address essential points to be considered when developing a crowdsourcing project. If popular papers in this domain evaluated crowdsourcing platforms and systems, our view is broader, regarding the evaluation of quality management and the essential elements of a crowdsourcing project. Furthermore, we present classifications and trends regarding the crowd, tasks, platforms, task results, products generated, and the quality management in crowdsourcing projects. Thus, one of our goals is to highlight the main concerns of crowdsourcing projects. These concepts provide crowdsourcers and researchers with essential information to successfully design their projects. This paper is, therefore, an up-to-date systematic literature review of crowdsourcing projects that aims to provide detailed information about this new research field. Hence, this review presents the following primary contributions:

- Classifications and tendencies of crowd types, task types, crowdsourcing platforms, and activities used to manage the quality of a crowdsourcing project.
- Observations about the limitations of traditional crowdsourcing platforms.
- Our view of the crowdsourcing workflow, regarding tasks and quality management.

This paper is organized as follows: Section 2 presents state-of-the-art background; Section 3 introduces the methods used in our approach; Section 4 presents the results of our review; Section 5 presents a discussion about all information obtained during the review process. Finally, Section 6 proposes future works and concludes this paper.

---

[1] ideajam.io
[2] ushahidi.com
[3] mturk.com
[4] crowdflower.com
[5] innocentive.com
[6] threadless.com