Human-centric Computing
and Information Sciences

REVIEW

Open Access

CrossMark

An investigation of unpaid crowdsourcing

Ria Mae Borromeo* and Motomichi Toyama

*Correspondence:
ria@db.ics.keio.ac.jp
Department of Information
and Computer Science,
Keio University, 3-14-1
Hiyoshi, Kohoku-ku,
Yokohama 223-8522, Japan

Abstract

The continual advancement of internet technologies has led to the evolution of how individuals and organizations operate. For example, through the internet, we can now tap a remote workforce to help us accomplish certain tasks, a phenomenon called crowdsourcing. Crowdsourcing is an approach that relies on people to perform activities that are costly or time-consuming using traditional methods. Depending on the incentive given to the crowd workers, crowdsourcing can be classified as paid or unpaid. In paid crowdsourcing, the workers are incentivized financially, enabling the formation of a robust workforce, which allows fast completion of tasks. Consequently, in unpaid crowdsourcing, the lack of financial incentive potentially leads to an unpredictable workforce and indeterminable task completion time. However, since payment to workers is not necessary, it can be an economical alternative for individuals and organizations who are more concerned about the budget than the task turnaround time. In this study, we explore unpaid crowdsourcing by reviewing crowdsourcing applications where the crowd comes from a pool of volunteers. We also evaluate its performance in sentiment analysis and data extraction projects. Our findings suggest that for such tasks, unpaid crowdsourcing completes slower but yields results of similar or higher quality compared to its paid counterpart.

Keywords: Crowdsourcing, Platforms, Sentiment analysis, Data extraction

Background

Computer technologies, which were once just used to aid individuals and organizations in their operations, have become a necessity. Due to the ability of computers to perform high speed and exact calculations, many human processes have been automated. Nevertheless, humans still perform better than computers in areas such as ideation, judgment, and perception. To maximize the strengths of both humans and computers, human computation, a computer science technique that is involved in the design or analysis of information processing systems in which humans participate as computational elements [1], has been widely studied.

Crowdsourcing is a form of human computation defined as the practice of obtaining information or services by soliciting input from a large number of people via the internet [2]. It has gained popularity over the past decade due to its ability to provide relatively cheap and fast solutions.

Borromeo and Toyama *Hum. Cent. Comput. Inf. Sci.* (2016) 6:11

Page 2 of 19

Based on the crowdsourcing taxonomy deduced by Hosseini et al. [3] there are four pillars of crowdsourcing: the requester, the crowd, the task, and the platform. The *requester* is the entity that has a set of tasks to crowdsource. The *task* is the activity to be done by the *crowd*, who is a group of people from the general public willing to participate with or without financial incentive. The crowdsourcing *platform* is the system used by requesters to publish tasks and by crowd workers to complete tasks.

Crowdsourcing platforms can either be paid or unpaid. In paid crowdsourcing platforms such as the Amazon mechanical turk (MTurk) [4], CrowdFlower [5], and Microworker [6], the crowd comes from a pool of paid workers. By contrast, in unpaid crowdsourcing platforms such as Crowd4U [7] and Zooniverse [8], the crowd comes from volunteers who receive no financial incentive.

Paid crowdsourcing platforms provide requesters with tools that promote their tasks to be completed by someone in a timely fashion, at equal or better quality than a full-time workforce. They enable the provision of monetary incentives along with other incentives, like recognition, which improves the likelihood of getting satisfactory results promptly [9].

Unpaid crowdsourcing platforms provide functionalities similar to paid platforms without built-in provision for monetary incentives. They employ volunteers as workers thus the completion time of tasks could be difficult to estimate. Nevertheless, since running a task in unpaid platforms does not require payment to workers, it may be an economical alternative for individuals and organizations who are more concerned about the budget than the task turnaround time.

In this study, we explore unpaid crowdsourcing by reviewing crowdsourced applications used in disaster response and relief, traffic management, and education that are powered by volunteer workers. We also implement sentiment analysis and data extraction applications in an unpaid crowdsourcing platform and compare their performance to their counterparts in paid platforms.

This paper is organized as follows. First, we briefly review the basics of crowdsourcing and introduce crowdsourcing platforms. Second, we discuss unpaid crowdsourcing and examples of its applications that are used in various fields. Third, we report the results of our sentiment analysis and data collection studies. After that, we share insights on our findings then finally conclude and present our future work.

Crowdsourcing essentials

The term *Crowdsourcing* was first published in 2006 in Jeff Howe's Wired Magazine article entitled *The Rise of Crowdsourcing* [10]. He further defined it as the act of taking a job traditionally performed by an employee and outsourcing it to an undefined, generally large group of people in the form of an open call [11]. Since then, varying definitions of crowdsourcing have emerged. In 2012, Estellés-Arolas et al. attempted to come up with an integrated definition of crowdsourcing. They analyzed 40 original definitions from research papers in the databases of ACM, IEEE, ScienceDirect, SAGE and Emerald, and came up with a definition that covers any crowdsourcing initiative. Their definition is as follows: *Crowdsourcing is a type of participative online activity in which an individual, an institution, a non-profit organization, or company proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary*

Borromeo and Toyama *Hum. Cent. Comput. Inf. Sci.* (2016) 6:11

Page 3 of 19

undertaking of a task. The undertaking of the task, of variable complexity and modularity, and in which the crowd should participate bringing their work, money, knowledge and/or experience, always entails mutual benefit. The user will receive the satisfaction of a given type of need, be it economic, social recognition, self-esteem, or the development of individual skills, while the crowdsourcer will obtain and utilize to their advantage what the user has brought to the venture, whose form will depend on the type of activity undertaken [12]. In addition to the four pillars of crowdsourcing, they pointed out two other elements: the incentive and the solution. The descriptions of the elements they identified are as follows.

- *Requester* An individual, an institution, a non-profit organization, or a company that has a problem to be solved or tasks to be completed
- *Task* The work to be done which is of variable complexity and modularity
- *Crowd* A group of individuals of varying knowledge, heterogeneity, and number who provides solutions or completes tasks; also known as workers
- *Platform* An application that provides crowdsourcing functionalities related to crowd and task management
- *Incentive* Satisfaction of a given type of need, be it economic, social recognition, self-esteem, or the development of individual skills
- *Solution* What the user has brought to the venture

A popular crowdsourcing example is the search for Jim Gray. Gray was a computer scientist who received a Turing Award for his seminal contributions to database and transaction processing research and his technical leadership in system implementation [13]. Early in 2007, he failed to return from his sailing trip around the Farallon Islands. One of the efforts done to find him and his boat was capturing satellite images of the ocean then asking volunteers from the general public to identify which images should be further examined. His colleagues posted tasks in the MTurk that asked volunteers to compare a few 300 by 300 pixels image sub-tiles to a template then provide a score for each sub-tile for evidence of features similar to the ones provided in the template [14]. In this example, the requester is Gray's colleagues; the task is to find images that should be further examined; the crowd consists of MTurk workers; the platform is MTurk; and the solution is the aggregate of results received for each sub-tile. There is no explicit incentive or extrinsic reward, but the workers were likely driven by an intrinsic motivation to help.

Asking people to count malaria-infected blood cells on images of a Petri dish using public displays is another example of crowdsourcing, specifically called *situated crowdsourcing* [15]. Situated crowdsourcing is a unique form of crowdsourcing, typically unpaid, where workers who are available serendipitously, perform tasks that are associated with a particular context. The context, which may be a single place, a single event or a single event in a single place, offers a resource for structuring the activity of the crowd [16]. In this example, the requester is a group of researchers; the task is to count malaria-infected blood cells on images of a Petri dish; the crowd consists of people with physical access to the public displays; the platform is a public display terminal; and the solution is the count of malaria-infected blood cells. Similar to the previous example, there is no

Borromeo and Toyama *Hum. Cent. Comput. Inf. Sci.* (2016) 6:11

Page 4 of 19

explicit incentive or extrinsic reward, but the crowd could have been motivated by the perceived altruism involved in the task.

Crowdsourcing platforms

A crowdsourcing platform is a system that connects requesters and workers. It is commonly a web application that provides functionalities for task and crowd management. For requesters, the choice of a crowdsourcing platform may depend on the nature of the projects they want to crowdsource and the incentive they are willing to provide to the workers. Some crowdsourcing platforms are specialized in particular projects e.g. InnoCentive [17] for research and development and ClickWorker [18] for managing e-commerce data. Other platforms are general-purpose such as MTurk, Microworker, and CrowdFlower.

General purpose platforms usually specialize in handling simple tasks or microtasks. Microtasks are tasks that require minimal time and cognitive effort but when combined can result in major accomplishments [19]. They are typically simple, repetitive, independent, and short. Examples of microtasks include labeling, transcription, text tagging, and sentiment analysis. However, there are more complex tasks that are context-heavy, interdependent, require more cognitive effort, and may take many hours to complete [20]. These tasks are called macrotasks. Examples of macrotasks include programming, document editing, and sentence translation. Macrotasks may be deployed on general purpose platforms in their original form or decomposed into microtasks. Larger tasks with more complex requirements such as software engineering and journalism are typically published in generic online outsourcing marketplaces or global online work platforms such as Upwork [21] and Freelancer [22].

We can also classify crowdsourcing platforms as either paid or unpaid. In paid platforms, requesters can tap into a large population of workers around the globe to accomplish tasks in a fraction of the time and money of more traditional methods [23]. On the other hand, requesters rely on volunteers or other crowd gathering techniques in unpaid platforms. In the following subsections, we will discuss examples of paid and unpaid platforms.

Paid platforms examples

The *Amazon mechanical turk* or MTurk for short is one of the most popular crowdsourcing platforms, available only to requesters in the United States. It has been used in studies from human linguistic annotation to image classification and has been a topic of interest for researchers on human–computer interaction (HCI), information retrieval, computer science, economics, and data mining [24]. As a marketplace for work, it provides requesters an on-demand and scalable workforce and allows fast turnaround time and relatively low costs of setting up a project [4]. It also implements a mechanism for filtering workers based on the workers' qualifications to uphold the good quality of work. According to Ipeirotis, approximately 80 % of the mechanical turk workers are from the US and 20 % are from India [25].

CrowdFlower is another general purpose crowdsourcing platform that distributes tasks to its workers and workers from different partner labor channels [26]. Partnering with a multitude of labor channels with workers from all over the world (e.g. clixsense

[27] and neobux [28]) significantly increases the scalability of the platform and diversifies the workforce [29]. For quality assurance, CrowdFlower screens workers by testing them using questions whose answers are already known (gold questions), by tracking contributor response velocity to answer distribution, and by assigning a confidence score to every unit of completed work [5].

Another platform similar to MTurk and CrowdFlower is the *microworkers* platform. Like CrowdFlower, it is open to requesters and workers regardless of geographic location. It employs task management and verification measures by using text Captcha test, gold questions, system and employer task verification, and individual or mass task rating [6].

Unlike the platforms mentioned above where all crowdsourcing processes can be done remotely through the Internet, *bazaar* is a paid platform for situated crowdsourcing. Requesters create and manage tasks using a web application while workers complete tasks using the touch screens in physical kiosks that can be placed in different locations [30]. Workers are paid with HexaCoins, a virtual currency which workers can exchange for money or goods.

Unpaid platforms examples

Crowd4U is a volunteer-based microtasking platform developed by the University of Tsukuba, Japan, whose crowd members consists of volunteers from universities or other research institutions. It provides a form-based rapid application development (RAD) tool to allow easy creation of commonly implemented tasks such as majority voting and translation [31]. To create more complex tasks, requesters must be knowledgeable in CyLog, a Datalog-like language that incorporates a proper feedback system for humans at the language level [32]. Tasks may be published and completed in the official Crowd4U website or embedded in other websites [33]. Additionally, tasks may also be performed in mobile devices by installing an application that requires a user to perform a microtask to unlock the screen.

Bossa is an open source software framework for volunteer-based crowdsourcing developed by the University of California, Berkeley. A Bossa instance must be hosted on a Linux server. Within the instance, requesters can generate, present and manage tasks by using PHP scripts while volunteers can perform tasks and interact with other volunteers. Bossa also provides support for dealing with the differences in the skills of volunteers by maintaining estimates of the skill level of volunteers and ensuring that for each task, there is a consensus of compatible results among a sufficient set of volunteers [34]. Bossa also integrates with Bolt, a web-based training framework, which could be used to train volunteers on how to perform tasks.

PyBossa is a free, open-source framework for crowdsourcing [35], developed by Sci-Fabric, a company that develops open source software for crowdsourcing research. As its name suggests, it is based on Bossa but rewritten in the Python programming language. Similar to Bossa, programming skills are required to create and manage tasks thus making it not as user-friendly as the popular paid crowdsourcing platforms. Nevertheless, it is equipped with features that help in the management of tasks and analysis of results. Requesters have the option to deploy their instance of PyBossa or publish tasks

Borromeo and Toyama *Hum. Cent. Comput. Inf. Sci.* (2016) 6:11

Page 6 of 19

in Crowdcrafting [36], a live instance of PyBossa, where tasks may be created, managed, and completed.

Zooniverse describes itself as the world's largest and most popular platform for people-powered research [8]. It started with Galaxy Zoo, a website where volunteers can participate in research by classifying galaxies from the Sloan Digital Science survey. Similar projects followed and as of writing, Zooniverse hosts dozens of projects, where anyone can participate in crowdsourced scientific research. Additionally, anyone can create a project in Zooniverse and tap on its community of volunteers.

CrowdButton [37] is a work-in-progress volunteer-based platform for situated crowdsourcing. The platform is made of a server and multiple Wi-Fi enabled reporting devices. The device consists of one replaceable question area, four arcade game buttons with built-in LED lights. The initial implementation supports multiple-choice questions, which volunteers can answer by pressing a physical button in a specific location. The answers are sent to the server where they are aggregated.

Unpaid crowdsourcing

In general, the use of unpaid crowdsourcing platforms is referred to as *unpaid crowdsourcing*. More specifically, unpaid crowdsourcing is a type of crowdsourcing wherein the workers are not incentivized by money but are motivated by other factors such as reputation, status, peer pressure, fame, community identification, and fun [38]. However, some non-monetary incentives depend on the nature of the task or the identity of the requester [39]. For example, unpaid sentiment analysis and data extraction crowdsourcing projects may not attract volunteers because they can be tedious and would primarily benefit the requester. To recruit volunteers in unpaid crowdsourcing, requesters must turn to other means of crowd gathering techniques such as requiring users, making users work in return for a service, and piggybacking on the user traces of a well-established system [40]. It appears tedious compared to paid crowdsourcing but through this approach, requesters can take advantage of a certain level of crowd control, an opportunity to get high-skilled volunteers, and a workforce without labor cost.

Crowd control may be implemented by making tasks visible only to certain people. For example, a crowdsourcing task within an educational institution can only be made available to students who fit certain criteria. On the other hand, requesters have the opportunity to get high skilled volunteers if a task is available to the public because all sorts of workers can work on a task, including high-skilled workers, who typically charge a high fee. It can be argued that these advantages have their corresponding disadvantages such as not getting enough workers and getting spammed by malicious workers. Nonetheless, there have been many successful unpaid crowdsourcing projects that we will tackle in the following subsections. We will discuss citizen science and several projects in disaster response and relief, traffic management, and education that relied on volunteer workers.

Citizen science

Citizen science is a form of research that involves the participation of the general public to aid in carrying out scientific research [41]. It is very similar to crowdsourcing as they both involve tasks to be accomplished and workers to accomplish the tasks through a platform. However, their objectives are different. The objective of crowdsourcing

Borromeo and Toyama *Hum. Cent. Comput. Inf. Sci.* (2016) 6:11

Page 7 of 19

projects could be personal, business or scientific, while the primary objective of citizen science projects is to advance scientific research [42].

Typically, the workers who are volunteers with or without specific scientific training, perform or manage research-related tasks such as observation, measurement or computation [43]. Many projects involve both citizen science and crowdsourcing. Nevertheless, not all citizen science projects include crowdsourcing, and not all crowdsourcing projects include citizen science [42].

While the paid and unpaid crowdsourcing platforms discussed in earlier can be used for citizen science, there are other specialized platforms where one can create citizen science projects, such as CitSci.org [44], Zooniverse [8], and iNaturalist [45].

Disaster response and relief

Volunteer-based crowdsourcing has been used in disaster response and relief projects. Disaster response and relief is a unique application of volunteer-based crowdsourcing because volunteers may not just be volunteers who solve computational problems but also volunteers who help in the actual response and relief operations. In Chu et al. [46] CROSS system, disaster surveillance data collection is crowdsourced through volunteers who explore threatened areas. CROSS works by calling out for volunteers through social networks such as Facebook and Twitter. It then plans exploration routes for volunteers who responded based on the volunteers' locations. CROSS continuously interacts with volunteers throughout their exploration by collecting reports, integrating messages and displaying useful information on a map.

Starbird's Tweak the Tweet (TtT) [47] was used in the aftermath of the Haiti earthquake in 2010. It is an idea for utilizing the Twitter platform for crowdsourcing information provisions during disasters and mass emergency events. Unlike other more formalized crowdsourcing systems deployed on crowdsourcing platforms, TtT operates completely within the existing functionality of Twitter. The idea is to ask volunteers who are exploring affected areas to add special hashtags into their crisis-related tweets to make the tweets machine-readable. Other crisis-related tweets without the proper hashtags are formatted by another group of volunteers into tweets with the special hashtags. As a result, the volunteers were able to provide a self-activated, self-organized layer of human computation that provides a mechanism for searching and filtering disaster-related information on social media.

Like TtT, the artificial intelligence for disaster response (AIDR) [48] system takes its input data from Twitter. The tweets are classified into a set of user-defined categories using machine learning techniques and crowdsourcing. It works by allowing the user to input keywords, geographic locations or languages to filter tweets. Based on the filtering criterion, it harvests relevant tweets. It crowdsources the labeling of the harvested tweets using the PyBossa platform. The labels from volunteers are then used as input to the system's machine learning algorithm to be able to classify tweets automatically. AIDR was successfully tested to classify informative tweets posted during the 2014 Pakistan earthquake.

Borromeo and Toyama *Hum. Cent. Comput. Inf. Sci.* (2016) 6:11

Page 8 of 19

Traffic management

In traffic management, the crowd is primarily used for collecting data regarding traffic conditions, in specific geographic locations. One of the most used traffic applications is Waze, a community-based traffic and navigation application [49]. It is a data collection platform that allows users to post observations of traffic incidents such as road construction, hazards, or accidents [50]. When the application is running on a mobile device, the user passively contributes traffic and other road data but can also voluntarily send more details about his or her location. The information collected from Waze users enables the application to give other users real-time traffic updates and suggest optimal routes for their destinations.

Another application of crowdsourcing in traffic management is in the improvement of traffic light timing in intersections. Riley et al. introduced a tool that allows mobile phone users to play a game challenging users to find the optimal light timing for a simulated traffic intersection [51]. The crowd is challenged to create new configurations for optimal timing or improve configurations from the high score list. The authors further plan to model an actual road intersection and see if the application can find an improved signal configuration.

Artikis et. al. [52] proposed an intelligent urban traffic management system that uses fixed sensors positioned on intersections, and mobile sensors installed in public transport vehicles. Sensors provide information to the system while humans serve as judges to resolve conflict in the data collected by the sensor. The system queries human volunteers or workers close to the location of the conflict to determine which data is real or not.

Education

The use of crowdsourcing in education has also been explored. In 2011, Bow et al. [53] developed a crowdsourcing model for creating tools to aid studying preclinical medicine at the John Hopkins University School of Medicine. They developed a simple Java program and used Google Drive to enable the crowd to create and edit flashcards simultaneously. The crowd, which consists of medical students in the class of 2014, populated a database with more than 16,000 flashcards. An analysis of the students' exam scores revealed that the students in the class of 2014 outperformed those students in the class of 2013 who did not have access to the system.

Aside from the creation of study tools, crowdsourcing has been used in grading or evaluation of students' requirements. CrowdGrader [54] is a system available to the general public that lets students submit and collaboratively review and grade homework [55]. Within the application, students can submit homework and evaluate several submissions of other students. The students are given an overall crowd score, which reflects the quality of their homework and the quality of reviews they give. The system is beneficial to students as they can benefit from the feedback of their peers and learn from the solutions submitted by others. Aside from those, CrowdGrader also helps instructors facilitate student learning and handle grading and evaluation of large classes.

In the two previous illustrations of crowdsourcing in education, the crowd consists of students. However, Dow et al. [56] studied how input from an anonymous online crowd affects student learning and motivation for project-based innovation work. The crowd's

Borromeo and Toyama *Hum. Cent. Comput. Inf. Sci.* (2016) 6:11

Page 9 of 19

role was to provide students with authentic users' opinions and realistic market forces in their projects. According to students, the online crowd helped them to identify needs quickly and inexpensively in the early stages of the innovation process. However, in the later stages, the students received a large, quantity of feedback with low quality.

Comparison of paid and unpaid crowdsourcing

In the previous sections, we differentiated paid and unpaid crowdsourcing based on the incentives they provide to workers and the platforms where they are deployed in. Only a few studies have compared the two based on the quality of results. In one of these studies, Mao et al. [57] adapted an annotation task that was originally performed by volunteers in the Planet Hunters citizen science project, to an experiment in MTurk with paid workers. They investigated how three types of payment schemes (pay per task, pay for time, and pay per annotation) influenced the behavior of paid workers compared to volunteers. Their findings show that given appropriate incentives, paid crowd workers might work at a faster rate and achieve similar accuracy compared to volunteers who are working on the same task [57]. Goncalves et al. [15] compared the performance of unpaid situated crowdsourcing for counting Malaria-infected blood cells with the performance of the same task deployed in MTurk. They observed that in unpaid crowdsourcing through public displays, the accuracy of results were lower but the rate of uptake of tasks was higher compared to MTurk.

Since related studies are still inconclusive, we perform experiments to investigate further the quality of results produced by paid and unpaid crowdsourcing. In the following subsections, we will discuss two crowdsourcing projects that are commonly done in MTurk and CrowdFlower: sentiment analysis and data extraction. We deployed these projects in both paid and unpaid platforms then compared their results based on completion time, crowd costs and quality of results. We used PyBossa as the unpaid crowdsourcing platform and CrowdFlower as the paid platform.

Case study 1: sentiment analysis

Sentiment analysis, which is also referred to as opinion mining, is defined as the task of finding opinions of authors about specific entities [58]. It may be conducted manually by experts determining the sentiment of a given text, automatically using machine learning algorithms and statistical methods, and by crowdsourcing. In this study, we performed sentiment analysis on student evaluation comments by paid and unpaid crowdsourcing. The comments, which were written in English, were collected from the 2006 to 2012 Student Evaluation of Teaching (SET) comments [59] for Professor Jonathan Cox, a Mathematics professor at the State University of New York at Fredonia. From the unstructured document that contained the comments, we were able to extract 418 comments with an average number of 46.52 words each and a standard deviation of 35.48.

Paid version

We first created the sentiment analysis project in CrowdFlower. CrowdFlower provides a template for sentiment analysis, which includes default settings such as the payment per task and the number of judgments per task. The default values are 0.01 USD and 3, respectively. Using the template, we only had to provide instructions and specify the

Borromeo and Toyama *Hum. Cent. Comput. Inf. Sci.* (2016) 6:11

Page 10 of 19

input data source. Fig. 1 shows the instructions and Fig. 2 shows a sample task. The cost of the project totaled 15.35 USD, which included the 0.01 USD payment per task and the platform service charges. The project was completed in 2.90 h by 86 workers. Each worker performed an average of 14.58 tasks with a standard deviation of 7.64.

For every task, CrowdFlower chooses the response which has the highest confidence score. The confidence score is based on a worker's trust score, a value that ranges from 0 to 1, where 0 is the lowest and 1 is the highest. The confidence score is calculated by adding the trust scores of contributors then dividing it by the sum of trust scores of contributors for a specific task [60]. The summary generated by CrowdFlower showed that the paid crowd detected 141 positive, 231 negative, and 46 neutral comments.

Unpaid version

We used the same settings, instructions, and input data as the paid version to create a crowdsourced sentiment analysis in PyBossa. Fig. 3 shows the landing page and Fig. 4 shows a sample task. We advertised the project on Twitter and Facebook and sent personal email messages to invite certain people to participate. Forty-six volunteers
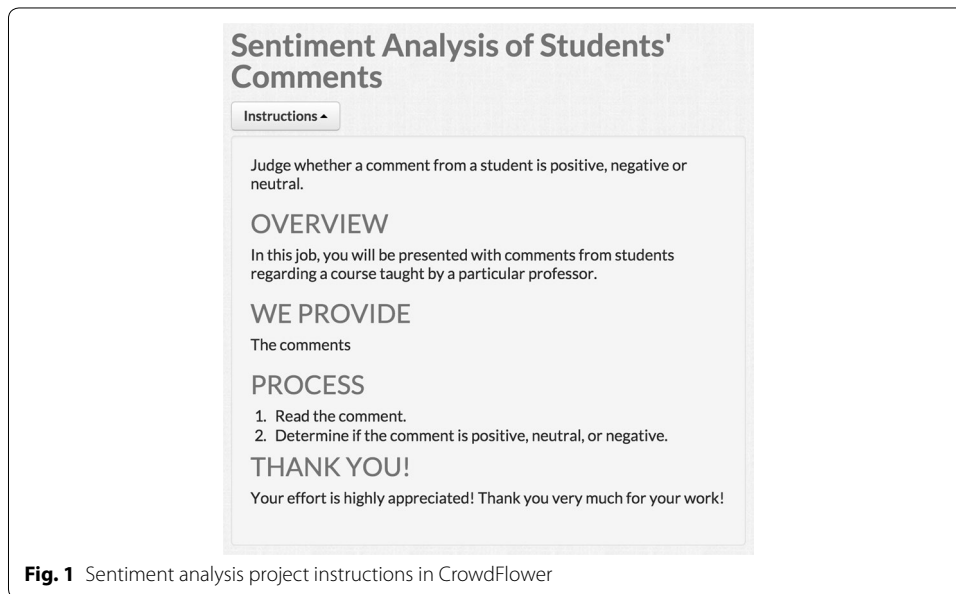
Fig. 1 Sentiment analysis project instructions in CrowdFlower
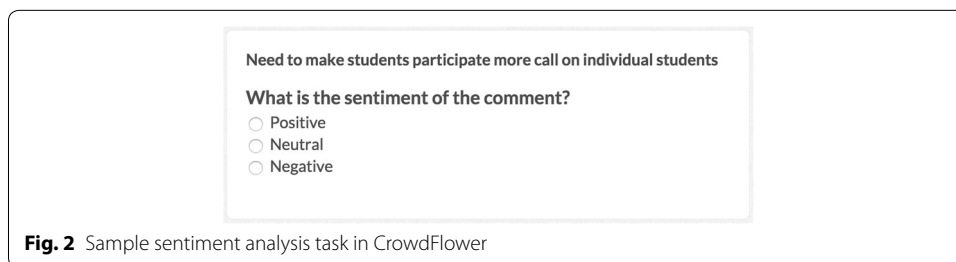
Fig. 2 Sample sentiment analysis task in CrowdFlower

Borromeo and Toyama *Hum. Cent. Comput. Inf. Sci.* (2016) 6:11

Page 11 of 19

Fig. 3 Sentiment analysis project instructions in PyBossa

Fig. 4 Sample sentiment analysis task in PyBossa
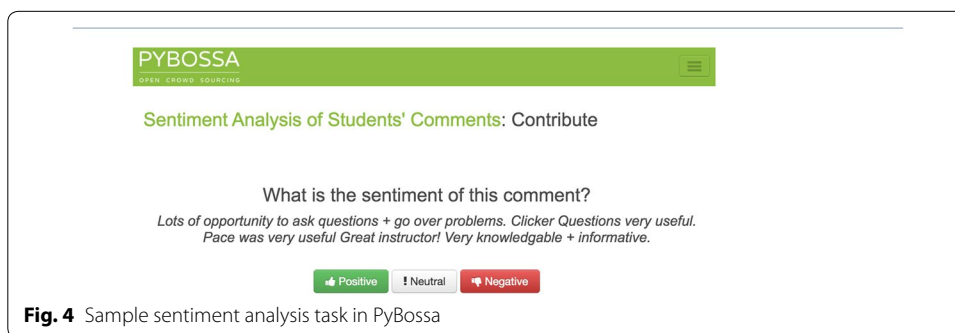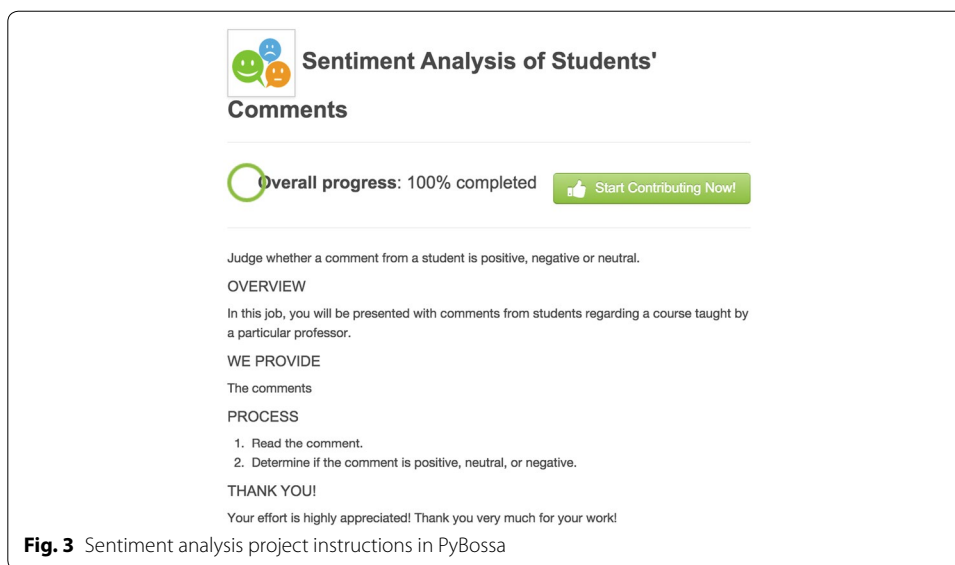
completed all the tasks in 44.8 h. Each volunteer performed an average of 28.50 tasks with a standard deviation of 60.65.

The output of the project was a CSV file containing the task runs. Post processing had to be performed to derive the sentiment of each comment. We adopted CrowdFlower's formula in deriving the final judgment. However, since we did not have the volunteers' trust scores, we assumed that they are all trustworthy and assigned them a trust score of 1. The resulting formula is equivalent to the rule of the majority. When a task received three different responses, it was classified as neutral. We derived 138 positive, 204 negative and 76 neutral comments from the unpaid crowd's responses.

Results

In Table 1, we summarize the comparison of the two methods in terms of crowd cost, completion time, and accuracy. *Crowd cost* is the amount paid to the platform and the crowd workers; *completion time* refers to the time from when the project was launched to the time when all required responses were received, and *accuracy* is the degree of similarity of the results compared to a gold standard. We use the manual evaluation of the same comments in [61] as the gold standard to measure the method's accuracy.

Borromeo and Toyama *Hum. Cent. Comput. Inf. Sci.* (2016) 6:11

Page 12 of 19

Table 1 Comparison of paid and unpaid Crowdsourced sentiment analysis

	Paid	Unpaid
Crowd cost (USD)	15.35	0.00
Completion time (hours)	2.90	44.80
Precision	76.08	76.32

As presented in Table 1, the paid method completed significantly faster than the unpaid version while the accuracy of the two crowdsourced methods is marginally similar.

Case study 2: data extraction

In the Database Laboratory at Keio University Faculty of Science and Technology, Japan, students take turns in studying research papers related to their research topics and presenting these papers to the entire laboratory. The presentation files are stored in a digital repository, which is available to all other students. From 2008 to 2015, 341 research papers have been presented. However, to date, there is no summary or index of the presentations, making the search for a particular paper presentation difficult. In this project, we aim to extract information from the digital repository and create a meaningful index that could be useful to the researchers. Since the project requires data to be extracted from PDF files that do not have a standard format, automatic data extraction is difficult, and manual data extraction is tedious. Nevertheless, in this case, crowdsourcing is a suitable option.

Aside from the benefits of the project to the research laboratory, we also want to compare the quality of results from paid and unpaid crowdsourcing for this type of task. To achieve this, we created a gold set of the 70 research papers presented in 2008 and 2009 by manually pasting the APA bibliographic information of each paper to a text file then running a script to get the desired information. We intend to use the gold set as a baseline to evaluate the quality of results from paid and unpaid crowdsourcing and as gold test questions when we crowdsource all the 341 research papers.

We designed a task to provide a worker with a link to the digital archive and ask him or her to extract the title, authors, source, year of publication, presenter, and presentation date of 70 research paper presentations from 2008 to 2009. It is important to note that one-third of the research papers presented was in Japanese, and the rest were in English.

Paid version

We published the project in CrowdFlower using one of the data collection and enrichment templates that the platform provides. Since one-third of the papers was in Japanese, we provided detailed instructions on how to decipher Japanese text. Nevertheless, we did not assume the workers to be able to make meaning of a different character set; thus, we expected that the accuracy of the responses from the CrowdFlower job to be at most 67 %. Figure 5 show the instructions given to the workers and Fig. 6 shows a sample task to be performed. Adopting CrowdFlower's default settings, we asked three responses for each paper. We paid 6.36 USD, which includes the 0.025 USD incentive for
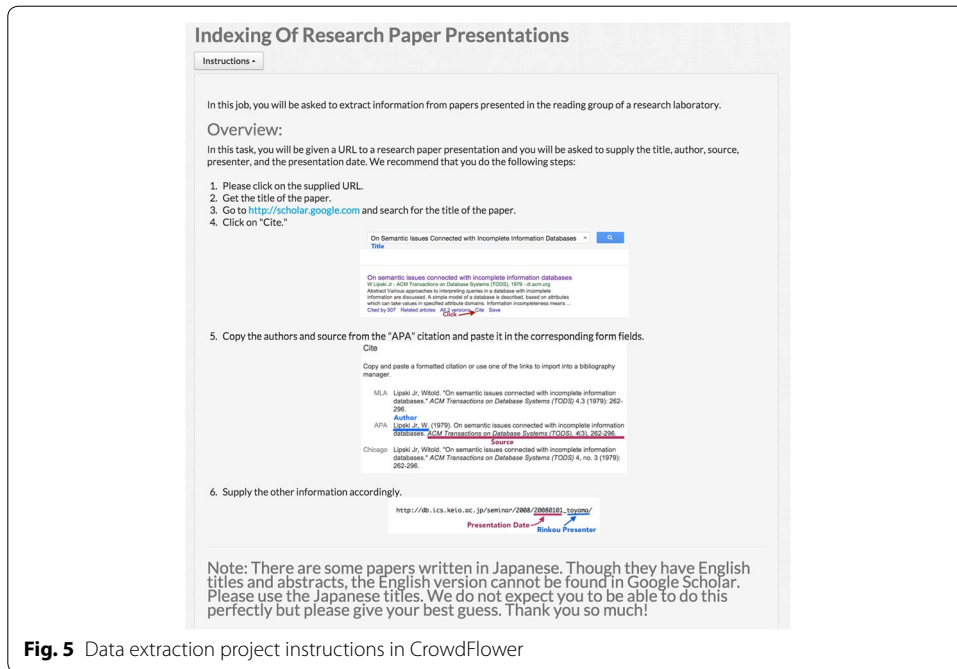
Borromeo and Toyama *Hum. Cent. Comput. Inf. Sci.* (2016) 6:11

Page 13 of 19

Fig. 5 Data extraction project instructions in CrowdFlower

Fig. 6 Sample data extraction task in CrowdFlower

every task and the platform service charges. The project was completed in approximately 2.50 h by 43 workers.

Unpaid version

We deployed the same project in PyBossa. Figure 7 shows the project's landing page and Fig. 8 shows a sample task. To gather crowd workers, we requested the participation of the members of a database research laboratory, consisting of 20 students and one professor, who are all equipped with basic English and Japanese language skills. We set a deadline 9 and 1/2 days or 228 h after the announcement of the project and the project completed within the allotted time.

Borromeo and Toyama *Hum. Cent. Comput. Inf. Sci. (2016) 6:11*

Page 14 of 19

Results

A data cleaning script was executed to ensure that the results from both crowdsourced methods did not have unnecessary white spaces or symbols. We then calculated the accuracy of the results by measuring their similarity to the gold set. In Table 2, we present the comparison of results in terms of crowd cost, completion time, and accuracy.

Fig. 7 Data extraction project instructions in PyBossa

Fig. 8 Sample data extraction task in PyBossa

Table 2 Comparison of paid and unpaid Crowdsourced data extraction

	Paid	Unpaid
Crowd cost (USD)	6.36	0.00
Completion time (hours)	2.30	228
Precision	27.05	89.61

Borromeo and Toyama *Hum. Cent. Comput. Inf. Sci.* (2016) 6:11

Page 15 of 19

Since the results from the paid crowd only achieved an accuracy of 27.05 %, we further examined the results and observed that 19.86 % of the responses were spam answers or random answers that did not attempt to answer what was asked.

The unpaid crowd achieved 89.61 % accuracy, which is significantly higher than that of the paid crowd. This accuracy could be attributed to the characteristics of the crowd: they had sufficient understanding of the project as they are potential beneficiaries of its successful completion, and they had sufficient language skills that are necessary to complete some of the tasks. Furthermore, they all belong to a research laboratory in a University, which implies that they have a specific degree of education as opposed to the anonymous paid crowd.

The completion time of the unpaid crowdsourcing project was remarkably slower than the paid version. However, we noted that the deadline influenced the completion time thus it could have been faster or slower depending on the set deadline.

Discussion

From our experiments, we observed that the quality of unpaid crowd work is similar to paid crowd work in the sentiment analysis project but is significantly higher than paid crowd work in the data extraction project. We also noted that in paid crowdsourcing, there are only two general steps required to complete a project. The first is to deploy the task and the second is to wait for results. On the other hand, in unpaid crowdsourcing, there is a very crucial additional step, which was the most challenging based on our experience: gather the crowd.

In addition to these observations, we experienced several crowdsourcing issues noted in the following subsections.

SPAM responses

In the sentiment analysis project, we were not able to determine whether a response is spam or not because of the limited answer choices, which were all valid. However, as reported earlier, 19.86 % of the responses from the paid data collection project were obviously spam. Spamming is a common problem in paid crowdsourcing because workers may still receive payment whether or not their answers are of good quality thus they have the option to hastily or randomly fill out input fields. Although spam detection and prevention mechanisms have been studied for paid crowdsourcing [62, 63], it has not been widely explored in unpaid crowdsourcing. In our case, there may be no obvious spam in the results of the unpaid data extraction project, but we cannot conclude that spam does not exist in unpaid crowdsourcing.

Task design

The crowd members in the unpaid data collection project were supposedly high skilled. Nevertheless, they only achieved 89.61 % accuracy. This observation led us to examine our task design. Although the task is just to fill up a simple web form, we noted later on that there are unnecessary form fields in our task. For example, the presentation date and the name of the presenter can be removed because they can be easily extracted from the URL of the digital archive. From this, we understood that ideally, only tasks that are extremely hard to automate are suited for crowdsourcing. The inclusion of Japanese

Borromeo and Toyama *Hum. Cent. Comput. Inf. Sci.* (2016) 6:11

Page 16 of 19

paper presentations may have also been a bad design choice and served as a demotivator for workers.

Platform choice

Based on these projects, we identified some factors to consider in choosing a platform for crowdsourcing: the type of task, the time requirement, and the budget. In paid platforms, there are templates for common tasks such as sentiment analysis, image tagging, and data extraction. They also allow relatively faster completion time but incur crowd costs. On the other hand, as we noted in the related works that we reviewed and the case studies we performed, there are types of tasks that are suitable for unpaid crowdsourcing. Additionally, monetary incentives to workers are not required thus it could be beneficial to requesters with budget constraints.

Limitations and future work

Quality control

Paid platforms such as CrowdFlower and MTurk have quality control mechanisms. For example, CrowdFlower relies on the process of generating gold units with known answers to provide targeted training feedback to workers and to prevent common scamming scenario [29]. This feature, however, is not built into PyBossa, the unpaid platform that we chose. Since we were unable to implement the same quality control feature in PyBossa, we did not use any quality control mechanisms on both platforms. In the future, we shall implement quality control mechanisms in an unpaid platform, and come up with a new set of comparisons.

The real cost of unpaid crowdsourcing

The experiments were also limited to comparing the crowd costs. However, there are other costs involved in unpaid crowdsourcing that needs to be measured such as the cost of the server setup and maintenance. If taken into consideration, it would be interesting to note if unpaid crowdsourcing is indeed an economical alternative to paid crowdsourcing.

Incentives and motivations

Regarding incentives, we mainly used values suggested by CrowdFlower. As for motivations, unlike the unpaid crowdsourcing projects we reviewed, the tasks in the case studies seem to be only beneficial to the requester and are likely unable to motivate volunteers. It is probable that the volunteers only participated in the projects because they were asked and it was a one-time thing. It would be interesting to observe the volunteers' engagement if we ask the same crowd to do more tasks of similar nature.

Platforms

There are many available crowdsourcing platforms but in this study, we only compared CrowdFlower and PyBossa. In the future, we could design a more advanced crowdsourcing experiment for more platforms, using different incentives and motivations, which would produce unbiased comparisons.

Borromeo and Toyama *Hum. Cent. Comput. Inf. Sci.* (2016) 6:11

Page 17 of 19

Summary

In this paper, we provided an introduction to crowdsourcing by reviewing research related to crowdsourcing definitions, taxonomy, platforms, and sample applications. In particular, we focused our investigation on unpaid crowdsourcing and its applications in various fields of study. Additionally, we compared unpaid to paid crowdsourcing by applying the two crowdsourcing methods to sentiment analysis and data extraction projects. Through the experiments, we noted the strengths and weaknesses of unpaid crowdsourcing in terms of project completion time, crowd costs, and quality. Additionally, we observed general crowdsourcing issues that are related to spam responses, task design, and platform choice. Finally, we identified further points of comparison between the two methods that involve quality control, cost analysis, incentives and motivations, and platforms.

References

1. Human computation. http://www.hcjournal.org
2. Crowdsourcing N. http://www.oed.com/view/Entry/376403?redirectedFrom=crowdsourcing
3. Hosseini M, Phalp K, Taylor J, Ali R (2014) The four pillars of crowdsourcing: a reference model. In: IEEE Eighth International conference on research challenges in information science (RCIS), IEEE, pp 1–12
4. Amazon mechanical turk — welcome. http://www.mturk.com/
5. Crowdflower company. http://www.crowdflower.com/company
6. Microworkers — work & earn or offer a micro job. http://www.microworkers.com
7. Crowd4U. A nonprofit "Open Crowd" platform for microvolunteering and crowdsourcing — crowd4U. http://www.crowd4u.org/en/
8. Zooniverse. https://www.zooniverse.org/
9. Frei B (2009) Paid crowdsourcing. Current state & progress toward mainstream business use, smartsheet.com report, Smartsheet.com 9
10. Howe J (2006) The rise of crowdsourcing. Wired Magazine 14(6):1–4
11. Crowdsourcing: a definition. http://www.crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html
12. Estellés-Arolas E, González-Ladrón-de-Guevara F (2012) Towards an integrated crowdsourcing definition. J Inf Sci 38(2):189–200
13. James 'Jim' Nicholas Gray. http://amturing.acm.org/award_winners/gray_3649936.cfm
14. Hellerstein JM, Tennenhouse DL (2011) Searching for jim gray: a technical overview. Commun ACM 54(7):77–87
15. Goncalves J, Ferreira D, Hosio S, Liu Y, Rogstadius J, Kukka H, Kostakos V (2013) Crowdsourcing on the spot: altruistic use of public displays, feasibility, performance, and behaviours. In: Proceedings of the 2013 ACM international joint conference on pervasive and ubiquitous computing, pp 753–762. ACM
16. Erickson T (2011) Some thoughts on a framework for crowdsourcing. In: Workshop on crowdsourcing and human computation, pp 1–4
17. Innovation management | crowdsourcing | challenges | competitions. http://www.innocentive.com
18. Your content provider—text creation, researches, tagging. http://www.clickworker.com
19. Morris RR, Dontcheva M, Gerber EM (2012) Priming for better performance in microtask crowdsourcing environments. IEEE Internet Computing 16(5):13–19
20. Haas D, Ansel J, Gu L, Marcus A (2015) Argonaut: macrotask crowdsourcing for complex data processing. Proceedings of the VLDB endowment 8(12)
21. Upwork, the world's largest online workplace. http://www.upwork.com
22. Hire freelancers & find freelance jobs online—freelancer. http://www.freelancer.com

Borromeo and Toyama *Hum. Cent. Comput. Inf. Sci.* (2016) 6:11

Page 18 of 19

23. Kittur A, Smus B, Khamkar S, Kraut RE (2011) Crowdforge: crowdsourcing complex work. In: Proceedings of the 24th Annual ACM symposium on user interface software and technology, ACM, pp 43–52

24. Chen JJ, Menezes NJ, Bradley AD (2011) Opportunities for crowdsourcing research on amazon mechanical turk. In: Proceedings of the CHI 2011 workshop on crowdsourcing and human computation. http://crowdresearch.org/chi2011-workshop/papers/chen-jenny.pdf

25. Ipeirotis P (2015) Demographics of mechanical turk: now live! (April 2015 Edition) | a computer scientist in a business school. http://www.behind-the-enemy-lines.com/2015/04/demographics-of-mechanical-turk-now.html

26. CrowdFlower labor channels. http://www.crowdflower.com/labor-channels

27. Make money taking surveys, earn free cash online, paid per click advertising | Clixsense. http://www.clixsense.com

28. Neobux The innovation in paid-to-click services. http://www.neobux.com

29. Oleson D, Sorokin A, Laughlin GP, Hester V, Le J, Biewald L (2011) Programmatic gold: targeted and scalable quality assurance in crowdsourcing. In: Proceedings of the 11th AAAI conference on human computation. AAAI Press, pp 43–48

30. Goncalves J, Hosio S, Ferreira D, Anagnostopoulos T, Kostakos V (2015) Bazaar: a situated crowdsourcing market. In: Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing and proceedings of the 2015 ACM international symposium on wearable computers, ACM, pp 1385–1390

31. Morishima A, Amer-Yahia S, Roy SB (2014) Crowd4u: an initiative for constructing an open academic crowdsourcing network. In: Second AAAI conference on human computation and crowdsourcing

32. Morishima A, Shinagawa N, Mitsuishi Y, Aoki H, Fukusumi S (2012) Cylog/crowd4u: a declarative platform for complex data-centric crowdsourcing. Proc VLDB Endowment 5(12):1918–1921

33. Morishima A, Morishima A (2013) Cylog/crowd4u: a case study of a computing platform for cybernetic dataspaces. In: Michelucci P (ed) Handbook of human computation. Springer, New York, pp 561–572

34. BossaIntro—BOINC. http://www.boinc.berkeley.edu/trac/wiki/BossaIntro

35. PyBossa. http://www.pybossa.com/

36. Crowdcrafting. http://www.crowdcrafting.org/

37. Huang Y Designing a micro-volunteering platform for situated crowdsourcing. In: Proceedings of the 18th ACM conference companion on computer supported cooperative work & social computing, ACM, pp 73–76

38. Hossain M (2012) Crowdsourcing: activities, incentives and users' motivations to participate. In: International conference cn Innovation management and technology research (ICIMTR), IEEE, pp 501–506

39. Horton JJ, Chilton LB (2010) The labor economics of paid crowdsourcing. In: Proceedings of the 11th ACM conference on electronic commerce, ACM, pp 209–218

40. Doan A, Ramakrishnan R, Halevy AY (2011) Crowdsourcing systems on the world-wide web. Commun ACM 54(4):86–96

41. Lanza EL, Sanz FS, Gascón CV, García, FS, Alonso PM, Villar JC (2013) A case of citizen science for cell biology images analysis. Biblioteca Digital Brasileira de Computação

42. Citizen science or crowdsourcing? Yes! http://www.neoninc.org/updates-events/update/citizen-science-or-crowdsourcing-yes

43. Luz EF, Correa FR, González DL, Grey F, Ramos FM (2014) The forestwatchers: a citizen cyberscience project for deforestation monitoring in the tropics. Human Computation 1(2):137–145

44. CitSci.org—home. http://www.citsci.org/

45. iNaturalist.org—a community for naturalists. https://www.inaturalist.org/

46. Chu ET, Chen YL, Lin JY, Liu JW (2012) Crowdsourcing support system for disaster surveillance and response. In: 15th international symposium on wireless personal multimedia communications (WPMC), IEEE, pp 21–25

47. Starbird K (2011) Digital volunteerism during disaster: crowdsourcing information processing. In: Conference on human factors in computing systems, pp 1–4

48. Imran M, Castillo C, Lucas J, Meier P, Vieweg S (2014)Aidr: artificial intelligence for disaster response. In: Proceedings of the companion publication of the 23rd international conference on world wide web companion, International World Wide Web Conferences Steering Committee, pp 159–162

49. Free community-based mapping, traffic & navigation app. http://www.waze.com

50. Keating M, Rhodes B, Richards A (2014) Crowdsourcing: a flexible method for innovation, data collection, and analysis in social science research. In: Hill CA, Dean E, Murphy J (eds) Social Media, Sociality, and Survey Research. Wiley & Sons, New Jersey, pp 179–201

51. Riley D, Nellen G, Barrera R, Quevedo J (2014) Crowdsourcing traffic simulation to improve signal timing. J Computing Sci Colleges 29(5):112–118

52. Artikis A, Weidlich M, Schnitzler F, Boutsis I, Liebig T, Piatkowski N, Bockermann C, Morik K, Kalogeraki V, Marecek J, et al (2014) Heterogeneous stream processing and crowdsourcing for urban traffic management. In: Proceedings of the 17th international conference on extending database technology, pp 712–723

53. Bow HC, Dattilo JR, Jonas AM, Lehmann CU (2013) A crowdsourcing model for creating preclinical medical education study tools. Acad Med 88(6):766–770

54. Crowdgrader: peer grading for your classroom.http://www.crowdgrader.org

55. De Alfaro L, Shavlovsky M (2014) Crowdgrader: a tool for crowdsourcing the evaluation of homework assignments. In: Proceedings of the 45th ACM technical symposium on computer science education, ACM, pp 415–420

56. Dow S, Gerber E, Wong A (2013) A pilot study of using crowds in the classroom. In: Proceedings of the SIGCHI conference on human factors in computing systems, ACM, pp 227–236

57. Mao A, Kamar E, Chen Y, Horvitz E, Schwamb ME, Lintott CJ, Smith AM (2013) Volunteering versus work for pay: incentives and tradeoffs in crowdsourcing. In: First AAAI conference on human computation and crowdsourcing

58. Feldman R (2013) Techniques and applications for sentiment analysis. Commun ACM 56(4):82–89

59. Teaching Evaluation Comments, 2006-2012, Dr. Jonathan Cox, Department of Mathematical Sciences, State University of New York at Fredonia. http://www.fredonia.edu/faculty/math/JonathanCox/TeachEvcommentsWeb.doc

60. How to calculate a confidence score. https://success.crowdflower.com/hc/en-us/articles/201855939-How-to-Calculate-a-Confidence-Score

Borromeo and Toyama *Hum. Cent. Comput. Inf. Sci.* (2016) 6:11

Page 19 of 19

61. Borromeo RM, Toyama M (2015) Automatic vs. crowdsourced sentiment analysis. In: Proceedings of the 19th international database engineering & applications symposium, ACM, pp 90–95
62. Difallah DE, Demartini G, Cudré-Mauroux P (2012) Mechanical cheat: spamming schemes and adversarial techniques on crowdsourcing platforms. Citeseer
63. Vuurens J, de Vries AP, Eickhoff C (2011) How much spam can you take? an analysis of crowdsourcing results to increase accuracy. In: Proceedings ACM SIGIR Workshop on Crowdsourcing for Information Retrieval (CIR'11), pp 21–26