Conceptual Modeling of Crowdsourcing Variations

Todor Branzov and Krassimira Ivanova

Institute of Mathematics and Informatics – Bulgarian Academy of Sciences,
Sofia, 1113, Acad. G. Bonchev Str., Block 8,
todor.branzov@gmail.com, kivanova@math.bas.bg

Abstract. The paper presents a method for application of conceptual modeling technics to studies on crowdsourcing in various research domains. The method uses tools by industry standard Business Process Model and Notation (BPMN) and is provided as a possible solution to the problems related with non-existence of universal definition of the crowdsourcing.

Keywords: Crowdsourcing, Business Process Modeling, Conceptual Modeling, BPMN, UML

1 Introduction to the research problem

The concept for "crowdsourcing" was introduced by Howe in 2006 [1] as a further development of the popular outsourcing concept - "…crowdsourcing represents the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call." Within the following years, the concept was elaborated and was widely and successfully applied in various fields – from graphic design, through data analysis in linguistics, to software product development. The original crowdsourcing concept was eagerly apprehended by business organizations, non-profits and researches - in most suitable for the particular case of application meaning. As a result, the term "crowdsourcing" is increasingly applied by authors with various background to describe almost every case in which a network of people is used to produce some result, thus successfully diminishing usage and blurring the boundaries with some similar, but much earlier concepts like open-source development model, citizen science or virtual community. This is noted by several authors, most notably Brabham, as one of the earliest and leading researchers of the crowdsourcing, who argues particularly about the differences between open-source model and crowdsourcing [2]. There are several attempts to fix the description and the load of the concept, however, universally recognized definition in 2018 still does not exist, and there are uncounted attempts for classification of the variations of crowdsourcing. One fine example of the lack of certainty is the attempt of Estellés-Arolas et. al. [3] to apply a validated methodology for definition creation, based on usage in texts by authors with vari-

ous background, to the concept of crowdsourcing – the result is a definition that consists of more than 130 words.

The research presented in this paper is driven by the suggestion that the lack of universally accepted definition complicates the research over the phenomenon of crowdsourcing. Also, present generalization of the term, although convenient for the general public, is burdening the concept with content that makes application and elaboration over existing results of studies of earlier concepts, such aforementioned open source model, or related phenomena, like virtual community, very uncertain or even impossible [4]. In attempt to ease these consequences, we propose application of a method based on preparation of conceptual models of various crowdsourcing process variations. Further, analysis and comparison of the models would present similarities and differences between any two or more variations, or between earlier concepts and a crowdsourcing variation, thus making results easier to validate.

2 Approach to conceptual modeling of crowdsourcing variations – tools and method

We introduced a suggestion that the application of the method of conceptual modeling and some relevant tools may present an approach to cope with uncertainties in definition of crowdsourcing concept. In order to present clues in favor of that assumption, the immediate goals of the research are to choose modeling tools, to prepare sample models of some of the crowdsourcing variations, to describe method of application and to present them to the community.

2.1 Choice of modeling toolset

Method of conceptual modeling, or more specifically – conceptual modeling of information systems is part of requirements engineering and describes an industrial practice of creation of abstract, hardware and system software independent model of some solution of (usually) complex data processing problem. With roots in 1960's and 1970's, the term gained wide acceptance since 1982 report of several ISO working groups, that cope with need of common terminology of databases [5]. Although the method does not require any specific modeling tool, the authors usually use some formalized set of graphics or symbols to express key ideas and to describe characteristics and parameters of particular concepts and their integration in the system that provides the solution [6]. Standardized toolsets were developed by various companies and working groups and were accepted by professional community in the following years. Review of all of them is far beyond the scope of current paper and we don't assume it will gave any additional value against stated research goals.

For our particular goal, we strongly considered two toolsets, both of which

we assessed as contemporary, mature, lively and rich enough to be useful for the task – Unified Modeling Language (UML) and Business Process Model and Notation (BPMN). UML is developed in late 1990's primarily as a tool for modeling of software systems developed within object-oriented paradigm, however it spread quickly and was adopted as one of de-facto standards for general software systems modeling [7]. Currently it is maintained as a standard by Object Management Group (OMG) – the same group maintains BPMN – a toolset aimed at modeling business processes and workflows.

Although both UML and BPMN are widely accepted, standardized, well documented and maintained and certainly suitable for the modeling task, we choose BPMN. Reasoning behind the choice is that by purpose BPMN is designed to be easily apprehended by wider audience then UML – by purpose it is meant to be understandable not only by software engineers and computer science domain experts [8] and as such we assume it has potential to be easier for acceptance by various research groups that work in the field of crowdsourcing.

Significance of this argument is demonstrated by analysis of distribution by "Research Areas" of publications with topics related with "crowdsourcing" in Web of Science (WoS). In period 2008-10.2018, there are 5599 records in the field, of which about 54% have been tagged with research area "Computer science"; among the records, about 30% are various engineering areas and about 20% are in humanities domain (see Fig. 1.) Since BPMN is intuitive for wider audience then UML, its choice will allow easier interdisciplinary research activities.

Field: Research Areas	Record Count	% of 5,771	Bar Chart
COMPUTER SCIENCE	3,075	53.284 %	
ENGINEERING	1,499	25.975 %	
TELECOMMUNICATIONS	643	11.142 %	
BUSINESS ECONOMICS	481	8.335 %	
INFORMATION SCIENCE LIBRARY SCIENCE	236	4.089 %	
REMOTE SENSING	162	2.807 %	
SCIENCE TECHNOLOGY OTHER TOPICS	143	2.478 %	
PSYCHOLOGY	138	2.391 %	
COMMUNICATION	124	2.149 %	
LINGUISTICS	122	2.114 %	
OPERATIONS RESEARCH MANAGEMENT SCIENCE	118	2.045 %	
ENVIRONMENTAL SCIENCES ECOLOGY	101	1.750 %	
SOCIAL SCIENCES OTHER TOPICS	97	1.681 %	
EDUCATION EDUCATIONAL RESEARCH	94	1.629 %	

Fig. 1. Distribution by research areas of crowdsourcing publications in Web of Science [9]

2.2 Task relevant upsides and downsides of BPMN

Along with already stated ease of use and apprehension by wide audience, one upside of BPMN, relevant to our goal, is that currently various process engines exist that assimilate the XML file which contains a BPMN diagram. Limited amount of other action is needed to present the researchers with a software implementation of the crowdsourcing variation workflow. Although certainly not fully functional software product, it may provide additional opportunities for study.

Among downsides of BPMN is that UML has richer toolset and is generally able to provide more detailed models; however, we suggest that if and when such need arise, it most probably will relate to development of some software product or prototype, so software engineering experts that have relevant knowledge may enter the research team. Also, when needed BPMN may be complemented with some other modeling tools like Decision Model and Notation (DMN) which is used for modeling decisions that are determined by business rules and Case Management Model and Notation (CMMN) – another compatible graphical notation used for modeling of cases which require various activities that may be performed in an unpredictable order.

2.3 Method of application

We suggest the following method of application (Fig. 2). The research team have to obtain models of studied crowdsourcing variation – either reuse existing ones or create their own. Then both definition and models are analyzed in parallel and similarities and differences are noted. Again, both definition and the models are used in reasoning of some results – either original or reuse by another author.

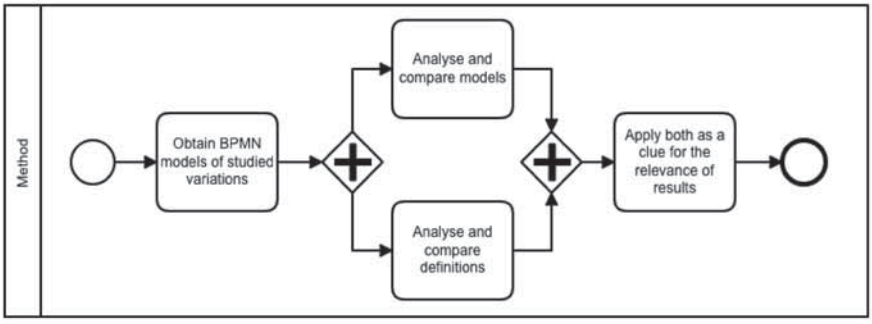

Fig. 2. Method of application

3 Sample Models of Crowdsourcing Processes

3.1 Model of Basic Crowdsourcing Process

We assume that "Basic Crowdsourcing" is the process that is described by Howe's definition (Fig. 3).

Two pools are used to model the participants in the process – the organization that crowdsource some function and the crowd participant. The organization prepares and broadcasts a call for some action, the crowd participant receives the call, perform some activities in order to complete the job and sends back the result. The model presents Howe's definition and is adequate enough to represent the key parts of it.

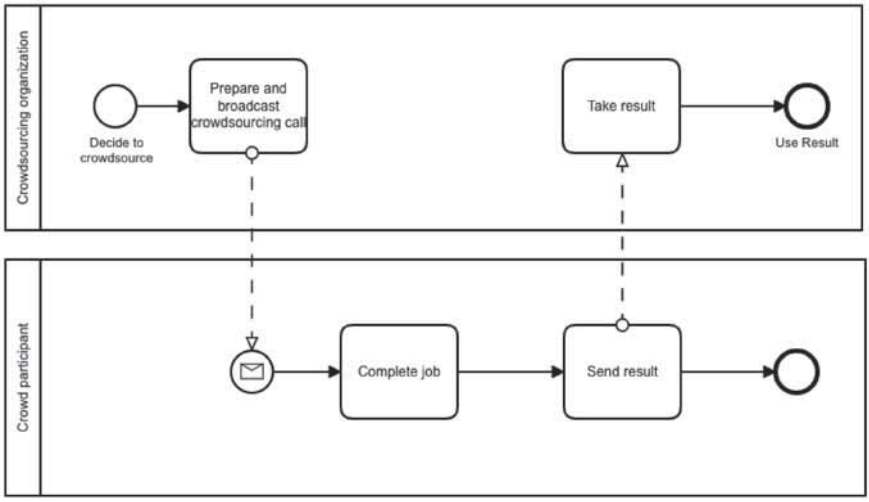

Fig. 3. Crowdsourcing process, according its original definition

3.2 Model of Basic Crowdsourcing Process in a platform

We develop this model as a variation of the basic process and we took one of the earliest platforms which implements it – Amazon's Mechanical Turk (MTurk) [10] as a model sample (Fig. 4). The key difference is that the "Take result" task from the basic model is replaced with a job verification sequence of tasks, that actually exist in MTurk. Exclusive gateway depicts the two possible verification scenarios – job may either be accepted, payed and positive feedback record prepared; or not accepted with negative feedback prepared. The sequence ends with a relevant feedback message to the worker.

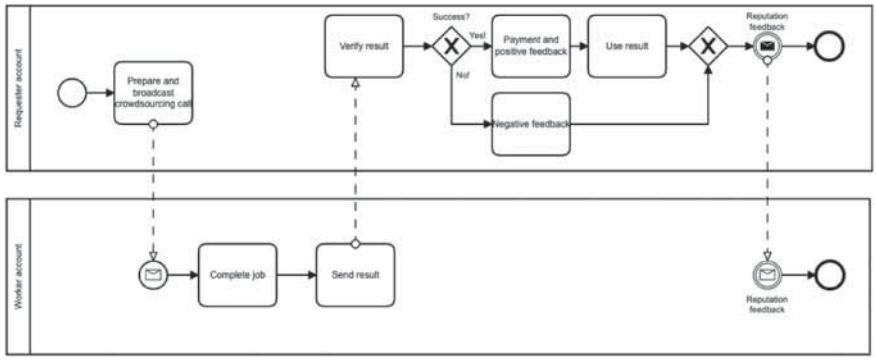

Fig. 4. Crowdsourcing process in a platform (MTurk)

If we compare both models we instantly receive an answer to one of the questions which arise with the basic Howe's model – what are the mechanisms to make a crowd member to want to take a job by a crowdsourcing organization? Within MTurk model we get at least two probable answers - payment and reputation.

3.3 Research domain specific model

In this section, we take a slightly different perspective – we use the MTurk model again, but we presume that some domain specific research requires detailed description. In the previous section, we noted that one of the probable reasons behind workers' participation is to rise their reputation. Research of this presumption lays in domains of sociology, psychology, management science. For this and similar domain specific studies some parts of the model may be remodeled with extra details (Fig. 5).

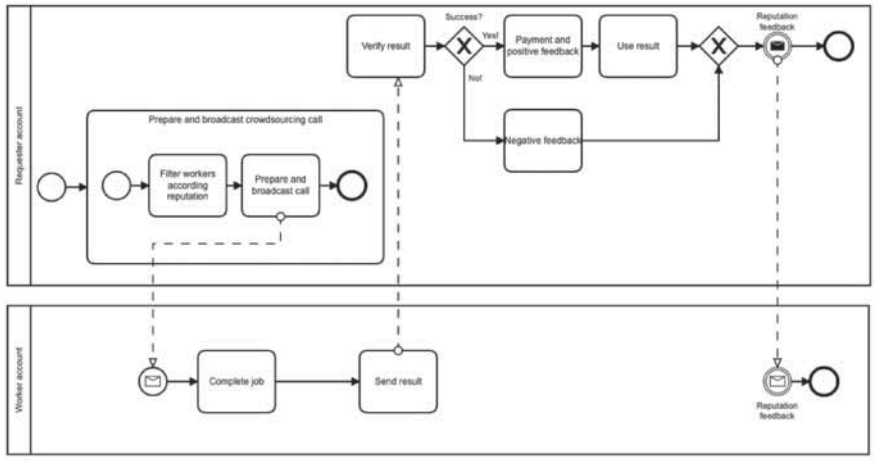

Fig. 5. Research domain specific model (MTurk)

If we compare both models now it is much clearer why would workers want a positive feedback – requesters have a mechanism to filter broadcast call only to those with sufficient reputation. Demonstrated technic of disaggregation is applied in order to make some task more detailed. The opposite technic - aggregation may prevent overcomplicating of models with details that are not needed. Both technics can be used according specific research domain needs.

3.4 Modeling crowdsourcing variation

Crowdfunding is a variation of the crowdsourcing, with purpose on obtaining financial funds needed to complete some project from the crowd. Within existing variations of the crowdfunding [11], we choose to model "the reward based crowdfunding" as it is implemented in the largest crowdfunding platform for creative projects – Kickstarter [12] (Fig. 5).

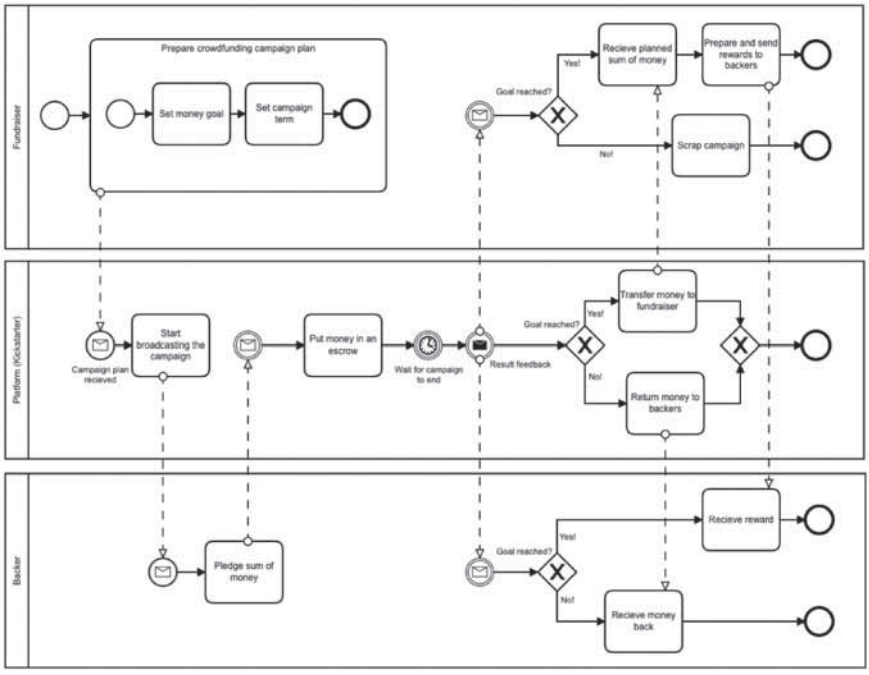

Fig. 6. Research domain specific model (Kickstarter)

This approach takes another perspective – in 3.2. and 3.3., the platform was the common environment for the participators which were performing functions in that environment, i.e. they were described as account types in the platform. Here, a new pool representing the platform is added. In this perspective, all of the participators are seen as independent entities. As such some additional aspects of their relations is possible to be modeled to the level of detail that is needed. This

perspective is suitable for studying the role of the platform in the model variation – for development purposes, for optimization of the process, or by any other cause of research interest.

4 Conclusion and future work

We demonstrated an approach that blends graphical and lexical description that may be applied to make clearer definitions of the numerous crowdsourcing variations that exist today. Although this is very common approach in broad range of research domains, we proposed specific usage of a standardized modeling tool – BPMN, that by design is developed for users with various background. Thus, we hope that it will be well accepted by wider research community and will be especially useful in interdisciplinary research, as well in educational activities.

As part of continuous research effort, we are planning to prepare a library of conceptual models for most of the crowdsourcing variations and to publish it for use by the community. Further results will be reported duly in conferences and peer-reviewed articles.

Acknowledgements. This work is supported by the project ДФНП-17-83/28.07.2017 under the Program for support of young scientists and PhD students of the Bulgarian Academy of Sciences.

References

1. Howe, J.: The Rise of Crowdsourcing. Wired. 14.06, (2006).
2. Brabham, D.: Crowdsourcing as a Model for Problem Solving. Convergence. 14, 75–90 (2008).
3. Estellés-Arolas, E., González-Ladrón-de-Guevara, F.: Towards an integrated crowdsourcing definition. J. Inf. Sci. 38, 189–200 (2012).
4. Branzov, T.: Community-sourcing in virtual societies. Serdica J. Comput. 10, 263–284 (2017).
5. Jardine, D.: Concepts and terminology for the conceptual schema and the information base. Comput. Stand. 3, 3–17 (1984).
6. Olivé, A.: 1.2. Conceptual modeling. In: Conceptual Modeling of Information Systems. p. 455. Springer Berlin Heidelberg (2007).
7. Fowler, M.: UML distilled: a brief guide to the standard object modeling language. (2004).
8. Chinosi, M., Trombetta, A.: BPMN: An introduction to the standard. Comput. Stand. Interfaces. 34, 124–134 (2012).
9. Web of Science [v.5.31] - Web of Science Core Collection Result Analysis, https://webofknowledge.com, accessed: 07.12.2018.
10. Buhrmester, M., Kwang, T., Gosling, S.D.: Amazon's Mechanical Turk. Perspect. Psychol. Sci. 6, 3–5 (2011).
11. Branzov, T., Maneva, N.: Crowdfunding Business Models and Their Use in Software Product Development. In: International Scientific Conference Informatics in Scientific Knowledge 2014, Varna, Bulgaria (2014).
12. Kickstarter, https://www.kickstarter.com, , accessed: 07.12.2018.