

# SYSTEMATIC MAPPING ON CROWDSOURCING DEVELOPMENT PROCESS

Gláucia Melo

PESC/COPPE/UFRJ, Graduate School of Engineering, Federal University of Rio de Janeiro

This study aimed to perform a systematic mapping on the software development process with crowdsourcing, aiming to know it, define it and demonstrate what the current literature can provide as knowledge. The research was done using the CAPES Portal, so all sources are published there. It was understood from this research that this process is treated according to what has been published until today, and where companies and people with an idea that can be transformed into a software, can use this methodology to obtain a product. Also seen were the restrictions on replacing outsourcing of IT specialists with this new process for system development. Another result of this research was to know what types of technology crowdsourcing has been integrated and which have made possible its use. We also saw the advantages and disadvantages of using the process and ways to manage it in specific cases.

## Introduction

The objective of this work is to perform a systematic mapping in order to obtain general knowledge about the implementation of a software development process with crowdsourcing, which requires new cultural standards and which is a new paradigm for IT.

The goal is to find out how crowdsourcing works and how it can be deployed, according to the research already done. This will add knowledge to companies and individuals who want to know the process and see if there is a possibility of using it commercially.

Almost 10 years ago, James Surowiecki, an American journalist, published *The Wisdom of Crowds* (Surowiecki, 2004), in which he explores the idea that a large group of people can provide better solutions than a small group, even this little group is made up of only brilliant people. In the book, the author explores diverse fields of knowledge, such as politics, medicine, popular culture, economics ... and artificial intelligence. It was then in 2013 that two Harvard professors, Boudreau and Lakhani, published an article called "Using the Crowd as an Innovation Partner" (Lakhani, 2013), in which they talk about the application of this idea in the corporate world.

Once it has become a widespread term and several companies are using the method successfully, crowdsourcing then becomes a factor of market differential. Today, almost all business magazines have already published on the subject, such as *IstoÉ* (<http://www.istoedinheiro.com.br/blogs-e-colunas/post/20140428/social-crowdsourcing-empresas/39.shtml>), *Época Negócios* (<http://epocanegocios.globo.com/Inspiracao/Empresa/noticia/2013/07/empresas-criando-novo-conceito-crowdsourcing-com-ajuda-das-multidoes-empresa-pode-inovar-mais.html>) and many others that speak of entrepreneurship. In the scope

of Software Engineering, although new, the method has been widely studied, and today there are articles published with, for example, proposals of frameworks for evaluation of existing crowdsourcing processes (An evaluation framework for software (Wu, Tsai, & Li, 2013)), use of collaborative software development platforms (Peng, Babar, & Ebert, 2014), systematic reviews on the subject and several other approaches. A systematic mapping was done on the subject, aiming to add to this world of research something more focused on the focus of crowdsourcing for Information Technology, in order to help those interested in the subject, if they want to know it and even use it .

Chapter 2 will present the theoretical reference of the work, that is, the state of the problem that was researched. Chapter 3 presents the research methodology, how the data were collected and analyzed, also the search criteria and methods for data extraction. In chapter 4 the analysis of the researched content and the discussion of the results are made and the conclusion is made in chapter 5.

## 2 BACKGROUND

### 2.1 THE ORIGIN OF CROWDSOURCING

The concept of cloudcomputing, the so-called cloud, has brought to the realm of technology new ways of providing various services. Servers have changed and today host your site or application in a few clicks, with fault tolerance, good performance. Databases as a service. File storage on the internet. There are many possibilities. Author of one of the best books for beginners in this area, John Rhoton explains all the concepts and philosophies, in business, in his book *Cloud Computing Explained: Implementation Handbook for Enterprises* (Recursive Press, 2010). He speaks of the enthusiasm around this subject, which promises economy and flexibility. Any technological novelty is a possibility of becoming a great market differential over competitors for any company. CEO around the world dream of a wise decision on a new subject that makes them a reference in the company, or have saved millions, or brought these same millions of profit. Today, the basis of business competitiveness is through technology. A new process or new software can make differences that can echo for decades. From this concept of distribution of responsibilities on the Internet, the possibility of collaboration was also born. That is, for a single goal several people will be able to apply to participate. Whether providing a solution or giving an idea.

With the help of the availability of centralized services on the internet and with the psychology of collaboration of large groups, crowdsourcing then created a new concept: crowdsourcing. Some authors, such as Clay Shirky (2008), call it "the power to organize without an organization". Another, Jeff Howe in

2009, raises the question: Why does the power of the masses direct the future of organizations? And the American James Surowiecki, who studies behaviors and had his articles published in The New York Times and the Wall Street Journal, argues in the aforementioned "The wisdom of Crowds" (Surowiecki, 2004) that decisions taken in a group are often better than taken by a single person in a group.

Thus, after publications from other areas such as economics and psychology were widely studied and used, the area of systems technology also found space to make use of this new type of process and in this work a review of the current literature was performed, for example, if every technology company can deploy this process to develop its software, what are the big advantages and disadvantages, and also how does it change the way people are managed to be the knowledge providers.

These will be the topics researched in the systematic mapping of literature that is object of this work.

## 2.2 SYSTEMATIC MAPPING PROCESS

The systematic mapping process is used for research, aiming to collect and evaluate evidence on a subject. Different from a usual process of literature review, systematic mapping is developed in a formal and systematic way, as the name suggests. Therefore, it follows a sequence of strict and well defined steps, according to a protocol developed (Biolchini, Mian, Natali, & Travassos, 2005).

There is also the concept of a systematic review, which is different from what was done in this work, since the review provides for evaluating and interpreting all available research relevant to a specific research question, or subject area, or phenomenon of interest, and the mapping allows a broader view of primary studies, making it dependent on mapping to reveal the research evidence. (Kitchenham, 2007).

The core of research is defined through concepts that are completely focused and structured to address a specific question. The methodology, evidence recovery strategies, inclusion and exclusion criteria of the research sources and the question are explicitly defined, so that it is possible for any person, using the same methods, in addition to reproducing the research, to judge whether the methods were either adequate or not (Biolchini, Mian, Natali, & Travassos, 2005) (Huhns, Li, & Tsai, 2015; Stol & Fitzgerald, 2014, Chiua, Lianga, & Turbanc, 2011; Group, Info-Tech Research, 2011).

## 2.3 CRITERIA FOR THE INCLUSION AND EXCLUSION OF ARTICLES

The criteria that were defined for the selection of articles were:

1. Must be articles published in the CAPES Portal;
2. The studies must be written in Portuguese or English;
3. They must contain the search strings defined;
4. Present the subject in the context of software engineering;
5. Must have been published after 2010;

## 2.4 RESEARCH QUESTIONS

The following questions should be answered during the survey:

Q1- What are the definitions of crowdsourcing used in the literature?

Q2 - How does the software development process work via crowdsourcing?

Q3- What are the advantages and disadvantages of using it?

Q4- How to implement and manage it so that software delivery is successful?

## 3 RESEARCH METHODOLOGY

### 3.1 TYPE OF RESEARCH

A systematic mapping was done on software development with crowdsourcing.

### 3.2 DATA COLLECTION AND ANALYSIS

Data collection and analysis was done through indirect documentation, which includes reading data about the phenomenon, as defined by the search criteria and the CAPES Portal articles.

The results will be interpreted according to the context, which should be turned to software engineering, summarized and documented in this document.

### 3.3 DEFINITION OF SEARCH CRITERIA

To perform the systematic mapping, it was necessary to determine some criteria that defined how the articles were found in the CAPES Portal. The search strings or search criteria are defined terms according to the theme that, when well used, will result in specific articles about the subject sought, so that they are analyzed. For this work the following were defined:

- \* crowdsourcing development process
  - \* crowdsourcing developmentprocess
  - \* framework
- Meaning
- \* advantages
  - \* disadvantages
  - \* not medical or medicine related

### 3.4 DATA EXTRACTION METHOD

Having applied the search strings and the criteria for inclusion of the studies, the title and abstract were read. Inclusion was determined by a superficial reading of the article. The included studies were meticulously read and summarized.

## 4 RESEARCH CONTENT ANALYSIS

When applying the strings in the search of the CAPES Portal, more than 70 articles were found that contained the search strings, and from these, 8 articles were extracted that, when reading the abstract of each one, fit the criteria previously defined for the selection. When read, key words were extracted from each article so that this information can be documented

here. The articles deal with diverse contexts in crowdsourcing, such as their definitions, tools, challenges, frameworks and others. The most relevant ones were mostly published after 2012. Table 1 lists the articles analyzed in this systematic mapping.

Code Title A1 Crowdsourcing and open source software participation (Olson & Rosacker, 2012) A2 Outsourcing to an unknown workfor

4.1 RESULTS OF THE RESEARCH

Q1- What are the definitions of crowdsourcing used in the literature?

Of the 8 articles reviewed, all report a definition for crowdsourcing. The association between crowdsourcing in Open Source Software (OSS) was cited by 3 articles, A1, A2 and A6. Focused on documentation, article A5 cited the NumPy tool. The difference and characterizations of outsourcing, offshoring and onshoring and their relations with crowdsourcing were cited by 2 articles, the A2 and A8.

In order to answer the second Q2 question, "How does the software development process via crowdsourcing work?" It is first necessary to define the term, and through this research it was possible to conclude that the definitions for crowdsourcing are diverse, but all follow the same pattern when talking about working community in order to meet a goal. Some call it a phenomenon, others a conceptual innovation. It is also spoken in competition or collaboration, and it is a fact that it is already being used, whether to build software or to document it, and the tendency is that it is increasingly discovered and used.

When associated with the software development process, crowdsourcing is characterized as a methodology in which there are no people or teams hired to do a task: the project is structured and broken into tasks, and these tasks are distributed to online communities where people with the skills inherent in that task will execute it. It's not just like doing outsourcing or offshoring. Since outsourcing means hiring another company to do a service for your company, and offshoring means dealing more geographically, it is a new concept where there is no relationship with another company, and your employees can be anywhere in the universe.

Q3- What are the advantages and disadvantages of using it?

Six articles somehow cited advantages and disadvantages in using crowdsourcing as a tool for software development, such as articles A1, A2, A3, A6, A7 and A8. Raymond classified the development process by teams and without teams because of their great psychological differences, and 3 articles cite this classification, Articles A1, A2 and A6 (Raymond, 2011).

Positive and negative aspects, which answer the second question in the survey "What are the advantages and disadvantages of using crowdsourcing?" Are cited and listed in Table 2. It is important to point out that these suit different niches such as marketing or building a new product, as well as for software engineering.

| Prós                                           | Contras                                     |
|------------------------------------------------|---------------------------------------------|
| O mundo inteiro poderia trabalhar em sua ideia | Pessoas dispersas geograficamente podem não |

## Prós

Funciona quando o tempo é curto e é necessário ter criatividade  
Reputação pessoal aumenta na área onde houve colaboração  
Redução de custos de projeto  
Aproximação dos clientes  
Quando associado a uma infraestrutura em nuvens, tem-se uma estrutura elástica.  
No caso da documentação, simples usuários da ferramenta tornaram-se desenvolvedores  
Possível melhora na qualidade de produtos (exemplo AppStore)

**Tabela 2 :** Prós e Contras do uso de desenvolvimento com *crowdsourcing*

Q4- How to implement and manage it so that software delivery is successful?

Regarding the management format, 6 articles mention that an efficient way to manage the participant resources, which were the A1, A2, A5, A6, A7 and A8, is necessary. While supporting crowds provide information to systems, the cloud infrastructure was cited by only 1 article, the A3.

Questions about how to motivate participants in crowdsourcing projects were cited in the eight articles, all of which referred to the fact that participants are not always looking for financial motivations.

The question of remuneration for this work is widely discussed, since for most of the great projects developed to date, the work was done by people who did not seek financial compensation, but rather recognition, status and importance in their specific community. Wikipedia is cited as the web site that gets a lot of participation from web users, and performs activities that a normal computer could not do. She maintains a regular staff of five who do content reviews, but the texts are all created by the community on the internet, and without any compensation. About the fact that there are no financial rewards, there are citations that indicate that getting people to get a project and developing the idea of grace is strange and may seem wrong. For software engineers, collaborating on widely used systems and having their name attached to such software can be very beneficial.

## 4.2 DISCUSSION OF RESULTS

One feature that is cited and related to standards to meet the tasks required of crowds is the question that all rules and standards for completing tasks must be clearly defined and widely disseminated. Since several people will be completing tasks, these tasks need to, after completion, talk to each other. It is these guidelines that will guide the participants to the same focus, with definitions that must be followed. Even so that new entrants do not have to waste time creating their own patterns. Considering requirements, the OSS community already supports the theory that requirements do not need to be described in detail, and that each developer is responsible for following the "an itch worth scratching" principle, which in a crude translation means "an itch should be a scratch," that is, a requirement must be dismembered and the person responsible for the task must develop the reasoning to serve it, and not only stick to what was requested, always going beyond.

(Raymond, 2011) describes in his book that he perceived a great difference between the development of traditional

software and the one that contrasts cooperation, and classifies the traditional one calling it "cathedral process", since the comparison made is that of this process it is like if it were the construction of a cathedral, where the team of architects works in isolation from other teams, as well as in software development, where analysis, programming and testing teams, for example, work in isolation. In a "bazaar process" cooperation format, teams are decoupled and multiple individuals work on various tasks, there is no division into teams. This description gave an idea that there was no formalization, but recently the crowdsourcing development process has been formalized, and this is clearly exemplified by the Apache and GNOME conferences and projects such as Zope / Plone and Pypy, where there are sprints and regular meetings between team members to coordinate and plan development.

Responding to question 3 of the survey, "How to implement the process and manage it?", There are fronts advocating that, once the OSS developer goes over time gaining marketing experience, he himself can choose what to do and manage his own time. Already on project management tools, they need to be appropriate to this dynamic community, and perform beyond the old well-known tasks, project estimates, planning and development, repository maintenance, among others, also incorporate tasks of ranking people, reputation and task assignment systems. There are tools in the market that are already capable of classifying the skills of professionals in the software development community.

On the market today there are platforms for development using crowdsourcing, such as TopCoder ([www.topcoder.com](http://www.topcoder.com)), CoFundos ([www.cofundos.org](http://www.cofundos.org)) that offer a service for developers to register and create communities online. These platforms were developed with the crowdsourcing model, where developers submitted their solutions to the development tasks. These platforms coordinate the relationship between companies and members. In TopCoder, a coordinator is selected and he is responsible for breaking the project into tasks and distributing them. However, not only these platforms but also the methodology itself, they still need to be better understood and developed to meet complex business projects.

## 5 CONCLUSION

### 5.1 MAIN CONTRIBUTIONS

This work helps in understanding the concept of crowdsourcing and where it can be applied, and also where it is already being used in the market, as well as discussing software development concepts and cases where crowdsourcing has been useful and has generated satisfactory results. The result of the research answers the 4 questions proposed through the analysis of articles searched in the CAPES Portal.

As a definition of the concept of crowdsourcing, which answers the first question of the research (Q1), there is a working community in order to fulfill a single objective.

The process used, which is the subject of discussion of the second question of the research (Q2) is characterized by a methodology in which there are no people or teams contracted to carry out a work, the tasks of a project are published and any interested person who has the ability to perform this task can apply for the job.

One can cite as an advantage the fact that the whole world could work on an idea, also that the process works when time is short and creativity is necessary, when the person involved and interested seeks to improve their reputation in a given community and also when seeking reduce project costs. But there are also disadvantages, for example, the fact that people are geographically dispersed and may not have knowledge about the way their fellow workers work, the quality of work that varies from person to person, the results are not accurate and the lack of can generate demotivation. These factors answer the third research question (Q3).

The fourth and final research question (Q4) deals with crowdsourcing project management, which is considered indispensable for this type of project, but should be conducted so that in addition to managing tasks and resources, it also stimulates interest in people to participate in the project. There are already tools on the market that are able to recognize the skill of the project participants, create a ranking for these people and give the manager knowledge to estimate and know the feasibility of solving tasks of different levels of difficulty.

## 5.2 FUTURE RESEARCH

As much as this new paradigm is in use in many areas of Software Engineering, it is still necessary that some fundamental principles behind software development through crowdsourcing be explored, such as:

1. Which parts of a complex system can be "crowd-sourced"?
2. What are the ultimate goals of using the development process through crowdsourcing?
3. Does using this method guarantee quality and maximize learning?
4. How to reward project participants and what are the costs of a ready solution?