

Crowdsourcing: State of the art

Bernhard Gößwein
some really, really important
information here
e8727334@student.tuwien.ac.at

Robert Kapeller
some really, really important
information here
e106884@student.tuwien.ac.at

David Riepl
some really, really important
information here
e0625016@student.tuwien.ac.at

ABSTRACT

The paper should provide an overview about the scientific state of the art in crowdsourcing. Note that the paper is not the documentation of your tool – it should discuss scientific papers related to this topic in the style of a seminar paper. Good starting points for finding related scientific papers are the sources cited in this text, Google Scholar 6, IEEE Explorer 7 or the ACM Digital Library 8. Use the ACM 'tight' conference style 9 (two columns), and keep it brief (3 pages). You do not necessarily need to install a LaTeX environment for this - you can use writeLaTeX 10, a collaborative paper writing tool as well.

General Terms

Crowdsourcing, some, more terms

Keywords

Crowdsourcing, some, more, keywords

1. INTRODUCTION

Write something useful about how crowdsourcing has been developed over time, history and the like. Come then to the current state of the topic and give a brief overview about the (possible) future.

2. DEFINITION

As also discussed in Section 3 the term “Crowdsourcing” has been introduced by Jeff Howe[5]. Howe itself described crowdsourcing as

[..]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.

Daran C. Brabham first defined *Crowdsourcing* in an scientific article [2] as:

[..]an online, distributed problem-solving and production model.

but this definition fails on two accounts: first - as history tells us - Crowdsourcing is not bound to online sources and second and more importantly it is not always used for problem solving.

Another rather older definition of the term has been stated in 2009 by Paul Whitley in his paper *Crowdsourcing and Its Application in Marketing Activities*[7]:

Crowdsourcing is a newly developed term which refers to the process of outsourcing of activities by a firm to an online community or crowd in the form of an ‘open call’. Any member of the crowd can then complete an assigned task and be paid for their efforts.

If we take for example wikipedia or the development of Linux into consideration as being a crowdsourcing effort as well, this definition is no longer true as those participants are obviously not paid and no *firm* is involved and no payment is done. It also does not take into consideration, that the effort of the crowd can be acquired implicitly as well (see Section 4).

As *crowdsourcing* is a relatively new concept this leads to the fact that it may be identified with many internet-based collaborative work initiatives and in fact it is not always clear where to draw a line between crowdsourcing efforts and something else. Consequently many definitions of the term *crowdsourcing* have been developed over time. Enrique Estellés-Arolas studied more than 40 papers with varying definitions of the term [4]. His definition based on his analysis tries to cover all types of crowdsourcing initiative:

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 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 that what the user has brought to the venture, whose form will depend on the type of activity undertake

This characterization however misses the fact, that crowdsourcing is not necessarily bound to an *online activity*. As we will show in Section 3 crowdsourcing existed before someone thought about such thing as the “WWW”.

3. HISTORY OF CROWDSOURCING

The term *crowdsourcing* has been coined by the american journalist Jeff Howe [5]. It has first been mentioned in the article *The Rise Of Crowdsourcing*, published in the Wired Magazin¹.

In this article Howe describes that and why crowd sourcing may be the next big thing in the internet 2.0. He connects the term *outsourcing* (giving work to well known and specialized group of persons) with *crowdsourcing*, meaning that the group of participants is open. The transition from outsourcing to crowdsourcing is described by examples like a site named iStockphoto², which raised by a image-sharing community. iStockphoto's micro-payment system allows a large community of more or less professional photographers to sell their pictures. Although this might not be a crowdsourcing system as it has been defined it certainly leads to it.

However crowdsourcing is much older and is not necessarily related to the internet. In the past crowdsourcing has often been used as a competition in order to discover a solution. The following examples are by no means complete and shall just show that crowdsourcing type of work has been around for a long time before the term has been established and even before the internet exists.

1714: The longitude price

In 1714 the British Government offered to a solution to a problem named *The Longitude Problem*³. Sailing was very dangerous at this times. One of the problems was to measure the current longitude. Eventually John Harrison won the main price by inventing the chronometer. Several other persons also benefited from the offer.

It shall be mentioned that quite similar offers have been done even earlier, in the late 16th century but by then nobody could solve the problems.

1936: Toyota Logo Contest

In 1963 Toyota held a contest to design a new logo. They received about 25.000 entries. The winning one was the well known three ellipses.

2001 - 2005: Youtube, Wikipedia ...

With the rise of the internet collaborative sites like youtube and wikipedia rose as well. The connection to the field of

crowdsourcing is obvious: a large group of people invest time and effort to create a larger whole.

2006 ff: Crowdsourcing

As already mentioned, the term *crowdsourcing* has been stated in 2006 by Jeff Howe. From mid 2000 until now a wide large number of different crowdsourcing sites has been established covering different domains. One can in fact see an explosion in crowdsourcing related web sites. www.crowdsourcing.org lists more than 2500 sites which are related to crowdsourcing in different categories and the list is constantly growing.

The next chapter will cover some of the sites in an exemplary way.

3.1 Prominent crowdsourcing sites

In general crowdsourcing means that a group of *workers* participate in finding a solution to a given problem, mostly by solving small parts of the problem. There are several web sites online which provide the technical and logistical infrastructure to support this kind of problem solving. The provided list is by no means complete but shall rather give a brief overview. Different sites cover different more or less specialized aspects of crowd sourcing. In Section 4 we will list and explain some of the most common crowdsourcing types.

Freelancer.com⁴

This site offers online jobs to freelancers who bid on projects or problems posted by companies or individuals on a set price. The kind of problems to solve range from programming tasks to design projects. Similar web sites are for example ScriptLance⁵, Elance⁶ or guru⁷

CrowdSpring⁸

A site specialized to graphic design projects. Similar sites are: 99designs⁹ or hatchwise¹⁰.

Utest¹¹

This site offers software testing, that relies on crowdsourcing. Similar pages are: UserTesting¹², Feedback Army¹³ (for web sites) and others.

Amazon Mechanical Turk¹⁴

One of the best known crowdsourcing sites. Best suited for small and simple tasks like finding or extracting of certain information, tagging pictures... Each task is really simple and donated by just a few cent.

InnoCentive¹⁵

⁴<http://www.freelancer.com/>

⁵<http://www.scriptlance.com>

⁶<http://www.elance.com>

⁷<http://www.guru.com>

⁸<https://www.crowdspring.com>

⁹<http://99designs.com/>

¹⁰<http://www.hatchwise.com>

¹¹<https://www.utest.com>

¹²<http://www.usertesting.com>

¹³<http://feedbackarmy.com>

¹⁴<https://www.mturk.com/mturk/welcome>

¹⁵<http://www.innocentive.com/>

¹<http://www.wired.com>

²<http://www.istockphoto.com>

³http://en.wikipedia.org/wiki/Longitude_prize/

On the contrary to the Mechanical Turk InnoCentive aims to solve complex and more comprehensive questions. Comparable to Innovation Exchange¹⁶.

Those are just a few of many, many crowdsourcing sites in the internet. A comprehensive list can be found at www.crowdsourcing.org

4. TYPES OF CROWDSOURCING

It turns out that a decent definition and typification crowdsourcing systems is not straight forward. As we have already seen the variety of CS-systems range from collaborative sites like wikipedia to micro-task sites like the Amacon Mechanical Turk. One could say that CS is to be defined as a system where a large group of collaborators participate to create a defined artefact, as it is the case at Wikipedia or Linux. On the other hand, this would ignore systems, that for example tasks (and pays) people to find certain information on the web (or else where). Obviously no lasting “artifact” is created in this case. So it seems not appropriate to distinguish the type of CS according to the target problem alone. Doan et. al [1] mainly distinguishes between *explicit* and *implicit* systems.

- **Explicit** systems let users collaborate explicitly, that is, users register, accept tasks, work on it and submit their solutions. This is the most obvious category and does not need further explanation.
- The intention of **Implicit** systems is, to motivate the user to collaborate without explicitly having the task in mind. Sometimes the user does not even know, that a current interaction might be used to gain additional information. For example Amazon-users implicitly solve the task to rate the usefulness of a product by rating it (same is true for purchaser).

Doan adds 8 more distinguishing dimensions: type of target problem, how to recruit and retain users, what users can do, how to combine their inputs, how to evaluate them, degree of manual effort, role of human users, standalone versus piggyback architectures. We will not go into detail for the other dimension. Please refer to the paper for details.

Daren C. Brabham[3] uses a rather problem based typology. He categorizes CS to:

- **Distributed Human Intelligence Tasking:** The crowd processes and analyses information.
- **Knowledge Discovery & Management:** The crowd finds and assembles information.
- **Broadcast search:** The crowd comes up with an objective solution of a problem.
- **Peer-Vetted Creative Production:** The crowd comes up with an subjective solution of a problem (e.g. design...)

Jeff Howe also stated categories of crowdsourcing. He attracts the problem on a higher level of abstraction. Howe

¹⁶<http://www.innovationexchange.com/>

proposes the categories *crowdvoting*, *crowdfunding*, *microwork*, *creative crowdsourcing*, *wisdom of the crowd* and *inducement price contests*.

Last but not least as an example I want to mention the web site www.hongkiat.com, which is dedicated to *Design, Inspiration, Techonolgy*, as it states. In his article *Crowdsourcing: Pros, Cons, And More*¹⁷ Darren Stevens simply distinguishes between *Crowdfunding*, *Crowdsourced design* and *Crowdwisdom*.

As one can see, different views to the problem create different solutions to it. Maybe one shall let the crowd decide on how to categorise crowdsourcing.

5. CHALLENGES WITH CROWDSOURCING

The idea of crowdsourcing is, that one can receive better quality results, since several people offer their best ideas, skills and support. It also let the crowdsourcer select the “best entries”. instead of receiving just one result from a single contractor. It is expected, that the results are delivered quicker and with better quality than with traditional methods. However this is not always and automatically the case. Workers are typically paid below market price and this leads to less motivation to provide good quality. Additionally worker could just cheat. This is possible especially when it comes to micro-work like structures as promoted by for example Amazons Mechanical Turk.

Doan et. al. discuss the following main challenges[1]:

How to recruit and retain users:

Doan identifies this as one of the most important problems. He outlines five major solutions, I want to discuss three of them briefly:

One would be to *require* people to contribute but this only possible if the crowdsourcer hires people which in turn somehow contradicts the idea of crowdsourcing. Another possibility to meet this challenge is to pay for the effort. Most of the web sites we showed in Section 3 follow this approach. Problem is, that for very simple the payment is very low (e.g. Mechanical Turk), so that the motivation to provide good work might be low.

The third possibility is to just *ask people* to contribute. This is also a widely used method but can only be established in fields where “workers” are interested in doing the work. For example open source development works this way¹⁸.

How to combine contributions:

Especially microwork sites like Amazon Mechanical Turk distribute a certain task among many people. All of them only solve a small part of the overall problem or task and submit their solution to this part. Sometimes the combination of the different submissions is quite easy and straightforward. For example the rating system of amazon just computes a medium of all the ratings and this might be enough for this kind of problem. There is no need to combine the outcomes of several users and to combine them to a greater whole. But this is no longer true for microwork sites. The challenges here are (among others): how to make sure, that the each part of the problem is tackled by a minimum num-

¹⁷<http://www.hongkiat.com/blog/what-is-crowdsourcing/>

¹⁸And it works quite well in many cases!

ber of workers? How to combine them to a final solution? An additional problem is, that different workers might work on the same part of the problem but with different outcome. How to distinguish between it? This types of problems arises whenever workers act collaborative. Most of the CS systems simply do not combine or rate contributions at all but let the crowdsourcer decide on how to interpret the single results.

How to evaluate users and contributions:

Many sites use But still there might be problems: how to recognize and eliminate notoriously bad ratings from users who just want to mess around?

Even worse is the scenario at microwork sites. As we already mentioned, the payment is quite

How well users perform depends on how motivated they are to work on a problem on the long run. Man-Ching Yuen et. al propose a method to support workers by selecting tasks on crowdsourcing platforms easily and effectively using a special task matching algorithm[6].

6. FUTURE OF CROWDSOURCING

7. CONCLUSION

Some final words about crowd sourcing in general and the future of it.

8. REFERENCES

- [1] R. R. AnHai Doan and A. Y. Halevy. Crowdsourcing systems on the world-wide web. *Communications of the ACM*, 54(4), April 2011.
- [2] D. Brabham. Crowdsourcing as a model for problem solving: An introduction and cases. *Convergence: The International Journal of Research into New Media Technologie*, 14(1), 2008.
- [3] D. C. Brabham. *Crowdsourcing*. MIT Press, 2013.
- [4] E. Estellés-Arolas. Towards an integrated crowdsourcing definition. *Journal of Information Science*, XX(X), 2012.
- [5] J. Howe. The rise of crowdsourcing. <http://www.wired.com/wired/archive/14.06/crowds.html>, June 2006. [Online; accessed 10-January-2014].
- [6] I. K. Man-Ching Yuen and K.-S. Leung. Task matching in crowdsourcing. *IEEE International Conferences on Internet of Things, and Cyber, Physical and Social Computing*, June 2011.
- [7] P. Whitla. Crowdsourcing and its application in marketing activities. *Contemporary Management Research / CMR / ISSN 1813-5498*, 5(1), 2009.