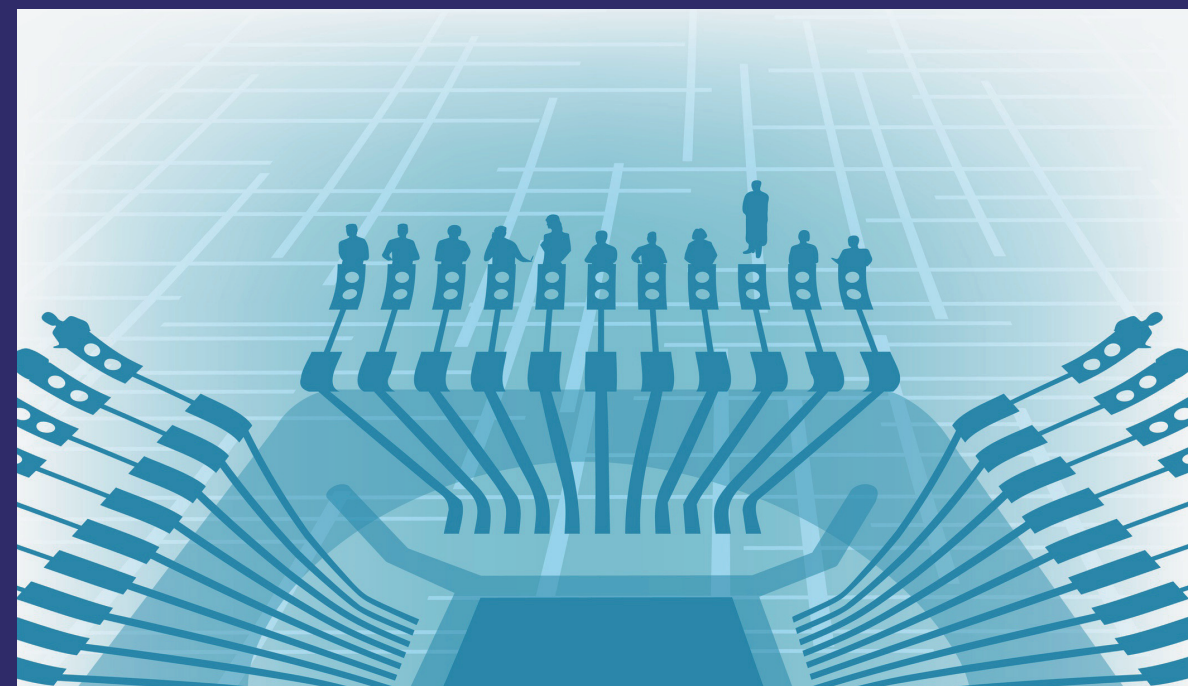


Extension of navigation applications to indoor use is a current research focus. There exist several concepts; however, all share the problem of ascertaining infrastructure data. Further, maintenance of the data is challenging. "Crowdsourcing" is an ansatz to outsource these tasks from the service provider to the user. There exist different derivates of crowdsourcing; however, no commonly accepted definition. An attempt of a definition is provided, resulting in a crowdsourcing taxonomy. Utilizing the taxonomy conceived, an architecture extension compliant to German data protection laws is introduced in order to support different classes of crowdsourcing. The extension conceived centres the "Indoor Navigation Server Access Network Entity" (INSANE) which acts as a proxy for crowdsourcing communication, effectively concealing the crowdsourcers' identity from the crowdfunders. Addressing security and privacy issues, a principal user management as well as concealing techniques are applied. Avoiding a single point of failure at the INSANE, utilisation of a distributed hash table is proposed. The concept is exemplarily implemented and proven to be of good performance and scalability.

Towards a Crowdsourcing Architecture



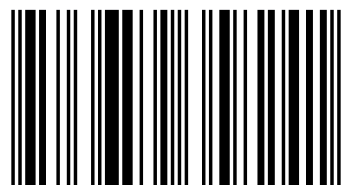
Tenshi Hara

Towards a reliable Architecture for Crowdsourcing



Tenshi Hara

Dipl.-Inf. Tenshi Hara studied Physics and Computer Science at TU Dresden (Saxony, Germany). He is currently employed as a research and teaching assistant by the Computer Networks Group at TU Dresden. His research and teaching activities focus on Computer Networks and E-Learning.



978-3-659-35824-1

Hara



Impressum / Imprint

Bibliografische Information der Deutschen Nationalbibliothek: Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Alle in diesem Buch genannten Marken und Produktnamen unterliegen warenzeichen-, marken- oder patentrechtlichem Schutz bzw. sind Warenzeichen oder eingetragene Warenzeichen der jeweiligen Inhaber. Die Wiedergabe von Marken, Produktnamen, Gebrauchsnamen, Handelsnamen, Warenbezeichnungen u.s.w. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutzgesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Bibliographic information published by the Deutsche Nationalbibliothek: The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at <http://dnb.d-nb.de>.

Any brand names and product names mentioned in this book are subject to trademark, brand or patent protection and are trademarks or registered trademarks of their respective holders. The use of brand names, product names, common names, trade names, product descriptions etc. even without a particular marking in this works is in no way to be construed to mean that such names may be regarded as unrestricted in respect of trademark and brand protection legislation and could thus be used by anyone.

Coverbild / Cover image: www.ingimage.com

Verlag / Publisher:

LAP LAMBERT Academic Publishing

ist ein Imprint der / is a trademark of

AV Akademikerverlag GmbH & Co. KG

Heinrich-Böcking-Str. 6-8, 66121 Saarbrücken, Deutschland / Germany

Email: info@lap-publishing.com

Herstellung: siehe letzte Seite /

Printed at: see last page

ISBN: 978-3-659-35824-1

Copyright © 2013 AV Akademikerverlag GmbH & Co. KG

Alle Rechte vorbehalten. / All rights reserved. Saarbrücken 2013

3 CROWDSOURCING

'You can't rule the world in hiding. You've got to come out on the balcony sometimes and wave a tentacle.'

The Fourth Doctor (Thomas Stewart Baker)



CROWDSOURCING

Crowdsourcing presents itself to be a rather young concept in the field of computer science, emerging in the late 1990s or early 2000s respectively. For example, the *SETI@home* project was released to the public on 17 May 1999, making the concept of a distributed calculation known to the public. As crowdsourcing is such a young concept, currently there exists no commonly agreed upon definition of what crowdsourcing actually is. There exists a vague consensus on basic ideas, so intuitively most people agree on what the ought to believe crowdsourcing is, but no agreeable formal definition has been accepted in general. Even though, crowdsourcing can be divided into two types: explicit crowdsourcing and implicit crowdsourcing; once again, not clearly defining what both of them are. Nevertheless, this chapter shall introduce an attempt at giving a definition of crowdsourcing as well as a taxonomy of crowdsourcing. Different derivatives shall be introduced and categorised according to the defined taxonomy. Finally, a context to the similar term ‘social sensing’ shall be established.

3.1 ATTEMPT AT A DEFINITION

Closely following Howe, Estellés-Arolas and González-Ladrón-de-Guevara [EG12, How06] in context of Quinn and Benderson [QB11] a common consensus on the definition of crowdsourcing shall be established. Even though it is hard to find a consensus as definitions vary broadly, it is imperative to agree on the common ground in order to build up an usable concept of crowdsourcing, especially a taxonomy of explicit crowdsourcing versus implicit crowdsourcing.

Comment

*Sometimes the equivalent term ‘crowd sourcing’ can be found in sources.
The author of this work attempts to consequentially use ‘crowdsourcing’.*

Currently, there is no generally accepted definition of crowdsourcing, only a large number of widely varying definitions. In order to have a starting point, Definition 3.1.1 [EG12] shall be given, which – according to self-made state-

ments – Estellés-Arolas and González-Ladrón-de-Guevara came up with after studying 40 definitions of crowdsourcing.

Definition 3.1.1 – Crowdsourcing

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 undertaken.

Thus, Crowdsourcing is a distributed online or offline process involving a flexible group of contributors (the ‘crowd’) and tasks to be (partially) completed by the contributors. The flexible group of contributors is not well defined and is subject to dynamic changes in composition, structure and taskability. Hence, crowdsourcing can be categorised as a distributed problem-solving model with different algorithms. The entity making use of the crowd and providing the task(s) is considered the ‘crowdsourcer’.

In most concepts – as is true for this work – the crowd is composed of humans which are at the same time users of a system which itself is the crowdsourcer. Often, the solutions submitted to the crowdsourcer are composed only of evaluable/analysable data, which are thereafter owned by the crowdsourcer, making the crowd take the role of service providers.

Often, either the crowd as a total or individual contributors are compensated monetarily¹⁸, with prizes¹⁹ or benefits²⁰, or with recognition²¹. In

¹⁸ Test subjects are often monetarily compensated in the field of pharmaceutical research.

¹⁹ Enquiries often offer the possibility of participating in a competition after submitting query-forms, promising equal chances of winning prizes to all participants.

²⁰ Dropbox Inc. offered users participating in the beta-test of their automated photo-upload function additional 500MB of online-storage, which would remain even after the beta-test concluded, for each 500MB of photos uploaded. <http://forums.dropbox.com/topic.php?id=53104> – Accessed 1 June 2012

²¹ Search for Extra Terrestrial Intelligence at home (SETI@home) offers their users the possibility to download and print

other cases, there is no reward as the crowd might not even know they are participating in the crowdsourcing process, e.g. Facebook Inc. gathers information on persons who are not even members of the community by simply analysing any data community members upload and putting them into relation; most of the Facebook users are not aware that they contribute personal data of *dégagé* thirds, such as telephone or address data, even though those thirds might give their consent to such contribution. The crowdsourcers generally benefit by large numbers of solutions/information being provided inexpensively or even for free; this of course, only if considered a crowd large enough.

3.2 TAXONOMY OF CROWDSOURCING

Different derivatives of crowdsourcing include, but are not limited to:

- *Crowdvoting* is the concept of a website gathering votes on a certain topic, e.g. which picture provided by a community shall be awarded 'picture of the year'.
- *Crowdwisdom* is the concept of collecting vast quantities of information, organising this information and deducing a commonly agreeable recollection of the picture behind the information. Wikipedia can be considered the example for crowdwisdom, as users contribute new information, update existing information or delete information considered irrelevant or false. As the crowd has different backgrounds²², the commonly agreed recollection can be considered broad and well founded.
- *Crowdfunding* is the concept of collecting vast financial resources by involving a crowd out of which each contributor only contributes a small monetary support. In general, but not as a rule, the goal of the fund-raiser is clearly defined. Kickstarter may be the example for crowdfunding as it is the biggest website for funding creative projects, having raised over 100 million USD. The resources contributed are only allocated to the crowdsourcer if the defined goal is reached, otherwise the resources are returned to the crowd.

out a certificate of appreciation for certain levels of credit points. http://setiathome.berkeley.edu/cert_print.php
– Accessed 1 June 2012

²² E.g. there might be veterinarians, truck-drivers, police force, etc. involved in maintaining an article about roadkills.

- *Crowdpurchasing* is the concept of leveraging the collective purchasing power of the crowd in order to attain a reasonable discount or even the best price a dealer or manufacturer is willing to offer in general. Letsbuyit.com is an example for this concept, where users are presented with several levels of achievable discounts, depending on the size of the crowd involved.
- *Crowdworking* is the concept of distributing parts of a dividable task to the crowd. In general, but not as a rule, the tasks at hand are considered challenging for a computer, but easy for a human. In 2006, the National Aeronautics and Space Agency asked people to review images returned by the Stardust project on NASA's website²³ in order to find images with 'interesting' dirt.

A more comprehensive look at some of the derivatives can be found in chapter 3.

Independent of the derivatives, there exist two concepts of crowdsourcing: explicit crowdsourcing and implicit crowdsourcing. The definition of what 'explicit' and what 'implicit' are do vary, making a general definition almost impossible. The two most commonly agreed distinctions of the two are:

- Explicit crowdsourcing lets users cooperate and actively contribute to the crowdsourcing process, while implicit crowdsourcing means that users solve a problem as a side effect of something else they are doing, and
- Explicit crowdsourcing gives users awareness of the crowdsourcing process and lets them decide when to participate in the crowd sourcing process, while implicit crowdsourcing keeps the crowdsourcing process transparent to the users, requiring no user interaction at all.

As for the MapBiquitous project and the creation of the original thesis and G. Bombach's assignment paper [Bom12], both distinction concepts shall be united, maintaining the agreed concepts of implicit crowdsourcing. Clearly, both distinctions follow independent differentiators: the immediacy of the crowdsourcing process on the one hand and the crowd's awareness of the crowdsourcing process on the other hand. As both differentiators are orthogonal to each other, the following taxonomy shall be proposed:

²³ <http://stardust.jpl.nasa.gov/home/index.html> – Accessed 1 June 2012

Definition 3.2.1 – Taxonomy of Crowdsourcing

There exist four types of crowdsourcing which can be placed within a taxonomy build on awareness and immediacy. Awareness is the concept of the crowd being aware of the crowdsourcing process taking place, whereas immediacy is the concept of acquiring required information either directly (strong correlation between collected data and derived information) or indirectly (loose correlation between collected data and derived/computed information). Hence, there are four types of crowdsourcing:

- **Aware Direct Crowdsourcing (ADC)**
Aware crowdsourcing directly correlated to the data required.
- **Unaware Direct Crowdsourcing (UDC)**
Unaware crowdsourcing directly correlated to the data required,.
- **Aware Indirect Crowdsourcing (AIC)**
Aware crowdsourcing loosely correlated to the data required.
- **Unaware Indirect Crowdsourcing (UIC)**
Unaware crowdsourcing loosely correlated to the data required.

The following conventions shall be valid: As *aware direct crowdsourcing* is explicit in awareness as well as immediacy, this type of crowdsourcing shall be labelled 'explicit crowdsourcing' within this work, whereas *unaware indirect crowdsourcing* is implicit in awareness as well as immediacy, so it shall be labelled 'implicit crowdsourcing' within this work.

The entire taxonomy as defined is depicted in Figure 3.2.2. Above introduced abbreviations (ADC, UDC, AIC, UIC) originate in the differentiators explicitness or implicitness, for which the first letter represents the awareness differentiator, and the second letter represents the immediacy differentiator.

Independent of the distinction between implicit and explicit crowdsourcing, commonly agreed upon (and hence often used), implicit crowdsourcing can take two forms. These forms of implicit crowdsourcing cannot clearly be mapped on either differentiator, but shall be mentioned here anyway, as there exists an obvious tendency towards the awareness differentiator, so – even though not fitting exactly – they can be mapped accordingly.

- Standalone implicit crowdsourcing allows the crowd to contribute as a side effect of the task they are actually solving or of the usage of the

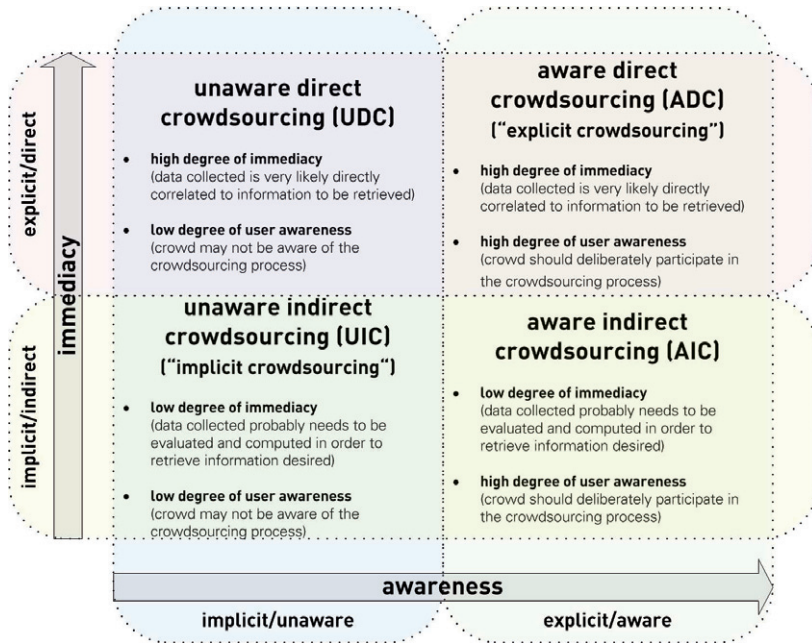


Figure 3.2.2: The taxonomy of crowdsourcing as to be used within the original thesis

crowdsourced system, whereas

- Piggyback implicit crowdsourcing takes information deducted from the crowd's general feedback and acquires data from them²⁴.

The tendency to the context of awareness is obvious, but it is not clear enough to justify naming *unaware direct crowdsourcing* 'standalone crowdsourcing' and *unaware indirect crowdsourcing* 'piggyback crowdsourcing' within the taxonomy.

Inexpediently, the four types of crowdsourcing may not always be clearly distinguished, making the crowdsourcing process present itself as a hybrid, which shall simply be called 'hybrid crowdsourcing'.

Corollary 3.2.3 – Refinement of the crowdsourcing taxonomy

The taxonomy of crowdsourcing implicitly includes hybrids and hence may be refined by adding information upon the awareness and immediacy:

- **Unaware Crowdsourcing**
The hybrid of UDC and UIC.
- **Aware Crowdsourcing**
The hybrid of ADC and ACI.
- **Direct Crowdsourcing**
The hybrid of UDC and ADC.
- **Indirect Crowdsourcing**
The hybrid of UIC and AID.
- **(Entirely) Hybrid Crowdsourcing**
The hybrid of all four types of (non-hybrid) crowdsourcing.

The refined taxonomy enriched by hybrids is depicted in Figure 3.2.4.

²⁴ Google stores a cookie on users' computer while surfing to Google's search engine. When later surfing to affiliated partner websites, the cookie can be used to trace users and deduct behavioural patterns, which in turn can be used to optimise advertisement cashback.

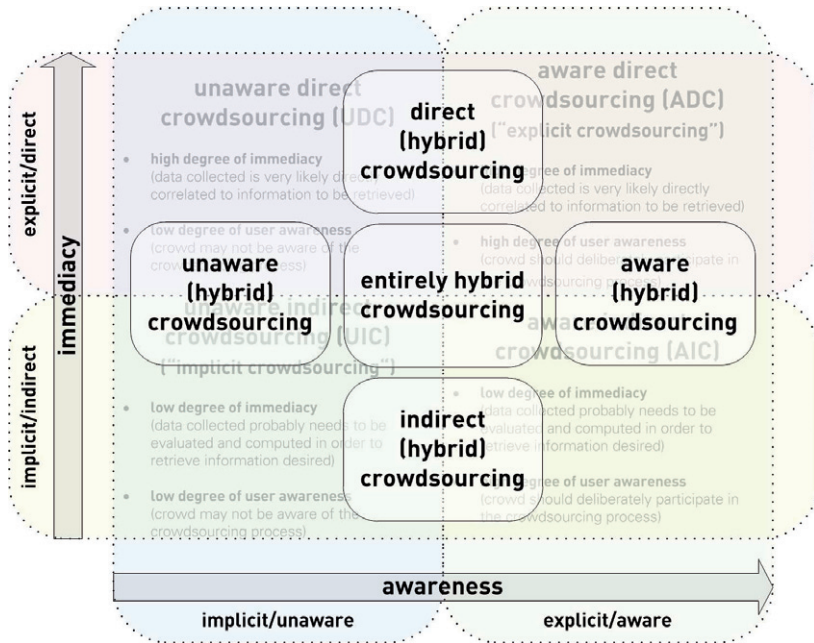


Figure 3.2.4: The taxonomy of crowdsourcing refined with hybrids

Corollary 3.2.5 – Further refinement of the crowdsourcing taxonomy

The refined taxonomy of crowdsourcing includes further hybrids beyond the hybrids introduced in Corollary 3.2.2 ('mostly unaware slightly indirect hybrid crowdsourcing', 'slightly unaware mostly indirect hybrid crowdsourcing', etc.).

It shall be noted, that data protection issues (refer to section 6.1) may arise. Hence, users should be asked whether they wish to contribute collected data, or not. This would automatically make the entire crowdsourcing process aware to the crowd, so the concept of awareness should be refined.

Definition 3.2.6 – Crowd Awareness

Awareness within a crowdsourcing process arises iff the crowd is regularly confronted with the fact that crowdsourcing is taking place.

In the sense of Definition 3.2.4, the crowd is unaware of the crowdsourcing process if they are asked for permission once at the beginning of the crowdsourcing process. This is very important when considering that humans tend to not read the fine-print when using software, etc.

With the so defined taxonomy, the above mentioned derivatives can be exemplarily categorised into the taxonomy types ADC (explicit crowd~) and UIC (implicit crowd~), without limiting to either of them.

• Crowdvoting:

- *Explicit crowdvoting*: Being presented a selection of choices, users pick their favourite. – The crowdsourcer is the designer of the selection trying to grasp the general opinion of a set of users, whereas the crowd is the set of users wanting to know, what thoughts others have on the topic in question.
- *Implicit crowdvoting*: Analysing the sales of products, a shop can create a ranking of their products. – The crowdsourcer is the shop wanting to sell high quantities of favourable products, whereas the crowd is the total set of customers simply buying the products.

• Crowdwisdom:

- *Explicit crowdwisdom*²⁵: In light of the annular solar eclipse that

²⁵ Information on this example provided by an article of the NHK World News on 25 May 2012.

took place on 20 May 2012²⁶, astronomers all over Japan observed the border of the annulus in order to find the exact²⁷ line of annular observability. With the information collected the exact circumference of the sun at that time could be calculated. – The crowd-sourcers were the scientific community of Japan wanting to determine the exact circumference of the sun at that time, whereas the crowd was the set of all observers sending in information on annular observability.

- *Implicit crowdwisdom*: Hardware manufacturer often offer a cash-back system, awarding end-users with financial benefits for returned hardware at the end of the hardware’s lifecycle. – This is of course an example for explicit crowdsourcing when considering only the cashback, but considering the manufacturer as crowd-sourcer wanting to acquire knowledge upon hardware quality or durability, the customers sending in their hardware can be considered the crowd implicitly providing the knowledge on quality or durability.

- **Crowdfunding:**

- *Explicit crowdfunding*: A start-up wanting to start production of a product may call for supporters, offering them a reduced special price when the product is finally available. – The crowdsourcer is the start-up requiring funding support, whereas the customers are the crowd providing the funds, expecting the reduced price as an award.
- *Implicit crowdfunding*: Users regularly visit a website with advertisements placed on it. – The crowdsourcer (explicit and implicit) is the website’s owner receiving money for each advertisement displayed and/or clicked, whereas the explicit crowd is the set of users clicking the advertisements (creating payable click impressions), and the implicit crowd is the set of users visiting the website and having the advertisement only displayed in their browser (creating payable view impressions).

- **Crowdpurchasing:**

- *Explicit crowdpurchasing*: A group of users may form an aggregation of purchasers in order to achieve the optimal price from a manufacturer. – The crowdsourcer is the manufacturer demanding

²⁶ UTC; 21 May 2012 in Asian time; eclipse 58 of Saros-cycle 128

²⁷ In some areas in Japan, observation units were set up every 180 metres.

to sell large quantities of their products, whereas the crowd is the aggregation of purchasers wanting to reach the price minimum.

- *Implicit crowd purchasing*: A supermarket sells articles at a fixed price, but buys them at varying prices. – The crowdsourcer is the supermarket willing to win the lowest purchase price, the crowd is the set of all customers buying the article from the supermarket. Depending on sells, the order amount varies; hence a purchase price fluctuation may occur. The supermarket may support this by advertising the article in order to push sells.

- **Crowdworking**:

- *Explicit crowdworking*: An academic work is put to public discussion. – The crowdsourcer is the author of the work asking for feedback, whereas the crowd is the set of all readers reading and commenting on the work.
- *Implicit crowdworking*: People are invited to test a new hiking course for free. – The crowdsourcer is the owner of the hiking course, whereas the explicit crowd is the set of people using the hiking course for free, while the implicit crowd is the same set of people compacting the grounds and foundations of the newly created hiking course.

Once again, it shall be mentioned that a more comprehensive look at some of the derivatives can be found in chapter 3.

3.3 SOCIAL SENSING

An approach at gathering information focussed on humans and their behaviour is 'social sensing'. The sole sources of information gathered are the human being and its surroundings. Humans themselves as monitors [ASS⁺10] gathering strongly correlated measurands as well as the humans' social interactions as monitors gathering loosely correlated measurands [MCLP10, Tel07] allow information retrieval in a social context. The approach is well distributed and crowd-based, so it is fair to assume it to be at least a derivative of crowdsourcing, which – as a reminder – is not limited to only humans.

Looking at the way social sensing is described to work in general (e.g.

in [ASS⁺10, MCLP10, Tel07]), a striking similarity to crowdsourcing can be observed. A party is interested in information, which shall be provided by observed (and distributed) humans. This model is an exact match to the crowdfunder/crowd-relation of crowdsourcing (refer to Figure 3.3.1).

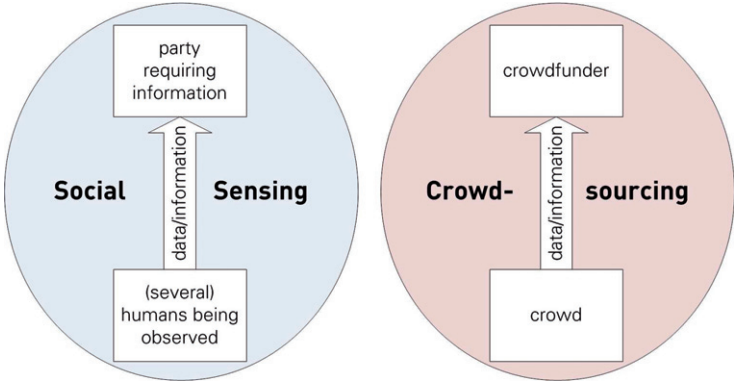


Figure 3.3.1: Comparison of social sensing and crowdsourcing; striking similarity

Further, the same taxonomy as introduced in Definition 3.2.1 can be applied accordingly. On the one hand, when social sensing takes place, it is possible to have the humans either be aware of the social sensing, or not be aware of it. Hence, following Definition 3.2.1, social sensing should also be able to be typified along the awareness differentiator, leading to *aware* and *unaware* social sensing. On the other hand, data acquired by social sensing can be strongly correlated to the information required²⁸, but also weakly correlated to the information required²⁹. Hence, also the immediacy differentiator as introduced in Definition 3.2.1 can be applied, leading to *direct* and *indirect* social sensing.

Taking all the similarities into consideration, it should be fair to proclaim that social sensing is in fact crowdsourcing, but limited to humans and data/information strongly correlated to humans and their social surrounding.

²⁸ E.g. the required information could be (in absolute numbers) an answer to the question, ‘How many female friends does a man aged 30 have in average?’

²⁹ E.g. the required information could be an answer to the question, ‘Does a larger circle of friends reduce the risk of allergic coryza?’

Comment

For the remainder of this work, it shall be assumed that the made proclamation is valid. Nevertheless, limiting the scope to social sensing would be unjust; hence, the broader crowdsourcing will be used in chapter 7.