

Chapter 2

Crowdsourcing Systems

While isolated examples for crowdsourcing approaches can be found throughout the centuries (Surowiecki 2005), the development of the Internet and Web 2.0 technologies has drastically leveraged the reach and the efficiency of connecting with large groups of people. Ever since, researchers and practitioners from various fields have begun to explore how to tap into the potential of crowds in a multitude of contexts.

It is only recently that such approaches have been considered as related phenomena. The term “crowdsourcing” (Howe 2006) has since attracted a lot of attention, even if its exact understanding still varies (Estellés-Arolas and González-Ladrón-de-Guevara 2012). Relevant research is rooted in fields as diverse as computer science (Doan et al. 2011) and management (Schenk and Guittard 2011) and is applied in many other areas that have discovered crowdsourcing as a potentially useful approach. Consequently, research questions mostly center on specific use cases and individual aspects of crowdsourcing. This fragmentation of the research landscape makes it difficult to establish a comprehensive knowledge base and to provide structured guidance for the design of crowdsourcing solutions.

Inspired by systems theory, this chapter introduces a more integrated, socio-technical perspective to study the information systems that realize Web-based crowdsourcing efforts. The resulting framework provides a conceptual structure to channel and relate crowdsourcing research. In particular, it can be used to gain a deeper and comparative understanding of different types of crowdsourcing systems and thus guide the systematic analysis and design of their components, e.g., with respect to task recommendation.

The following sections summarize the product of a series of previous studies (Geiger et al. 2011a, b, 2012). Section 2.1 first establishes a socio-technical definition of crowdsourcing systems and their individual components. Section 2.2 then proceeds to analyze the organizational functions of crowdsourcing systems, identifying four system archetypes with distinct characteristics. Section 2.3 briefly discusses the usefulness of the resulting typology.

2.1 A Socio-Technical Perspective

A system is a set of interrelated elements or components that work together to achieve an overall objective. Systems have a clearly defined boundary and exist as components or subsystems of other systems, also called the environment of a system. Most systems are open, i.e., they interact with their environment via interfaces. Systems are ubiquitous and can, for instance, be of biological, technical, or social nature (Ackoff 1971; Bertalanffy 1972; Churchman 1968).

Information systems (IS) are subsystems of an organizational system that provide an organization with information services needed for operations and management (Davis 2000; Falkenberg et al. 1998; Heinrich et al. 2011). Understandings of the term *information system* differ widely in the extent to which they emphasize social vs. technical concerns (Alter 2008; Carvalho 2000; Falkenberg et al. 1998). While some academics hold a primarily technical view (e.g., Ein-Dor and Segev 1993), the majority of the IS community view information systems as socio-technical systems that integrate human and machine components (Davis 2000; Heinrich et al. 2011; Kroenke 2011; Land 1985; Valacich et al. 2011; WKWI 1994). Well-founded socio-technical approaches study information systems within their organizational context and hence ensure that the individual elements, e.g., the IT component, and their design are aligned with this context and with each other (Alter 2008; Carvalho 2000; Lyytinen and Newman 2008).

As one particular socio-technical approach, the work system approach defines an information system as “a system in which human participants and/or machines perform work (processes and activities) using information, technology, and other resources to produce informational products and/or services for internal or external customers” (Alter 2008, p. 451). Informational products are understood in a broad sense and include, for instance, the production of digital goods. The work that is performed in such a system, i.e., its processes and activities, is devoted to a generic function of processing information, which involves “capturing, transmitting, storing, retrieving, manipulating, and displaying information” (Alter 2008, p. 451). As a means to describe and analyze an information system, the work system framework identifies, among others, the four basic components that are involved in performing the work: processes and activities, (human) participants, information, and technologies. The work system view of the function and elements of information systems can easily be mapped to the definitions in most of the IS textbook literature (Davis 2000; Ferstl and Sinz 2008; Gray 2005; Heinrich et al. 2011; Huber et al. 2006; Kroenke 2011; Rainer and Cegielski 2010; Valacich et al. 2011).

Building on this definition, *crowdsourcing systems* can be perceived as socio-technical systems that provide informational products and services for internal or external customers by harnessing the diverse potential of large groups of people, usually via the Web. All of these systems employ variations of a generic crowdsourcing process that relies primarily on contributions from human participants to transform existing or produce new information. Information technology is used to enable this process and, where possible, support the activities performed in the system. Figure 2.1 summarizes the nature of the main components in a crowdsourcing system.

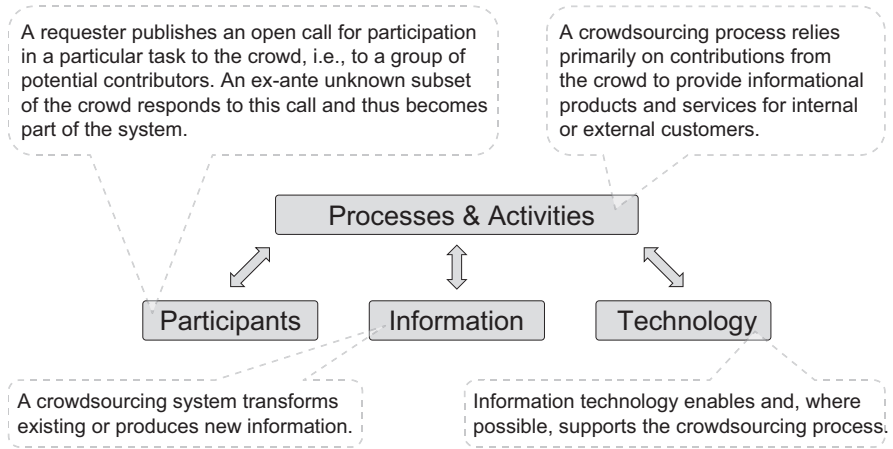


Fig. 2.1 Components of a crowdsourcing system

A crowdsourcing process starts with a crowdsourcing organization, often called requester, that publishes an open call for participation in a particular task to the *crowd*, i.e., to a typically unrestricted group of potential contributors. In response to this call, an ex-ante unknown subset of individuals in the crowd decide if they are willing and capable of contributing to the respective task. This essential self-selection part of the process is also called “self-identification of contributors” (Howe 2009). Individuals that choose to participate submit their contributions, which are then aggregated and selected in different ways (Schenk and Guittard 2011), depending on the function of the crowdsourcing system.

The open nature of their processes enables crowdsourcing systems to reach a large number of potential contributors and, therefore, to scale very well. At the same time, interested individuals may come from entirely different backgrounds, which creates the potential for a high level of diversity in the respective crowd. It is this unique scalability and diversity that, if implemented correctly, ensure the large success of crowdsourcing systems for a variety of organizational functions.

2.2 Organizational Functions

According to general systems theory, a system cannot be understood by analyzing its constituent parts alone, but only by determining its function in the supersystem, i.e., the surrounding system of which it is a part (Ackoff 1993; cited in Silver et al. 1995). Like any information system, a crowdsourcing system thus needs to be considered in terms of its function within the organizational environment before one can derive details on its “features and component parts” and how they enable this function (Silver et al. 1995).

Consequently, in order to facilitate a better understanding of the respective socio-technical designs, this section lays out a distinction of fundamental crowdsourcing system archetypes along their organizational function. The development of typologies provides structures to organize the body of knowledge and to study relationships among otherwise disorderly concepts (Glass and Vessey 1995; Nickerson et al. 2013). By enabling a more differentiated perspective, a typology of systems can be used to channel research results and to attribute divergences, e.g., with respect to the design of task recommendation approaches, to systematic differences (Sabherwal and King 1995).

2.2.1 *Typology Development*

Nickerson et al. (2013) describe a structured approach to the development of classification schemes, which is based on a comprehensive literature survey of 73 taxonomies in IS and on methodological guidance from related fields. Note that, according to Nickerson et al., the terms *taxonomy* and *typology* are often used interchangeably and can refer to both a set of characteristics and a set of classified objects. One of the main traits of the method is the definition of a meta-characteristic as a first step to the development of any taxonomy:

The meta-characteristic is the most comprehensive characteristic that will serve as the basis for the choice of characteristics in the taxonomy. Each characteristic should be a logical consequence of the meta-characteristic. The choice of the meta-characteristic should be based on the purpose of the taxonomy. (Nickerson et al. 2013, p. 8)

The purpose of this typology, as stated above, is to distinguish archetypes of crowdsourcing systems along their organizational function. Following the work system definition, the general function of any information system can be perceived as the processing of information in order to provide informational products and services (Alter 2008). As opposed to traditional information systems, the product or service that is provided by a crowdsourcing system is essentially determined by the input of its various human contributors. The most comprehensive characteristic to differentiate between the organizational functions of these systems therefore considers how a system derives value from crowd contributions to deliver the aspired result.

As prescribed by the method, an iterative combination of empirical-to-conceptual and conceptual-to-empirical approaches was employed to develop the typology. On the empirical side, this process involved analyzing a sample of over 50 crowdsourcing systems. On the conceptual side, inspiration was drawn from concepts used in systems theory and from existing attempts at categorizing the crowdsourcing landscape in academia and industry. The iterative approach involved a broad consideration of candidate dimensions, which were tested for their individual relevance to the meta-characteristic and their collective distinctive potential. Nickerson et al. provide a number of objective and subjective criteria that

serve as both ending conditions for the development process and evaluation criteria for the resulting artifact. The next section presents this artifact, followed by a brief discussion of its compliance with the evaluation criteria.

2.2.2 System Archetypes

The developed typology covers the specified meta-characteristic—how a crowdsourcing system derives value from crowd contributions to achieve its organizational function—by differentiating two fundamental dimensions: (i) whether a system seeks homogeneous or heterogeneous contributions from the crowd and (ii) whether it seeks an emergent or a non-emergent value from these contributions.

- i. A system that seeks *homogeneous* contributions values all valid contributions equally. Homogeneous contributions that comply with the predefined specifications are seen as qualitatively identical; the system is geared to mere quantitative processing. In contrast, a system that seeks *heterogeneous* contributions values these contributions differently according to their individual qualities. Relevant qualities are determined by the specific task and vary from objective measures, such as test results, to subjective perceptions, such as clarity or esthetics. Heterogeneous contributions are seen as alternatives or complements and are processed accordingly. This dimension is inspired by the notion of heterogeneous components (or components perceived as such), which is studied in various systems (Heinrich et al. 2011, p. 16). A particular focus on heterogeneity can be found, for instance, in the context of distributed computer systems (Maheswaran et al. 1999) or agent models in economic systems (Hommes 2006).
- ii. A system that seeks a *non-emergent* value from its contributions derives this value directly from all or some of the individual contributions in isolation. In such systems, an individual contribution delivers a fixed value, which is independent of other contributions. A system that seeks an *emergent* value from its contributions, however, can only derive this value from the entirety of contributions and the relationships between them. An individual contribution therefore only delivers value in combination with others. Emergence is a philosophical concept that is, among others, central to systems theory to denote properties of a system that are not possessed by its isolated components but rather depend on the relationships among them in a composition (Bunge 2003, p. 12 ff.; Checkland 1988, p. 243; Heinrich et al. 2011, p. 15; Weber 1997, p. 37).

The combination of these two dimensions yields four fundamental types of crowdsourcing systems as depicted in Fig. 2.2. Each of the archetypes was given a label that reflects its distinct organizational function and, accordingly, the type of product or service delivered by the system.

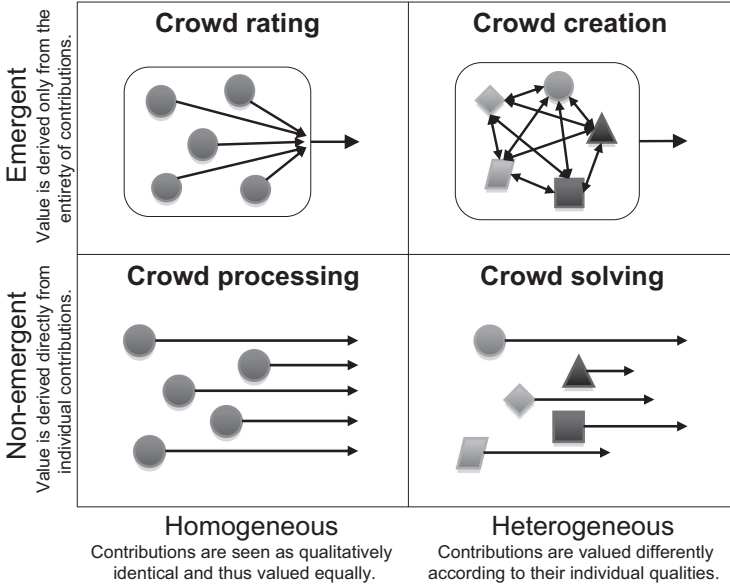


Fig. 2.2 The four archetypes of crowdsourcing systems

Crowd processing systems seek non-emergent value that derives directly from large quantities of homogeneous contributions. Valid contributions represent qualitatively identical and therefore equally valued chunks of work. These systems utilize the additional bandwidth and the scalability provided by a crowdsourcing approach for quick and efficient batch processing, thus minimizing the use of traditional organizational resources (Doan et al. 2011). The diversity of interests within the crowd enables these systems to recruit sufficient contributors for a variety of tasks, most of which harness basic abilities of the human brain. Prominent examples for crowd processing systems include Mechanical Turk,¹ Galaxy Zoo² (Lintott et al. 2008), and Recaptcha³ (von Ahn et al. 2008).

Crowd rating systems also rely on large amounts of homogeneous contributions but seek a collective value that emerges only from the entirety of contributions. As homogeneous contributions are considered qualitatively identical, the aspired value emerges from the mere quantitative properties of the collective input. Contributions hence represent ‘votes’ on a given topic. By aggregating a sufficient number of these votes, crowd rating systems deduce a collective response, such as a collective assessment on TripAdvisor⁴ or a collective prediction on the Hollywood Stock

¹ <http://www.mturk.com/>

² <http://www.galaxyzoo.org/>

³ <http://www.google.com/recaptcha/>

⁴ <http://www.tripadvisor.com/>

Exchange.⁵ Typically, the larger and more diverse a crowd that rating systems can assemble, the more accurate the results become. In fact, Surowiecki (2005) states diversity as one of the prerequisites for the “wisdom of crowds” to emerge.

Crowd solving systems seek non-emergent value that derives directly from the isolated values of their heterogeneous contributions. Contributions are qualitatively different and thus represent alternative or complementary solutions to a given problem. These systems can be built around hard problems with well-defined evaluation criteria, such as FoldIt⁶ (Cooper et al. 2010), kaggle⁷ (Carpenter 2011), or the Netflix Prize (Bennett and Lanning 2007). They can also be used to approach soft problems that do not have an optimal solution, such as ideation contests on InnoCentive⁸ (Allio 2004) or ideaBounty⁹ and ‘make-to-order’ digital product contests on 99designs¹⁰ or Naming Force.¹¹ Crowd solving systems benefit from large and diverse crowds as every individual potentially provides new insights, ideas, or abilities and therefore increases the chance of finding a (better) solution. In some contexts, this phenomenon is referred to as the “wisdom in the crowd” (Dondio and Longo 2011, p. 113).

Crowd creation systems seek a collective value that emerges from the accumulation of a variety of heterogeneous contributions. In contrast to crowd rating systems, the aspired value emerges not only from quantitative but primarily from qualitative properties of the collective input. Contributions have a complementary share in the collective outcome depending on their individual qualities and their relationship with others. Large numbers of diverse contributors enable crowd creation systems to harness multiple perspectives, distributed knowledge, or different skills, and to aggregate the corresponding contributions into highly comprehensive artifacts. Typical examples are user-generated content platforms (e.g., YouTube¹²), the ‘make-to-stock’ production of digital content (e.g., stock photography), or knowledge aggregation (e.g., Wikipedia¹³).

As stated above, the described systems are archetypes with distinct organizational functions. In practice, many crowdsourcing efforts are built on hybrid systems that combine some of these functions, often relying on quantitative and qualitative components. Many systems that are based on heterogeneous contributions, for instance, rely on a crowd rating function in form of a collective vote as an

⁵ <http://www.hsx.com/>

⁶ <http://fold.it/>

⁷ <http://www.kaggle.com/>

⁸ <http://www.innocentive.com/>

⁹ <http://www.ideabounty.com/>

¹⁰ <http://99designs.com/>

¹¹ <http://www.namingforce.com/>

¹² <http://www.youtube.com/>

¹³ <http://www.wikipedia.org/>

indicator for the quality of individual contributions. Some examples include Dell's IdeaStorm,¹⁴ iStockphoto,¹⁵ or Google's App Store.¹⁶ Additionally, some of these systems make use of crowd processing functions such as tagging in order to organize the set of input elements.

2.3 Discussion

According to the design science paradigm, a “search for the best, or optimal, design is often intractable for realistic information systems problems” and should instead aim to discover effective solutions (Hevner et al. 2004, p. 88). On this basis, Nickerson et al. (2013) argue that taxonomies can only be evaluated with respect to their usefulness. They propose a set of qualitative attributes that form the necessary conditions for a useful taxonomy: it needs to be concise, robust, comprehensive, extendible, and explanatory.

For the scope of this work, the developed typology satisfies these criteria. The number of characteristics is *concise* enough to be easily applied, yet they provide a *robust* differentiation of distinct system archetypes. The typology is *comprehensive* because its disjoint characteristics partition the set of crowdsourcing systems, which allows the classification of every system in the used sample. While future work could *extend* the typology by identifying further subtypes, the current version has sufficient *explanatory* power with respect to the essential mechanisms and the organizational functions of the classified system instances.

In addition to these generic, necessary conditions, however, Nickerson et al. note that the sufficient conditions for usefulness depend on the expected use of a specific taxonomy. As classification schemes are not an end in themselves, their usefulness can only be evaluated by observing their use over time, with regards to their respective purpose. Following that idea, the next chapter proceeds to apply the typology to the comparative study of personalized task recommendation approaches in crowdsourcing systems.

¹⁴ <http://www.ideastorm.com/>

¹⁵ <http://www.istockphoto.com/>

¹⁶ <http://play.google.com/store>

Personalized Task Recommendation in Crowdsourcing
Systems

Geiger, D.

2016, XI, 108 p. 23 illus., 9 illus. in color., Hardcover

ISBN: 978-3-319-22290-5