POLITECNICO DI MILANO

Facoltà di Ingegneria dell'Informazione

Corso di Laurea Specialistica in Ingegneria Informatica

A crowdsourcing methodology for a semantic sentiment analysis engine

Relatore:

Prof.ssa Chiara Francalanci

Correlatore:

Prof. Eugenio Capra

Tesi di Laurea Specialistica di

Francesco Pongetti

Matricola 739280

Anno Accademico 2010/2011

To all the people I met in my life that showed me the correct handling of knives, books and wine such as my tools, wit and taste may be kept sharp

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

SOMMARIO

La *Sentiment Analysis* consiste nell'applicazione di tecniche automatiche di elaborazione del linguaggio naturale (NLP) per estrarre le opinioni espresse in un testo su un oggetto o un argomento da parte dell'autore. La *Sentiment Analysis* è spesso utilizzata per analizzare i contenuti provenienti dai *social media* sul Web ed estrarre le opinioni in materia di *brands* e prodotti. Questo processo prende il nome di analisi della *Web Reputation*. Strumenti semantici di analisi del sentimento richiedono un'approfondita conoscenza di dominio per eseguire le loro analisi. Nella maggior parte dei casi il database della conoscenza va costruito e mantenuto aggiornato manualmente. In questo lavoro si propone un nuovo approccio automatico per raccogliere dati al fine di costruire una conoscenza di dominio. Il nostro approccio sfrutta il *crowdsourcing*. Il *crowdsourcing* è un modello che utilizza tecnologie IT, come software e reti di telecomunicazione, per costruire una piattaforma virtuale, in cui dei *tasks* sono svolti in *outsourcing* da un insieme distribuito di individui che sono disposti a eseguirli. Il nostro approccio si basa sull'idea di acquisire la conoscenza di dominio sfruttando una *web application* e una comunità online di individui. In questo lavoro si è proceduto in primo luogo studiando il fenomeno del *crowdsourcing* e costruendo una serie di strumenti di modellazione per la sua applicazione. Abbiamo utilizzato questi strumenti per sviluppare una metodologia di *crowdsourcing* per un *tool* di analisi semantica del sentimento. Abbiamo identificato, quindi, una serie di diversi possibili design che una metodologia definitiva avrebbe potuto assumere. Il passo successivo è stato costruire un esperimento consistente nel test di quattro diverse metodologie, rappresentate da quattro differenti *crowdsourcing web applications*. Ciascuna

metodologia si differenzia dalle altre proponendo diversi tipi di compiti agli utenti nella propria comunità e sfruttando paradigmi distinti sia per incoraggiare la partecipazione degli stessi sia per raccogliere dati. In particolare, ogni metodologia premia in modo diverso gli utenti in cambio delle loro informazioni (compensazione monetaria, esperienza di gioco, ecc.). Infine, abbiamo chiesto a un gruppo di prova di 51 utenti di utilizzare le nostre applicazioni e fornire le loro opinioni attraverso un questionario. In questo modo, abbiamo valutato le quattro metodologie rispetto alla variabile rappresentata dalla soddisfazione degli utenti. I dati provenienti dall'esperimento sono stati analizzati allo scopo di elaborare una proposta definitiva di metodologia di *crowdsourcing*, capace di raccogliere dati da un esteso insieme di individui e popolare, in questo modo, il dominio di conoscenza di uno strumento di analisi semantica del sentimento. Questo modello finale massimizza la soddisfazione degli utenti, cioè la metrica di giudizio da noi scelta, e di conseguenza risulta validato secondo i nostri parametri e obiettivi.

ABSTRACT

Sentiment Analysis refers to the application of natural language processing techniques to extract the sentiment expressed by an author in a text, over an object or a topic. Nowadays, the sentiment analysis is widely adopted to analyze the content coming from online Web social media, and extract the people's opinions regarding brands and products. This process is named Web Reputation analysis. Semantic sentiment analysis tools require extensive domain knowledge to perform their analyses. The majority of the actual software designs require to manually provide and keep up to date, the knowledge database. In this work we propose a novel automatic approach to gather and collect data in order to build a domain knowledge. Our approach exploits the crowdsourcing. Crowdsourcing is a model in which a collection of IT technologies, such as software and telecommunication networks, are exploited to build a virtual platform for outsourcing a certain collection of tasks to a distributed pool of individuals that are willing to perform these tasks. Thus, our approach is based on the idea of acquiring specific domain knowledge by addressing a large online community of individuals through a web application. In this work we proceeded firstly by studying the crowdsourcing phenomenon and by building a set of modeling tools for crowdsourcing applications. Then, we used these tools to develop a crowdsourcing methodology for the sentiment analysis. We identified a set of several options from which we could choose the features of the final crowdsourcing methodology. Thus, we built an experiment consisting on the testing of four different methodologies, represented by four different crowdsourcing web applications. Each methodology differs from the others by exploiting different types of task executed by the users in its community, and distinct paradigms for

encouraging user participation and the data collection. In particular, each methodology rewards the users for giving their knowledge, in a different way (monetary compensation, gaming experience, etc.). We addressed a test group of 51 users and we asked them to use our applications and provide their opinions by means of a survey. In this way, we assessed the four methodologies with respect to the user satisfaction variable. Finally, we analyzed the data coming from the experiment to come up with a final proposal of a crowdsourcing methodology to collect data from a large set of individuals in order to populate the knowledge domain of a sentiment analysis tool. This final model maximizes the user satisfaction metric that we chose as quality indicator and thus it results validated according to our parameters and goals.

ACKNOWLEDGMENTS

First, my utmost gratitude to my advisors, Prof. Chiara Francalanci and Prof. Eugenio Capra, for supporting me during this work and for giving me so much freedom to explore and discover new areas of research.

I would like to thank my family: my father Marzio, my mother Mirella, my brother Andrea, my uncle Walter and my aunt Giuliana. My mother has always been very supportive during these years and her help has given me power and strength. My brother has been a continuous and unlimited source of inspiration since I was a child. My father has always been my model of the correct behavior in life. My uncle's teaching has contributed to my growth more than all my studies together. My aunt has always been so devoted to me and my brother.

I would like to thank Raffaella, her continuous help and encouragement has been fundamental to arrive here. I cannot find the words to express my gratitude to such an unlimited support.

I would like to thank my many friends and colleagues at Politecnico di Milano with whom I have had the pleasure of studying over the years.

Finally, I would like to thank my friends in Senigallia and in Milan for always providing a welcome distraction from school.

Francesco

Chapter 1

Introduction: scope of research and methodology

The following thesis is the result of our researches in the fields of *crowdsourcing* and *sentiment analysis.*

Sentiment analysis, or opinion mining, refers to the application of natural language processing, computational linguistics, and text analytics to identify and extract subjective information in source materials. In particular, it usually refers to the analysis of human-redacted text to detect the attitude (positive, negative or neutral) expressed by the author over an object. Several software applications have been developed for the *sentiment analysis*. At this regard, our research focused on a specific subset of these tools, capable of analyzing the content coming from various social platforms (e.g. Twitter) and detecting the sentiment expressed by the authors over a particular brand or product. This process is named *Web reputation* analysis. The most valuable *sentiment analysis* tools exploit semantic text processing techniques and thereby need comprehensive domain knowledge to perform their analyses. One of the shortcomings of the actual design of the majority of these instruments is that the knowledge domain must be manually provided by the administrators or by the users of the tool. Thus, this approach requires devolving resources in order to customize the knowledge domain before using the application. Moreover, it requires devolving time to keep the knowledge domain up to date and to extent it with new information that may come out over the time.

In this thesis we propose a new automatic approach to the building and maintenance of the domain knowledge of a semantic *Web reputation* analysis tool, based on the application of the *crowdsourcing.*

Crowdsourcing is a novel research topic and it has been poorly covered by the literature, even if much has been said about it by several sources. We define *crowdsourcing* as *"a model in which a collection of IT technologies, such as software and telecommunication networks, are exploited to build a virtual platform for outsourcing a certain collection of tasks to a distributed pool of individuals (the crowd) that are willing to perform these tasks"*. Being such an innovative domain, in order to harness the *crowdsourcing* in our tool, we first had to perform a widespread and comprehensive study of the phenomenon. We proceeded by developing a complete model for analyzing and studying the *crowdsourcing.* The resulting model is composed of two parts: the *descriptive framework* and the *prescriptive framework.*

The *descriptive framework* is our modeling instrument. It provides ten dimensions of analysis and the related metrics, to classify, describe and model *crowdsourcing* systems. Examples of these dimensions are: the User Type and the Task Type, that describe the type of users (amateur or professional) and tasks (simple, complex, game) of a *crowdsourcing* platform; the Community Size, that describes both in a quantitative and qualitative way the size of the user-base of a *crowdsourcing* system; or, finally, the Reward and Incentive dimensions, that assess the type of reward and incentive mechanisms that the *crowdsourcing* system implements to encourage user participation. The ten dimensions are the result of our theoretical and experimental researches in the *crowdsourcing* domain. We put particular focus on the developing of a theoretically well-grounded descripting model. The *descriptive framework* fulfills this requirement by aggregating and presenting both our theoretical research, and our experimental experience acquired by analyzing real *crowdsourcing* platforms currently on the market.

The *prescriptive framework* moves our work a step further. It applies the concepts introduced by the *descriptive framework* for analyzing a vast and heterogeneous empirical dataset composed of 61 real *crowdsourcing* applications. Examples of *crowdsourcing* applications that we analyzed in the empirical dataset are: Wikipedia, Amazon Mechanical Turk and Threadless. The *prescriptive framework* analyzes the links between the dimensions in the *descriptive framework* looking how the values of each dimension are mapped in the values of the other dimensions, using the dataset as source of information. In this way it provides a characterization of the most common practices implemented by real and successful *crowdsourcing* systems on the market.

The development of these tools gave us a sufficient knowledge of the *crowdsourcing* domain, in order to properly exploit it. We then returned at our main problem of building an innovative methodology to populate the domain knowledge of a semantic sentiment analysis tool. At this regard, we developed four test cases of *crowdsourcing* methodologies aimed at this goal. The test cases are examples of real *crowdsourcing* web applications and we named them *crowdsourcing* scenarios. In particular, all the four scenarios are *Prediction markets* web applications. *Prediction Markets* are virtual markets created for the purpose of making predictions. The user of these markets makes prediction concerning events in the future by answering to questions or placing new questions. Then, the answers coming from a large user-base of web users are aggregated by the web application and the result is the final prediction. We modeled and chose the scenarios according to the *descriptive framework* and by using the knowledge acquired from the *prescriptive framework.* Each scenario exploits a different set of values for several dimensions in the *descriptive framework* (Rewards, Incentive, Task Type, etc.). After building them, we experimentally tested the applications by addressing a group of test users.

We can summarize the methodology that we followed in this thesis and its organization in a set of successive steps. Each step is covered by one of the following chapters.

The first step has been studying the literature on *sentiment analysis* and *crowdsourcing* to understand the boundaries of our research, the main concepts, the current development in these fields, and where we can insert our work in the scientific playground. Chapter 2 covers this part.

The result of this critical literature review moved us in the second step. We understood that we needed a set of tools to support us in the domain of *crowdsourcing.* Thus, we developed the *descriptive* and *prescriptive frameworks* presented in Chapter 3.

Chapter 4 describes the experimental phase of our work. What we wanted was building an effective *crowdsourcing* methodology for a sentiment analysis engine. Many parameters can be used to assess the quality of a *crowdsourcing* methodology: the resources needed for implementing it, the quality of the data that it produces, etc. The researches that we carried out in the context of the *frameworks* suggested us to turn our attention toward the user satisfaction, i.e. the level of satisfaction that the users of a *crowdsourcing* application express over it. Indeed, we understood that maximizing the level of satisfaction could maximize also the level of participation. Thus, we made the hypothesis that the user satisfaction could be a valid indicator for harnessing the quality of a *crowdsourcing* system. In the wake of this, we set up our experiment. Following the recommendations coming from the *frameworks,* we developed the four different *crowdsourcing* scenarios. We then assessed the user satisfaction parameter by asking to a test group of users to test the scenarios and answers to a survey. Finally we analyzed the experimental data in order to build a *crowdsourcing* methodology for a sentiment analysis engine that maximizes the user satisfaction variable.

Chapter 5 contains the final conclusion of this thesis and the answer to the question if our research methodology has been valid for solving our research problem. Moreover, it outlines the possible future developments of the research.

Finally, the Appendix A contains a detailed description of the empirical dataset of *crowdsourcing* applications and the Appendix B contains all the disaggregated results of the survey, together with the list of questions. These appendixes can be used as reference while reading this thesis.

Chapter 2

Background review and state of art

2.1 Introduction

In this chapter we will present the fundamental concepts used throughout all this research. In particular, we will widely discuss the concepts of *wisdom of the crowds, crowdsourcing* and *sentiment analysis.* The aim driving our efforts is to provide a complete set of definitions and ideas that are needed to understand the analyses and discussions carried out in the remaining part of thesis. This is the first step of our research methodology as we outlined it in the fist chapter.

First, we will briefly outline the topic of the *wisdom of the crowds* presenting a series of related work and research. The *wisdom of the crowds* is one of the concepts that stay behind the *crowdsourcing* phenomenon. Thus, we must spend some words at this regard before addressing the *crowdsourcing* in order to give the reader a complete background context.

Crowdsourcing is the second topic discussed in this chapter. It is the central topic and driver of all our research and thereby, we will take care of presenting and describing it in details. In particular, we will offer an introduction to the topic containing the main definitions used throughout the entire thesis. Then, we will carry out a review of the other works that have been developed in this field. The goal of this review is to show where this research can be put in the scientific playground and how it differs from other related works. Finally, we will provide some examples of *crowdsourcing* applications that will help the reader in understanding and correlating all the presented concepts.

The last part of this chapter is dedicated to presenting the *sentiment analysis.* The *sentiment analysis* is, together with the *crowdsourcing,* the key concept from which we started developing our research. We will explain what is the *sentiment analysis* and how is correlated to the *crowdsourcing.* We will conclude presenting several works that tried to address some of the same problems we faced.

2.2 Wisdom of the crowds

2.2.1 What is the wisdom of the crowds

In his book, Surowiecki discusses in detail the concept of *wisdom of the crowds* and propose to define it as the process of taking into account the collective opinion and knowledge of a group of individuals rather then a single expert, to answer a question or find a solution to a problem (Surowiecki, 2004). The term *crowds* will be widely used throughout this thesis with various meanings that will be clarified each time. For the moment, we can provide a definition, provided by Surowiecky (2004) as well. This definition is rather abstract and general but already suitable for our purposes. When presenting the *descriptive framework* we will propose instead an operational definition more suitable for analytical purposes. The following is the definition by Surowiecky (2004):

> *"A crowd is any group of people who can act collectively to make decisions and solve problems. So, on the one hand, big organizations, like a company or a government agency, count as crowds. And so do small groups, like a team of scientists working on a problem. But even more interesting are groups that aren't really aware themselves as groups, like bettors on a horse race or investors in the*

stock market. They make up crowds, too, because they're collectively producing a solution to a complicated problem"

The idea of studying the behavior of big groups of people taking decisions is rather old. At the beginning of the last century, the British scientist Francis Galton already tried to analyze how collective judgments arise among group of individuals at a county fair, and to assess their quality compared to other form of decisions (Surowiecki, 2004). He surprisingly found out that the quality of collective judgments was usually higher with respect to decisions made by single experts. These findings were later confirmed by several other experimental studies (Steyvers *et al.*, 2009; Vul & Pashler, 2008; Kittur & Kraut, 2008).

Surowiecki relates to diverse collections of independently deciding individuals, rather than crowd psychology as traditionally understood. His central thesis, that a diverse collection of independently deciding individuals is likely to make certain types of decisions and predictions better than individuals or even experts, draws many parallels with statistical sampling. Surowiecki argues that, under certain conditions, the collective judgments of large groups of people are more accurate than the judgments of any individual, even an expert (Surowiecki, 2004). He names this phenomenon the *wisdom of the crowds* effect.

In the 90', due the Internet revolution and the increasing number of people connected by digital technologies, the idea has started acquiring more value and being fashionable in the academic and business context. Internet, the Web and cheap IT connections, have made finally possible to design systems capable of interconnecting and interacting with huge groups of users. The IT technologies and software applications have given the way to harness the *wisdom of the crowds* to a large extent and scale. Moreover, these technologies come at a low cost, adding more value and appeal to the projects trying to exploit the collective knowledge.

In the computer science domain many projects have tried to tap into the collective knowledge of the people. The *wisdom of the crowds* effect has been used by many famous projects such as Wikipedia, Amazon Mechanical Turk, etc. (these projects will be covered in the following paragraphs as examples). Moreover, the *wisdom of* *the crowds* is the idea running many social web applications such as the Folksonomies. A Folksonomy (the term has been coined by Thomas Vander Wal and is a portmanteau of folks and taxonomy) is a system of classification derived from the practice and method of collaboratively creating and managing tags to annotate and categorize content (Peters, 2009); this practice is also known as collaborative tagging, social classification, social indexing, and social tagging (Lambiotte & Ausloos, 2006). Tagging, which is one of the defining characteristics of *Web 2.0*[1] services, allows users to collectively classify and find information, thus it widely taps the *wisdom of the crowds* as we presented it until now.

The idea holds so much appeal for the social software community that "*the wisdom of crowds*" has become a part of the vernacular. Unfortunately, most of these claims are misguided. The problem is that people tend to forget the "under certain conditions" part of Surowiecki's theory. Indeed, Surowiecki in his researches, put some constraints and hypotheses aimed at allowing the *wisdom the crowds* effect to effectively work. Not all *crowds* (groups) are wise. Consider, for example, mobs or crazed investors in a stock market. Indeed, according to Surowiecki (2004), the requirements of “openness”, “peering” and “heterogeneity” are key factors for the emerging of the *collective intelligence.* These general requirements can be declined in four elements, required to form a wise *crowd:*

- Diversity of opinion
- Independence
- Decentralization
- Aggregation

[1] The term Web 2.0 is associated with web applications that facilitate participatory information sharing, interoperability, user-centered design and collaboration on the World Web. A Web 2.0 site allows users to interact and collaborate with each other in a social media dialogue as creators of user-generated content in a virtual community, in contrast to websites where users are limited to the passive viewing of content that was created for them. Examples of Web 2.0 include social networking sites, blogs, wikis and folksonomies.

Diversity of opinion is fundamental. Theoretically, each person should have private information even if it's just an eccentric interpretation of the known facts. Practically, this is impossible to achieve and the requirement is usually relaxed: in order to effectively being able to exploit the *wisdom of the crowds* effect, the *crowd* must present, at least, a good grade of heterogeneity.

Independence is another important requirement. People's opinions don't have to be determined by the opinions of those around them. As we already discussed for the diversity of opinion, this property is also impossible to be completely ensured in practice, but large online virtual communities, composed of millions of users spread in the world, can achieve a good grade of independence.

Decentralization means the people are able to specialize and draw on local knowledge, i.e. the people don't have a common social, cultural and environmental background. This property is easily ensured by the fact the virtual communities are virtually accessible from everywhere in the world.

Finally, aggregation is the last requirement. It means that the systems should implement some mechanisms for turning private judgments into a collective decision.

Moreover, still according to Surowiecki, the *wisdom of the crowds* effect can occur only when members of a *crowd* hold another fundamental property:

- They are presented with a clearly defined problem

This constraint means that the *wisdom of the crowds* is effective only when tapped for solving well-defined problems, i.e. problems where a resolution methodology is already shaped and the information coming from the *crowd* are the input of this resolution process. We can link this requirement to the previous of aggregation. Indeed, only for clearly defined problems is possible to design an aggregation mechanisms. The systems (online or offline) cannot exploit the *wisdom of the crowds* effect to find novel resolutions mechanisms. The aggregation algorithms must already employ one of these mechanisms. According to this, the *wisdom of the crowds* can be seen just as a way of collecting a great amount of disparate information to be used as input of a resolution process. This is a fundamental concept, often misunderstood and discarded by the researchers and innovators in this field.

Until now we presented the concepts of *wisdom of the crowds* and *crowds* and we showed which are the requirements that must be enforced to effectively tap them. However, we didn't address the problem of establishing which are the theoretical groundings that run the *wisdoms of the crowds* effect. In other word, we didn't explain, from a theoretical point of view, how is possible for the *wisdom of the crowds* to work. Bayn (2008) reformulates this problem in the question "Where does the *wisdom of the crowds* come from?". Surowiecki (2004) doesn't offer any theoretical explanations instead he merely provides a comprehensive list of *wisdom of the crowds* working examples without trying to address the problem. Other authors instead try to provide some explanations. For instance, Bayn (2008), at this regard, recalls the Law of Large Numbers: the trick is that truly diverse, i.e. random, opinions will always vary around the mean of the solution; so when you aggregate a whole lot of random opinions, you get a deceptively precise average. Steyvers *et al.* (2009) argue that that what powers the *wisdom of the crowds* effect is actually the aggregation mechanism and thus, they state that the real "intelligence" arises from this algorithm and not from the *crowd.* On the other hand, Kittur & Kraut (2008) think that the coordination of the people's individual intelligences makes it possible for a collective intelligence to arise and perform better than the single individuals.

Moreover, some authors showed the weakness of the *wisdom of the crowds* approach and how it can easily be undermined in various ways. For instance, Lorenz *et al.* (2011) showed that social influence among a community could easily destroy the validity of the outcomes of the *wisdom of the crowds* effect. On the same wake, Kittur *et al.* (2007) made a study and showed that as the online communities get older and bigger, small subgroups of "elite" users emerge within them and this aspect can greatly undermine the quality of the results. Finally, we can conclude with Surowiecki (2004) who cites many more potential sources of bias: group polarization, conformity, in-group bias, framing and groupthink. All these sources come from the study of Social Psychology which, and should be clear from now on, is related to the *wisdom of the crowds* and the *crowdsourcing* phenomena as much as the Computer Science.

2.3 Crowdsourcing

2.3.1 The Crowdsourcing: introduction and description

Coined by Jeff Howe in the June 2006 issue of Wired magazine (Howe, 2006), the term *crowdsourcing*, in its original fashion, describes a new web-based business model that harnesses the creative solutions of a distributed network of individuals through what amounts to an open call for proposals. Howe (2006) offers the following definition:

"Simply defined, 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. This can take the form of peer-production (when the job is performed collaboratively), but is also often undertaken by sole individuals. The crucial prerequisite is the use of the open call format and the large network of potential laborers"

According to this definition, *crowdsourcing* is mostly related to the business world and the term itself is a portmanteau of crowd and sourcing, making clear the link with the word outsourcing that belongs to the economics and business jargons. Therefore, it shouldn't be surprising that most of the sources that covered this phenomenon from the beginning have seen it as something concerning the business domain (see Chanal & Caron-Fasan, 2008; Whitla, 2009). The main difference between the *crowdsourcing* and the traditional outsourcing is that in the later case we have a subject (a company, an institution, etc.) that hires a single external agent to perform a task while in the former case the same subject makes an open call to a vast and distributed network of agents that concur, at the same time, in accomplishing the task, until the subject doesn't find the best solution(s) for himself and close the call. Therefore, in this business exception, *crowdsourcing* describes a process of organizing labour, where firms parcel out work to some form of online community, offering rewards of some kinds for anyone within the *crowd* who completes the task the firm has set (Whitla, 2009).

The description we outlined until here derives from the traditional business point of view. However, from its origin the phenomenon has deeply changed its borders and characteristics and now the term usually refers to a broader idea that embraces the field of computer science, business and behaviorism. Thus, nowadays, *crowdsourcing* is a term representing a whole set of IT systems, technologies and designs, built for harnessing distributed networks of individuals with the aim of accomplishing some kind of tasks. *Crowdsourcing* isn't anymore just a business model but it represents a whole new way of designing IT systems interacting with large online communities. Moreover, in this exception, the requirement for a precise subject outsourcing the tasks to the *crowd* disappears because the *crowd* itself may self-organize in an online community that performs some kind of jobs.

Finally, the original Howe's definition refers to web-based systems (Howe, 2006), while, nowadays, the market proposes several desktop applications capable of exploiting the *crowdsourcing*. At this regard, we will discuss the example of computer games developed for *crowdsourcing* information from the players (see Chapter 3). At a later stage, Howe himself defines the *crowdsourcing* in more general term (Howe, 2008):

"Crowdsourcing is the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call"

In addition to this, we too have to provide an operational definition of the *crowdsourcing.* It is a fundamental passage in order to clearly understand, from the beginning, the limits of the research field in which we are moving. Thus, we define the *crowdsourcing* as

"a model in which a collection of IT technologies, such as software and telecommunication networks, are exploited to build a virtual platform for outsourcing a certain collection of tasks to a distributed pool of individuals (the crowd) that are willing to perform these tasks"

In our definition the individuals can either concur to make the best solution emerge among the many others, or they can collaborate on the same task to develop and evolutionary improving solution (we will discuss in detail these two approaches in the *descriptive framework,* Chapter 3).

We can draw a parallelism between the concept of *wisdom of the crowd* presented in the previous paragraph and the *crowdsourcing.* Indeed, beyond the fact that the word *crowd* appears in both the terms, the commonalities are many. Before proceeding on this topic, we need to underline the fact that in this passage we are referring to the *wisdom of the crowds* as a model, namely a problem-solving model, and, as consequence, we are not focusing our attention on the *wisdom of the crowds* effect that we widely discuss previously. Actually, we can see the *wisdom of the crowd* model as the model that harnesses the *wisdom of the crowds* effect and in this way we will consider it in relation to the *crowdsourcing*.

The *crowdsourcing* is a more general model that includes within its borders also the *wisdom of the crowds*. The *wisdom of the crowds* model requires to rely on the collective intelligence of groups of individuals to collect information in order to solve particular problems. Surowiecki states that the *wisdom of the crowds* model can be used only for problems that have a defined right answer ("cognition" problems), such as, for instance, forecasting the weather of the next week. The model instead isn't good for solving problems of skills; for instance, it means that we can't exploit the *wisdom of the crowds* model to perform a surgery (Surowiecki, 2004). Therefore, the *wisdom of the crowds* model is a declination of the *crowdsourcing* model in which the task consists in gathering, aggregating and presenting information from the *crowd.* The *crowdsourcing* model is more general because it isn't bounded to any particular kind of tasks. Concluding the parallelism, we can summarize the relationship between the *crowdsourcing* and the *wisdom of the crowds* arguing that the *crowdsourcing* is a general model of "task-resolution" while the *wisdom of the crowds* is a problem-solving model that fits just for what we named "cognition" problems. Thus, *crowdsourcing* is a broader model and the *wisdom of the crowds* model is a particular instance of the general case.

We have now to discuss the nature of the tasks that are usually crowd-sourced. This may help to understand the phenomenon itself. In our research we saw many different kind of *crowdsourcing* platforms and the range of crowd-sourced tasks varies to a great extent accordingly. In the *descriptive framework* we will propose a precise and complete categorization of these *crowdsourcing* systems. In this section we just want to provide some practical examples. The following table provides a list of *crowdsourcing* systems with a description of their scope, tasks and why we can consider them *crowdsourcing* platform according to our definition.

Table 1: Several *crowdsourcing* systems and their description

Name	Description
Amazon Mechanical Turk	Amazon Mechanical Turk2 (AMT) is a popular crowdsourcing marketplace, introduced by Amazon Inc. in 2005. AMT is a marketplace for small tasks that cannot be easily automated. Tasks are usually jobs that computers cannot easily accomplish and thus require human intelligence; for instance, telling if two different descriptions correspond to the same product, tagging an image with descriptions of its content, or transcribing with high quality an audio snippet. In the marketplace, employers are known as requesters and they post tasks, called human intelligence tasks, or HITs. The HITs are then picked up by online users,

[2] The marketplace is named after an 18th century "automatic" chess-playing machine, which was handily beating humans in chess games. Of course, the robot was not using any artificial intelligence algorithms back then. The secret of the Mechanical Turk machine was a human operator, hidden inside the machine, who was the real intelligence source.

	referred to as workers, who complete them in exchange for a small payment, typically a few cents per HIT. AMT is basically an online crowdsourcing virtual labor platform that acts as intermediary between the demand of work time and its supply from the users in the crowd. (Ipeirotis, 2010b)
Wikipedia	Wikipedia is a free, web-based, collaborative encyclopedia project. Its over 19 million articles have been written collaboratively by volunteers around the world. Wikipedia is an example of crowdsourcing project in which the task to be accomplished by the crowd is the collective writing and editing of articles. Moreover, Wikipedia is an example of crowdsourcing system unlinked to the business world: the users in the crowds are volunteers and the reward mechanism is completely not monetary. (Kolbitsch & Maurer, 2006).
InnoCentive	InnoCentive is an "open innovation" company that takes research problems in a broad range of domains such as engineering, computer science, math and business and frames them as "challenge problems" for anyone to solve them. It gives cash awards for the best solutions to solvers who meet the challenge criteria. InnoCentive calls the people who attempt the problems "solvers" and the companies these problems come from as "seekers". By 2011, InnoCentive's online community exceeds 250,000 solvers. The cash awards for solving challenge problems are typically in the $10,000 to $100,000 range. InnoCentive is a crowdsourcing platform according to our definition. It is somehow similar to Amazon Mechanical Turk but the tasks that it proposes are of a different kind being complex scientific problems. Thus, the reward mechanism is different and it will be studied in the descriptive framework as well (see Chapter 3). (Brabham, 2008a; Lakhani et al., 2006).
Threadless	Threadless is a crowdsourcing online apparel store. Members of the Threadless community submit t-shirt designs online; designers upload their t-shirt designs to the website, where visitors and members of the community score them on a scale of 0 to 5. According to our definition, Threadless is a crowdsourcing project where the tasks performed by the crowd is the designing and selection through voting of new t-shirts. (Brabham, 2008a, 2009)
Waze	Waze is a social mobile application providing free turn-by-turn GPS navigation based on the live conditions of the road. The users in its online community have to

	install the Waze's software in their mobile phone and in this way give real-time information used to build the maps and the traffic condition. According to our definition, Waze is a crowdsourcing project where the users are asked to give information while moving with their mobile phone. A collection of algorithms is then used to aggregate this data and build a knowledge base.
Kickstarter	Kickstarter is an online threshold pledge system for funding creative projects such as movies or music bands. An amount of money is set as threshold to reach for the specified purpose and interested individuals will pitch in, keeping the donation in an escrow fund. When the threshold is reached, the contributions are retired from the escrow fund and a contract is formed so that the collective good is supplied. If the chosen target is not gathered by the deadline, no funds are collected. Kickstarter is a crowdsourcing system where the task is contributing in a monetary way to fund a project or a cause. There are many other examples of systems similar to Kickstarter, and they usually come with the name of crowdfunding platforms.

As we stated, the range of possible tasks varies to a great extent and some of the systems share just few similarities with the purposes and scope that we set for our research (see Introduction). In particular, crowdfunding systems and online labor platforms don't even involve the sharing and the crawling of information. Although they are *crowdsourcing* systems according to our operational definition, they are not mean for building any kind of knowledge base. On the other hand, Wikipedia or Waze, have more similarities with our original purposes: they involve the collecting of information from an online community to build a knowledge base of some kind (for instance a freely available encyclopedia). Thus, it's legitimate to wonder why we decided to focus our attention toward such a wide field, including platforms with very different features and goals. The reason is that we wanted to encompass the whole *crowdsourcing* phenomenon, in all its declinations and formulations. We thought that thanks to this analysis we could get more data and knowledge in order to better understand and thereby exploit, the concept for our goals. We wanted to understand which kind of people compose the *crowd*, that following our definition is basically an online community, and what motivate these people to join it. In the wake of this, we decided to study also the reasons that move the people to join the communities of online labor platforms and crowdfunding and check to what extent their reward mechanisms can be applied to our case. Moreover, we thought that only the studying of the whole *crowdsourcing* phenomenon would have made us able to understand which are the underlying key factors that power both the *crowdsourcing* model and its formulation in term of *wisdom of the crowds*.

For the same reasons we also briefly analyzed the open-source environment. Indeed, it is clear that it is possible to draw some kind of parallelisms between the two phenomena and various authors pointed it out (Brabham, 2008a; Enkel *et al.,* 2009). According to our operational definition, open-source can be considered a particular application of the *crowdsourcing* model in which we lack the external subject allocating the tasks but instead the developers' community regulates itself and choices the tasks to be accomplished. The open-source movement is interesting to us because the study of the motivations that move people to join open-source communities share a lot of similarities with the study of the motivations that encourage people engagements in many others *crowdsourcing* projects.

In the next paragraph we will discuss more in deep the role that our work holds in the *crowdsourcing* scientific playground and what links it shares with other researches.

2.3.2 Literature review on crowdsourcing

In this section we will try to offer a brief but comprehensive overview of the researches that have been carried out in the field of *crowdsourcing* and that we have covered for our purposes. Moreover, we will draw several links between these studies and studies belonging to other fields of research such as open-source, psychology and online communities. Our final goal is showing in which position we can put our work in the scientific playground. In this paragraph we will not cover the studies over the *sentiment analysis* because we will discuss them in the next section.

Crowdsourcing is a relatively new phenomenon but various authors already tried to carry out some studies on this topic. The majority of these studies lack a complete vision of the *crowdsourcing* concept. Instead, they focus on specific case studies that

they analyze in details, usually providing many considerations and statistical facts. Examples of this approach are the works focusing on specific *crowdsourcing* communities like Brabham (2008b, 2009) and Chanal & Caron-Fasan (2008). All these researches offer an introduction to the *crowdsourcing,* with definitions and descriptions, but they don't try to completely analyze the phenomenon. Their goal is not developing a comprehensive model of the *crowdsourcing* but providing detailed studies of how some successful *crowdsourcing* systems achieved their missions.
On the other hand, the work of Brabham (2010), Schenk & Guittard (2011), Erickson (2011) and Buecheler *et al.* (2011) tried to offer a first complete insight on the *crowdsourcing.* They tried to model the phenomenon and find the underlying key factors. Their main goal is developing frameworks to describe the *crowdsourcing* projects that may help designers and developers in this field. These works have been the main subjects of our researches because we share with them the goal of building a comprehensive framework. This framework should, in our mind, allow us to effectively model new *crowdsourcing* applications with strong theoretical groundings (see Introduction). The reason why we decided to build our own new model (namely the *descriptive* and *prescriptive frameworks* presented in Chapter 3) is that we thought that these earlier works didn't provide a frame complete enough for the *crowdsourcing* environment. Often they propose too simple taxonomies to embrace all the possible facets that the *crowdsourcing* can assume. For instance, Buecheler *et al.* (2011) propose a framework in which the projects are analyzed according to four main dimensions: what, who, why and how. Malone *et al.* (2009) propose a larger model with more categories and declinations act at describing collective projects. Still also these models lacks what we considered to be a fundamental passage: the analyses of the motivations and the reasons that move the people to engage with *crowdsourcing* communities. Indeed, we wanted to analyze how to design *crowdsourcing* application capable of stimulating as much as possible the user participation. All these studies offer several brief comments on this topic. But no work is strong enough on this. We thought that there was the chance of developing a better framework embracing all the previous experiences of earlier researches. Each of the surveyed studies focused on a small subset of the underlying

factors that run the *crowdsourcing*. Schenk & Guittard (2011) focus on the type of tasks that can be crowd-sourced and the several techniques that can be used to choose the solutions among the many proposed by the *crowds* (see Chapter 3). Brabham (2010) focuses more on the motivations of the users joining *crowdsourcing* communities. Finally, Laubacher & Dellarocas (2009) introduce the concept of data quality mechanism.

What we wanted was a framework that not only included all the results of these works but also tried to show which are the links between the various features of the *crowdsourcing* (that we call dimensions in our frameworks) and how they influence each others, always keeping in mind the main goal of stimulating user engagement. Finally, we wanted to add a stronger theoretical grounding to the researches on the *crowdsourcing* phenomenon. Indeed, the *crowdsourcing* has been often covered without the sufficient scientific attitude. The studies usually lack a common jargon and use different terms to refer to the same concept. We wanted to develop a framework providing a common ground from which starting the discussions over the *crowdsourcing*. Moreover, we wanted to link the analysis of the *crowdsourcing* with the researches carried out in other fields in this way adding even more theoretical roots. At this scope, we surveyed the works on the incentives and rewards mechanisms used in corporation and management theory (for instances, von Hippel & von Krogh, 2006; Lakhani *et al.*, 2006) looking for a correlation with the rewards used in online communities. For the same reason, as we already partly stated, we also took into account studies on the open-source movements (for instance, Bonaccorsi & Rossi, 2004; Lakhani & von Hippel, 2003; Lakhani & Wolf, 2003), on the online communities phenomenon itself (for instance, Nov, 2007) and of psychology (Calder & Staw, 1975).

Finally, we were driven also by the consideration that no study analyzed all the possible types of *crowdsourcing* applications. Indeed, all the studies that we surveyed focus just on specific categories of *crowdsourcing* systems such as crowdfunding or online labor platforms. Erickson (2011) drafts a raw taxonomy of the various types of *crowdsourcing* applications but his work is just at an early stage and at the moment it's not possible to check if it will be really effective. Instead, we tried to embrace all the possible facets offering a categorization of the various forms that the *crowdsourcing* can assume.

Our research also had a strong experimental part. For what we know almost all the academic literature on the *crowdsourcing* phenomenon lacks on this side. One attempt is the work of Ling *et al.* (2005) in which they tried to statistically study which are the most effect method to motivate users in online communities. We completed our theoretical results with an experimental analysis of a vast dataset of real *crowdsourcing* applications and with the results of an experiment we conducted to test various *crowdsourcing* scenarios. In this way we also drew a link between the theoretical studies over the *crowdsourcing* with the practical analysis of IT systems, such are the web applications we developed as possible *crowdsourcing* scenarios (see Chapter 4). Indeed, although the *crowdsourcing* is a topic belonging to the IT domain, we found out that the majority of the studies don't consider this connection and prefer to study it as a mere social phenomenon completely unlinked to its natural environment, i.e. the Internet. Good exceptions to this approach are the work of Bretzke & Vassileva (2003) and Cheng & Vassileva (2005b).

2.4 Sentiment Analysis

2.4.1 Introduction and description

Communication through the Web in all its forms (websites, blogs, forums, social networks) has become an important way of forming and spreading opinions. For this reason the analysis of web content has become an important field of study for several kinds of applications, such as industrial marketing or anti-terrorism.

The *sentiment analysis* is a topic belonging to the broader field of the natural language processing (NLP, Natural Language Processing). *Sentiment analysis* is about analyzing human-redacted text in order to understand the attitude (positive, negative or neutral) of the author on the subject he is writing about. The recognition of opinions is not an easy problem because it requires the semantic comprehension of a text, domain knowledge and high-quality language processing skills. We can consider the traditional text classification and the *sentiment analysis* as two orthogonal issues: in the traditional text classification, the focus is on identifying the topic discussed in the document, while in the sentiment classification the focus is on the assessment of the sentiment expressed by the author of the document (Aue & Gamon, 2005). The later task is more complex for several reasons: we need to identify the object of which the author expresses his sentiment; we must identify if in a document the author expresses more than one sentiment on the same object and, in case, how this group of opinions is correlated; the *sentiment analysis* must deal with sarcasm, irony and rhetorical forms. We can conclude that the *sentiment analysis* involves all the steps that are usually performed in text classification, plus all the required processes to assess the sentiment.

Sentiment analysis is becoming a fashionable topic because of the great amount of human-redacted texts that are available today thanks to the *Web 2.0* platforms. The users of *Web 2.0* platforms daily provide a great amount of information about the most disparate subjects by writing on blogs and posting on social networks or on web forums. This amount of data can be analyzed through sentiment analysis techniques to extract information useful for several purposes, such as marketing or political surveys. Industrial marketing perhaps, is the major beneficiary of these analyses. Indeed, much of the content shared by Internet users contains opinions on products and services on the commercial market. These opinions can be gathered to study the overall sentiment of the people about goods and commercial brands. We call this process *Web reputation* analysis or *opinion mining*. The automatic *Web reputation* analysis allows the firms to stop leveraging the traditional instruments of collection of the customer satisfaction and to exploit faster and cheaper mechanisms based on IT technologies. Moreover, the *sentiment analysis* and its declination in automatic *Web reputation* analysis allows the gathering of real-time data and this can represent a vital competitive advantage for the companies acting in aggressive markets.

The algorithms for performing *sentiment analysis* can be classified according to various dimensions: the task of the analysis, i.e. the classification of text into positive or negative categories; the level of classification (sentence or document); whether the analysis is mainly based on syntax (Pang *et al.*, 2002), mainly using dependency parsing (Matsumoto *et al.,* 2005) or semantics, using corpora (such as WordNet) and then extending them manually or automatically (Whitelaw *et al.,* 2005); the techniques used in the algorithms, e.g., Support Vector Machine (SVM) or Naive Bayes (NB).

2.4.2 Sentiment Analysis Engine: description

In the previous section we presented the field of *sentiment analysis* and we discussed its meaning and scope. The majority of the semantic sentiment analysis tools have a common architecture design that is the result of the evolution of the academic and industrial research in this field. Thus, in this section we will discuss this general architecture.

As we already stated, we will focus specifically on tools for the *Web Reputation* analysis, as we described it in the previous paragraph. However, often throughout this thesis we will generally refer to a sentiment analysis tool. The readers must keep in mind that in that cases we are anyway referring the subset of tools of *sentiment analysis* that perform *Web reputation* analysis.

Web reputation analysis tool can be used to analyze the social media (Twitter[3], LonelyPlanet[4], TripAdvisor[5]) and extract the sentiment concerning a brand. These tools monitor the platforms, crawling the data posted by the users, and automatically extracts the information concerning a subject (the brand). Moreover, the analysis is

[3] Twitter is an online social networking and microblogging service that enables its users to send and read text-based posts of up to 140 characters, informally known as "tweets." [http://www.twitter.com]

[4] Lonely Planet's online community is used by over 600,000 travellers for trade tips and advice. The Lonely Planet website includes blogs, a groups platform and the ability to rate and review sites and restaurants. [http://www.lonelyplanet.com]

[5] TripAdvisor is the world's largest travel site that assists customers in gathering travel information, posting reviews and opinions of travel related content and engaging in interactive travel forums. It is a pioneer of user-generated content. The website services are free to users, who provide most of the content. [http://www.tripadvisor.com]

not just about an overall brand, but these tools can also discern among posts about various topic related to the brand.

The tools supporting the *Web Reputation* analysis can be split into two categories: semantic tools, and not semantic tools. The semantic tools analyze the text and information crawled from the Web, according to a semantic interpretation of the natural language. On the other hand, the tools belonging to the second category cannot interpret the natural language, but instead they try to be competitive by fast crawling and analyzing a great amount of data. Surely, these tools are more limited that the semantic ones because the analysis is performed without taking into account the semantic of the information.

We focus only on tools of *Web reputation* or *sentiment analysis* that can exploit semantic analysis. The Figure 1 shows the typical architecture of one of these tools. We will brief describe it. It is not our goal to propose a detailed description of these instruments because, as we stated, our research focuses on the harnessing of *crowdsourcing* techniques to build a knowledge domain to be used by a generic tool (see Introduction). Thus, we will focus on the building of domain knowledge through *crowdsourcing* and we will discuss this topic in the next paragraph.

The architecture is composed of four main parts that are by themselves divided in smaller components, as it's possible to see in Figure 1. The main parts are: crawling, data quality, sentiment analysis and user interface. They correspond to software modules. The crawling software module takes care of gathering the information from the social platform and storing it in raw format in a database. The crawler module is itself split in three components matching the three platforms that are monitored by our example tool. The data quality module takes care of taking the raw data coming from the crawler and performing some transformations to allow a better *sentiment analysis*. It assesses the quality of the information according to their sources, checks the language detecting slang and jargon words and finally prunes the database from unneeded data. The main module is of course the sentiment analysis one. It is the module that actually analyzes the text trying to understand the sentiment of the opinion expressed. Moreover, it also exploits semantic analysis to understand the object to which the sentiment refers. Finally, at the end of the chain we have the user interface for performing and presenting the analyses. We can see that the sentiment analysis module has a component called domain data. This is the knowledge base used by the tool to perform semantic analyses. Indeed, to be able to understand the meaning of the information, the tools must be aware of a lot of information concerning brands, names, products, etc. For instance, the tool must be aware that if a tweet posted on Twitter contains the word "Milan A.C." this is referring to the football club and not to the city of Milan itself. Thus, it appears clear that the domain knowledge is a fundamental database and that it is of great interest to keep it updated and as much as possible exhaustive. In the next section we will show how the domain data is designed and we will introduce a methodology to use *crowdsourcing* for enhancing it.

Figure 1: A *Web reputation* analysis tool typical architecture design.
The arrows represent the data flux and the box represents the main components

2.4.3 Domain modeling and crowdsourcing

The majority of researches developed in the *sentiment analysis* and *Web reputation* field focus on the core problem of understanding the attitude expressed by a natural language text. Much less has been said on the minor problem of building a model around the subject (the brand) whose we want to analyze the reputation on the Web. In order to better understand the topic first we need to present the terminology and give the definition of the terms brand, model, etc. Then we will propose a literature review of the works on this field and we will trace a link with the *crowdsourcing* phenomenon showing which is the direction of our research effort.

Roughly speaking, the model of a brand is a formal description of a brand, including the set of characteristics and features that we want to monitor. Consider for instance, as subject for the *sentiment analysis* (what we named the brand), the city of Milan. Thus, in this case, the model of the brand is the set of categories in which a *Web reputation* analysis tool splits the crawled information coming from the users. A brand is always collocated in a well-determined domain, thus designing a model for a brand also means building a representation of its domain. We can definite the domain as the part of the real world that refers to the brand. It is composed by the elements that are correlated to the brand in the real world. For better understanding this concept consider the following examples: if we want to study the *Web reputation* of a mobile phone company, the brand will be the company name and/or nicknames, the model of the brand will be the set of features we want to study, such as the quality of its products or costumer services, and the domain will be represented by the names of its phones, by the components of its phones, by the categories of products it sells and so on. The concept of domain described here is the same concept that we named knowledge domain in the previous paragraph. Thus, in our example architecture, the information concerning the domain of a brand is stored in the domain data component of Figure 1. To each domain corresponds a specific vocabulary of terms that are used in its context. The tool uses this vocabulary to solve the lexical ambiguities in which it may encounter while analyzing a text. We already discussed the example of the word “Milan” that can represent the city or the football club.

Another example is the Italian word “calcio”: it may refers to the sport, to the act of kicking or finally, to the chemical element. It’s clear that only an updated and comprehensive vocabulary can allow a *Web reputation* analysis tool to perform high quality sentiment analyses.

Some studies, while not explicitly focusing on domain modeling, still try to determine the elements that characterize a particular application domain and they do so through feature extraction systems, i.e. bottom-up methodologies to perform mining on text expressing opinions on a brand in order to deduce the main features of it. One of these studies is Balahur & Montoyo (2009) that takes as input a user query containing the brand of which the user wants to know the *Web reputation*. Then the system, using the knowledge extracted from ConceptNet[6] or WordNet[7], automatically discovers a series of features typical of the brand. Spangler, Proctor & Chen (2008) use a different approach that attempts to infer from the volume of speech the most important features of the brand, without relying to any a priori knowledge. The end result is a set of distinct categories of characteristics that describe the brand of interest. These categories cluster the terminology of a specific brand (like products names, nicknames and components).

We can see from these studies that there is an agreement in the academic literature that a semantic system of *sentiment analysis* is highly dependent on knowledge domain (Jeong *et al.*, 2009). Thus, a sentiment analysis tool is usually vertical focused on a specific set of brands belonging to the same domain, although there even been attempts to develop systems that easily adapt to new contexts (Andreevskaia & Bergler, 2008 ; Aue & Gamon, 2005).

The approach to domain knowledge building of the majority of sentiment analysis tools currently available is manual. A tool of this type doesn’t automatically build a domain representation using some of the automatic techniques presented previously,

[6] WordNet is a lexical database for the English language. [http://wordnet.princeton.edu]

[7] ConceptNet is an artificial intelligence project whose goal is to build and utilize a large commonsense knowledge base from the contributions of many thousands of people across the Web. [http://www.media.mit.edu/~hugo/conceptnet/]

but the domain must be manually provided by the administrators before starting to use tool. Thus, before using the tool it must pass a customization phase. This approach presents some shortcomings. First, it requires a great effort to design a comprehensive and exhaustive model. Second, it requires to keep the domain knowledge up to date. Moreover, the architecture of these tools, usually involves that the crawler component uses part of the domain knowledge to crawl new data from the social platforms. Indeed, the crawling is based on keywords detection in the analyzed text. These keywords derive from the data included in the domain knowledge (they correspond to the sub-brands). Table 2 shows the typical representation of the domain knowledge used by a *Web reputation sentiment analysis* tool.

Table 2: An example of domain knowledge representation for a *Web reputation* analysis tool

Concept	Description
Brand	The object of sentiment analysis. Examples are: Milan, Berlin, London or Nokia, Vodafone, etc.
Category (and sub-category, with as many levels as needed)	Taxonomy of the concepts related to a brand. For instance in the case of the brand of the city of Milan this can involve the category of “Food & Drink” and the sub-category of “restaurant”. The brand Nokia can involve the category “mobile phone” and the sub-category “UMTS mobile phones”
Sub-brand	Sub-brands are the concepts contained in the categories and sub-categories. For instance, if the brand is Apple and the category is “mobile phone”, a possible sub-brand can be “Iphone 4”. The sub-brands are used by the crawler component, together with the brand name, as keywords for crawling data from the social platforms.

We propose a different approach to populate, keep updated and extend the domain knowledge. Indeed, we saw that *crowdsourcing* techniques can be used to gather knowledge coming from the people. Moreover, we saw that a large network of individuals usually detains a great amount of knowledge. We think that these two factors together can be exploited to build an automatic system capable of extending and keeping updated the a domain knowledge for a sentiment analysis tool, through the harnessing of the data coming from large a large *crowd*. In Chapter 4, we will show in details how we built our approach.

In particular we think that a *crowdsourcing* online community can be used to extract information regarding the brands and the sub-brands. We studied several different mechanisms to tap *crowdsourcing* in order to obtain domain knowledge All these mechanisms (the scenarios, see Introduction), will be presented in Chapter 4.

Other studies in the academic literature tried to tap the *wisdom of the crowds* and the *crowdsourcing* to build a knowledge base for specific domain or of common-sense fact. Von Ahn & Dabbish (2004) developed a *crowdsourcing* game to acquire common-sense knowledge from the players (see also von Ahn *et al.*, 2006). They addressed the problem of collecting a database of “common-sense facts” using a computer game. Informally, a common-sense fact is a true statement about the world that is known to most humans: “milk is white,” “touching hot metal hurts,” etc. Several efforts have been devoted to collecting common-sense knowledge for the purpose of making computer programs more intelligent. Such efforts, however, have not succeeded in amassing enough data because the manual process of entering these facts is tedious. They therefore introduced Verbosity, a novel interactive system in the form of an enjoyable game.

Hsueh *et al.* (2009) instead focused on tapping the *crowdsourcing* to build a knowledge base of sentiment annotations (positive, negative, neutral) of snippets. Annotation acquisition is an essential step in training supervised classifiers. However, manual annotation is often time-consuming and expensive. Thus, they investigated the possibility of recruiting *crowds* of annotators through Internet online communities as an appealing option that allows multiple labeling tasks to be outsourced in bulk, typically with low overall costs and fast completion rates. They particularly focused on the problem of assessing the over quality of this process and of the data coming from the *crowds*. Moreover, they conducted an empirical study to examine the effect of noisy annotations on the performance of a sentiment analysis tool.

Chapter 3

Modeling crowdsourcing applications

3.1 Introduction

In the previous chapter we discussed the concepts of *Wisdom of the crowds, Sentiment Analysis* and *Crowdsourcing*. In particular we outlined several important definitions that are already well grounded in the academic literature, and significant examples of the application of *crowdsourcing* techniques in a wide variety of circumstances.

In the wake of the literature review carried out in Chapter 2, we realized that there is a substantial dearth of a deep and complete empirical study of the current applications of *crowdsourcing*. In this chapter we fulfill this lack by developing a research framework for modeling *crowdsourcing* projects. This is the second step of our research methodology as we outlined it in the first chapter.

Our research should follow the same path that has been traced by Lakhani & von Hippel (2003) and Lakhani & Wolf (2003) for the open-source software. In their researches they assess the open-source collaborative environment focusing on the motivations and the incentives that encourage the people to engage in collective projects and provide their resources. Two are the goals we set for this part of our work:

- provide a *descriptive framework* for assessing and modeling existing *crowdsourcing* projects
- provide a *prescriptive framework* in the form of *best practices* obtained from the analysis of existing *crowdsourcing* projects according to the previous framework. The idea behind is that these *best practices* should be used as guidelines while designing new applications exploiting *crowdsourcing,* and provide a further set of information to aid the design of the experiment (see Chapter 4)

First, we build an empirical dataset of existing *crowdsourcing* applications and projects. Second, we outline the set of dimensions that shape our *descriptive framework*. These dimensions come from a comprehensive inquiry of the existing literature on *crowdsourcing* while taking into account also contributions from other research domains that we regarded as meaningful (Psychology, Theory of Management and Economics).

Last, we try to draw some conclusions from the analysis of the empirical dataset exploiting our *descriptive framework*. In particular we outline a taxonomy of the *crowdsourcing* projects included in our empirical dataset. The synthesis of these analysis and conclusions forms the *prescriptive framework*.

3.2 Empirical Dataset

3.2.1 Selection of the sources

While choosing the members of our empirical dataset we faced the problem of having a precise and coherent definition of what can or can't be considered *crowdsourcing*. We have already cited in the second chapter the Howe's (2008) brief description of *crowdsourcing*:

"Crowdsourcing is the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call."

Moreover, in Chapter 2 as well, provided our operational definition of what is the *crowdsourcing*.

"Crowdsourcing is a model in which a collection of IT technologies, such as software and telecommunication networks, are exploited to build a virtual platform for outsourcing a certain collection of tasks to a distributed pool of individuals (the crowd) that are willing to perform these tasks"

These definitions give a general idea of what means exploiting the *crowdsourcing* and we provided many details and examples in the previous section. Nevertheless, for our aim, namely gathering a selection of sources for an empirical dataset, these definitions are still too vague and generic. Eventually, in order to populate our empirical dataset we decided to include only applications that match these key characteristics:

1 They are crowdsourcing projects according to Howe's (2008) and our operational definitions
2 They provide one or more coherent and systematic mechanisms to support the exchange of the information from and to the crowd
3 They are based on software programs (including Web-based applications)
4 They are actively used by a community of people, well-established and with a stable implementation
5 They are task-oriented
6 They are open to everyone

The first characteristic is listed because the Howe's definition of *crowdsourcing* was the starting point of our research and also because it is widely accepted as characterization of the term. As a result the other outlined requirements are used to better specify the membership function of our empirical dataset and to further restrict our field of research. In particular we focus only on computer applications (3° characteristic). Indeed, as discussed in the previous chapter, throughout history there have been various examples of *crowdsourcing* applications not exploiting the Web or even the information technology (IT) (see Surowiecki, 2004). These legacy examples are out of the context of our research.

The second characteristic states that the applications in the empirical dataset must have an interface to support the *crowdsourcing*. Namely it means that the applications must exploit the *crowdsourcing* in a systematic way using a standardized set of tools (interfaces, algorithms, technology backend, etc.). Obviously this is an unavoidable requirement to allow us to perform a coherent scientific analysis on the dataset. The characteristics 4 and 6 are self-explicative: we focus just on systems that are currently active and that are not subject to revolutionary renovations. Moreover they should be freely available to everyone allowing us to test and study them. Something more should be discussed about point 5. This requirement states that the applications must allocate the work to the *crowd* in the form of *tasks*. For our matters a task is a unit of work or a set of actions that can be accomplished in a finite amount of time. A task can be complex or simple, time-consuming or trivial, etc. The definition doesn't commit to any specific type of work (Schenk & Guittard, 2011).

It's possible to raise the question if the list of requirements constraining the membership to the empirical dataset can undermine the validity of the *descriptive framework* and, consequently, of the conclusions built over it. Indeed this problem can be ruled out: the dataset, even bounded by the previous assumptions, still contains a significant share of the existing *crowdsourcing* applications. Indeed any *crowdsourcing* project meets the first characteristics by definition. Furthermore the majority of the projects are either Web-based or simple desktop applications as showed by the examples in the previous chapter. This property forces them to meet the second characteristic almost in any case (counter-examples can be found in Surowiecki, 2004, and in Jones & Rafaeli, 1999). The requirements 4, 5 and 6 don't change the situation significantly. As a result the dataset still remains relatively heterogeneous and representative of the landscape.

The list of candidates has been drawn looking for existing catalogs of applications and through a survey of the available literature. Wikipedia (2011), aBitAbout (2011) and crowdsourcing.org (2011), all offer comprehensive lists of *crowdsourcing* projects. Likewise Brabham (2008a), Frei (2009), von Ahn & Dabbish (2004), Chanal & Caron-Fasan (2008), Shaw *et al.* (2011), Malone *et al.* (2009), Schenk & Guittard (2011), present some other interesting candidates in their works.

Only a subset of the candidates has been included in the empirical dataset, namely the ones matching the features we have already discussed and of which we could obtain enough data of our aim.

3.2.2 Dataset Description

In this section we list some statistical information concerning the empirical dataset.

The dataset consists of 61 *crowdsourcing* applications. The greatest majority of them (56/61) are Web-based projects. The remaining part is made of desktop applications (games, as we will discuss in the next section). Table 3 summarizes the data so far presented.

Table 3: Distribution of the type of applications in the dataset

Web-based applications	56
Others applications	5
Total	61

We also collected the country of origin of the elements in the dataset. The country of origin of a *crowdsourcing* project is the nation where the firm or the developing team that created the application is placed. For a small part of the set we couldn't find exact information: some projects don't have a precise developing team or the team is spread in various locations or the project is not backed by any company. This is

mainly true for open-source collaborative project like Wikipedia or Distributed Proofreaders[8]. All the applications in the dataset, independently from the country of origin, use English as main language.
Figure 2 shows this information in a chart.

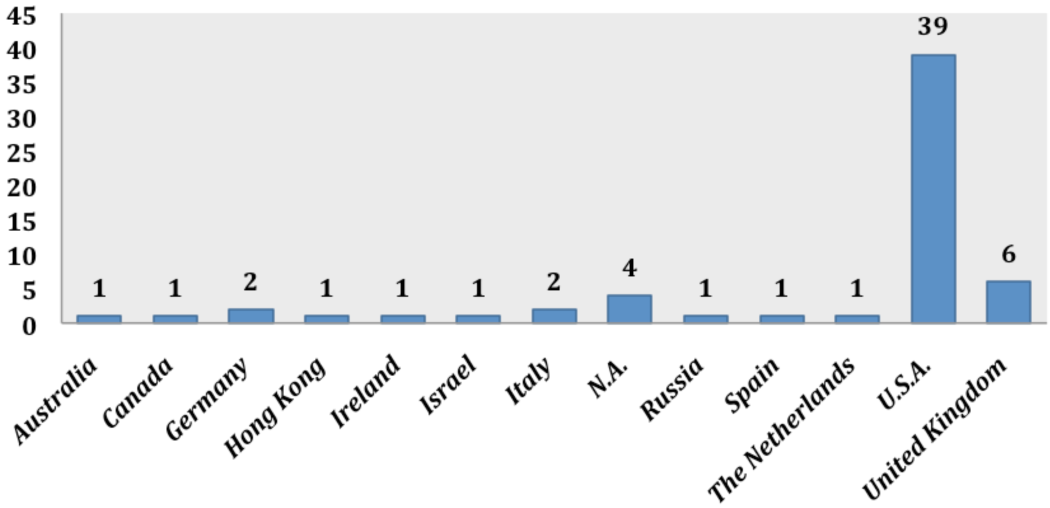

Figure 2: Geographical distribution of the applications in the dataset

N.A. stands for not available

While collecting data for the elements in the dataset we couldn't get precise numeric information in all the cases. In some cases we had to estimate these missing information or proceed without them. We can already state that the missing data doesn't affect the overall quality of the framework. Indeed the missing information usually is the country of origin or the size of the user-base (this will be discussed in the next section). These records are not imperative to perform all the analysis. Table 4 provides some statistics on this topic.

[8] [http://www.pgdp.net] Distributed Proofreaders is a web-based project that supports the development of e-texts by allowing many people to work together in proofreading drafts of e-texts for errors.

Table 4: Quality of the data concerning the applications in the dataset.
***Due to rounding, the percentages may not add up to 100%**

Applications of which we have complete precise data	33	54%
Applications of which we estimated some data	4	7%
Applications of which we lack some data	28	46%

Appendix A contains a detailed list of the elements in the empirical dataset with attached the sources used for gathering the records. Appendix A contains also a short description of the functionalities of every application in the set. The description takes the form of a list of keywords. These keywords are self-explicative and allow the readers to easily understand which are the aims of the *crowdsourcing* project under observation. For instance, Table 5 shows the keywords for two elements in the dataset: Amazon Mechanical Turk and Wikipedia (see Chapter 2 for their descriptions).

Table 5: List of keywords for two *crowdsourcing* projects in the dataset

Amazon Mechanical Turk	Task allocating, job marketplace
Wikipedia	Collective encyclopedia

Following a common practice introduced by the *Web 2.0*, Figure 3 shows a *tag cloud*[9] of the keywords describing the elements in the dataset.

[9] A *tag cloud* is a visual representation for text data, typically used to depict keyword metadata (tags) on websites, or to visualize free form text.

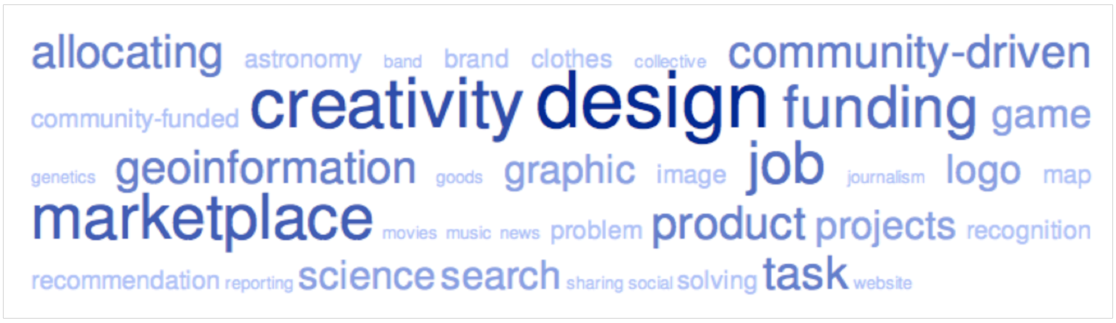

Figure 3: ***Tag cloud*** **of the keywords describing the elements in the dataset**

3.3 Descriptive Framework

3.3.1 Introduction

The *descriptive framework* is the novel systematic approach that we propose to assess and analyze real *crowdsourcing* applications. Our research focused on selecting and identifying the most fitting variables that have to be taken into account while studying the *crowdsourcing* phenomenon. The result is a set of ten dimensions and for any of them the associated metric. These dimensions come from a comprehensive inquiry of the previous academic works in the fields of *Web 2.0, Psychology, Online Communities, Knowledge Management, Knowledge Sharing, etc.* As we have already outlined in the second chapter *crowdsourcing* is not just a sub-domain of the computer science but instead, as far as it intrinsically requires a profound interaction with a user-base (the *crowd*), it has a great deal also with human factors that transcend the mere technology. User motivations, incentives, rewards are key concepts that can be understood only taking into account topics that cross the borders of IT. Lakhani & Wolf (2003) already pointed out this conclusion while dealing with open-source phenomenon. As consequence several dimensions in the *descriptive framework* are new adaptations of already existing concepts in domains, at first glance, distant from our fields. We didn't take existing concepts *as-is* but we renewed them to meet the requirements that the modeling of *crowdsourcing* applications brings. This approach from one-side gives a solid scientific background to our analysis and from the other side doesn't commit us to a too inflexible model. Indeed some dimensions are not reflexed in any previous research but come out from our considerations. The result is a *descriptive framework* that encompasses if not the whole range of *crowdsourcing* projects, at least the majority of them. Obviously it seamlessly covers the empirical dataset we presented in the previous sections. This is straightforward as the dataset played a double role in our research: it was the test field for assessing the goodness of the dimensions imported from domains other than the *crowdsourcing* and it gave us a set of real cases from which extract new dimensions and their metrics.

In the next section every dimension will be discussed illustrating the rationale and the analysis process behind it with punctual references to the academic literature when needed. We will also try to offer some operational guidelines and "real life" examples to better explain how to use the metric of the dimensions in the framework. As a summary Table 6 shows the list of all dimensions specifying also their metrics.

Table 6: The dimensions of the *descriptive framework*

Dimension Name	Metric
Categorization	Collective Knowledge, Knowledge Sharing, Collective Creativity, Cloud Labor, Knowledge Acquisition, Crowdfunding, Open Innovation, Problem Solving
Crowdsourcing Type	Integrative, Selective
Required Knowledge	Low, Medium, High
Community Size (Quantitative)	>0, N.A.
Community Size (Qualitative)	Small, Medium, Big, N.A.
User Type	Amateur, Professional
Task Type	Simple, Complex, Game
Main Reward	Enjoyment-based, Opportunistic, Prestige-oriented
Minor Reward	Enjoyment-based, Opportunistic, Prestige-oriented, None

Remuneration (Quantitative)	Numeric Range, N.A.
Remuneration (Qualitative)	Low, Medium, High
Incentive	Sharing of the result, Sharing of the goal, User ranking and voting systems, Position inside community and user power scaling, Money, Competition
Data Quality Mechanism	Group Evaluation [Voting], Group Evaluation [Averaging], Group Evaluation [Consensus], Reward Accuracy, Competition, Surveillance, None

3.3.2 Dimension Description

Categorization

Categorization is the first and most important dimension of the *descriptive framework*. As the name suggests it already provides a first taxonomy of the *crowdsourcing* applications assessed by our novel framework. Its importance lies on the fact that correctly categorizing a *crowdsourcing* project is a key step before moving to the other dimensions. In the next section, when the result of the analysis on the dataset will be presented, it will clearly emerge that the category to which a member belongs, significantly influences the values of the other dimensions.

While taking into account these remarks the framework tries to provide an as much as possible reliable characterization of the dimension. To fulfill this requirement it offers eight different values to describe a *crowdsourcing* application.

The framework doesn't try to invent from scratch new categories but instead tries to follow the prevalent ideas that come from the industry and the academia. Nevertheless the framework distinguishes itself from previous works because it brings a rigorous and systematic approach to the dealing of the problem. The main effort is aimed at preciously defining the values for the dimension.

The possible values for the metric are: Collective Knowledge, Knowledge Sharing, Knowledge Acquisition, Cloud Labor, Collective Creativity, Crowdfunding, Open Innovation, Problem Solving.

In the *crowdsourcing* projects that fall in the Collective Knowledge category the *crowdsourcing* is used to develop and aggregate knowledge and information through open Q&A, user-generated knowledge systems, social news systems, social forecasting, etc. (Crowdsourcing.org, 2011). In other words in these applications the *crowdsourcing* is used to acquire and/or share information from and to the *crowd.* For instance, Wikipedia falls in this category (for a description of Wikipedia internals and structure see Kolbitsch & Maurer (2006) and Malone *et al.* (2009)).
Knowledge Sharing and Knowledge Acquisition are two possible specifications of the Collective Knowledge category and they are bound to it. In simple words we can state that in Knowledge Sharing projects the knowledge coming from the *crowd* goes back to the *crowd.* In Knowledge Acquisition projects this doesn't happen and the flux of information is one-way, going from the *crowd* to a different agent. Wikipedia is an example of Knowledge Sharing, because the *crowd* can access to the generated knowledge. Get A Slogan[10] instead is an example of Knowledge Acquisition because the *crowd* can't access to the knowledge generated by other contributors.
Cloud Labor is the leveraging of a distributed virtual labor pool, available on-demand to fulfill a range of tasks from simple to complex. *Crowdsourcing* is used to connect labor demand and supply (Crowdsourcing.org, 2011)). For instance, Amazon Mechanical Turk falls in this category (Alonso *et al.*, 2008; Ipeirotis 2010a, 2010b) offer a comprehensive introduction to Amazon Mechanical Turk design). At support of Cloud Labor as a separated category we can recall Quinn & Bederson (2009); indeed they claim that Cloud Labor ("Mechanized Labor") completely changes the dynamics of *crowdsourcing* and proposes to categorize it in a separate cluster.
Problem Solving is a specialization of the Cloud Labor category and is strictly bound to it. Problem Solving embraces the *crowdsourcing* applications in which the focus is not on the information but on the execution of tasks. In this paradigm the *crowd* is

[10] Get A Slogan is a crowd-sourced slogan development service. [http://www.getaslogan.com]

not called to share their knowledge but to help in solving problems of different kinds. CrowdSpirit[11] is an example (Chanal & Caron-Fasan, 2008).

Collective Creativity *crowdsourcing* applications tap into a creative talent pools to design and develop original art, media or content. *Crowdsourcing* is used to tap into online communities of thousands of creatives to develop original products and concepts, including photography, advertising, film, video production, graphic design, apparel, consumer goods, etc. (Crowdsourcing.org, 2011).

Open Innovation is a concept similar to Collective Creativity but it declines in a different way. We define Open Innovation as the usage of the *crowdsourcing* paradigm to address sources outside an entity or a group in order to generate, develop and implement new ideas (Schenk & Guittard, 2011). This refers mainly to scientific research. An example is InnoCentive (see Chapter 2 for a description). The central idea of Open Innovation is that in a world of distributed knowledge, companies should not only rely on their own research and development but also try to use the resources coming from outside their borders (Chesbruugh, 2003)). According to this view Open Innovation falls inside the *crowdsourcing* paradigm and for this reason the framework inserts it among the possible categories.

Crowdfunding is an approach to the raising of monetary capital for new projects and activities (including business) by soliciting contributions from a large number of stakeholders following several types of models: donations or sponsorship where there is no expected financial return, lending or investment in exchange for equity, profit or revenue sharing ((Crowdsourcing.org, 2011)). An example of this paradigm is Kickstarter (see Chapter 2 for description).

It's clear that the values may overlap in some cases and that the borders between one category and the other are not that sharp. So it's possible that an application falls into more categories at the same time. For this reason the framework allows using multiple values for this dimension.

---

[11] CrowdSpirit is a *crowdsourcing* community built around designing electronic products. Users submit ideas for innovative electronic products that the community votes on. The best ideas rise to the top where investors provide financing. [http://www.crowdspirit.com]

Table 7 can help while categorizing a *crowdsourcing* application according to the taxonomy proposed in our *descriptive framework.* It consists of a set of questions and examples to be used in assessing a system.

Table 7: Questions for correctly assessing the category of a *crowdsourcing* system

Category	Questions	Examples (see Appendix A)
Collective Knowledge	• Does the system provide a mechanism exploiting *crowdsourcing* to gather and aggregate information and/or knowledge coming from the user community? • Does the system use user-generated knowledge systems such as wikis or Q&A? • Does the system allow the users to signal news, websites or other form of information?	- Wikipedia - Waze - Delicious - Urtak
Knowledge Sharing	• Does the system allow the users to access freely to the knowledge base built through *crowdsourcing*?	- Wikipedia - Waze
Knowledge Acquisition	• Does the system collect the information coming from the *crowd* and pass it to other agents outside the *crowd* without sharing freely the knowledge built by its community?	- Article One Partners - Urtak
Cloud Labor	• Does the system offer a mechanism to allow agents demanding work-time and job/tasks execution to meet crowds supplying it?	- Amazon MT - CrowdSpring
Collective Creativity	• Does the system offer a mechanism to create online communities of creatives developing original products (art, design, movies, advertising, music, etc.) allowing agents outside the community to tap this creative pool?	- Threadless - BoobB - Zooppa - iStockphoto
Crowd-funding	• Is the system used to globally collect and raise financial fund from people interested in investing or donating their money for some goals?	- Kickstarter - Pledgemusic
Open Innovation	• Is the system used to exploit *crowdsourcing* in order	- InnoCentive

	to collect ideas or advanced solutions globally? • Is the system used to ask people to study scientific problems? • Is the system used to bring innovation and new developments in any research field using idea coming from the crowds?	- Philoptima
Problem Solving	• Is the system used to ask the crowds help in order to solve problems such as software testing, data collecting or testing of models?	- TopCoder - uTest

As concluding remarks Figure 4 shows the Categorization of the elements in the empirical dataset.

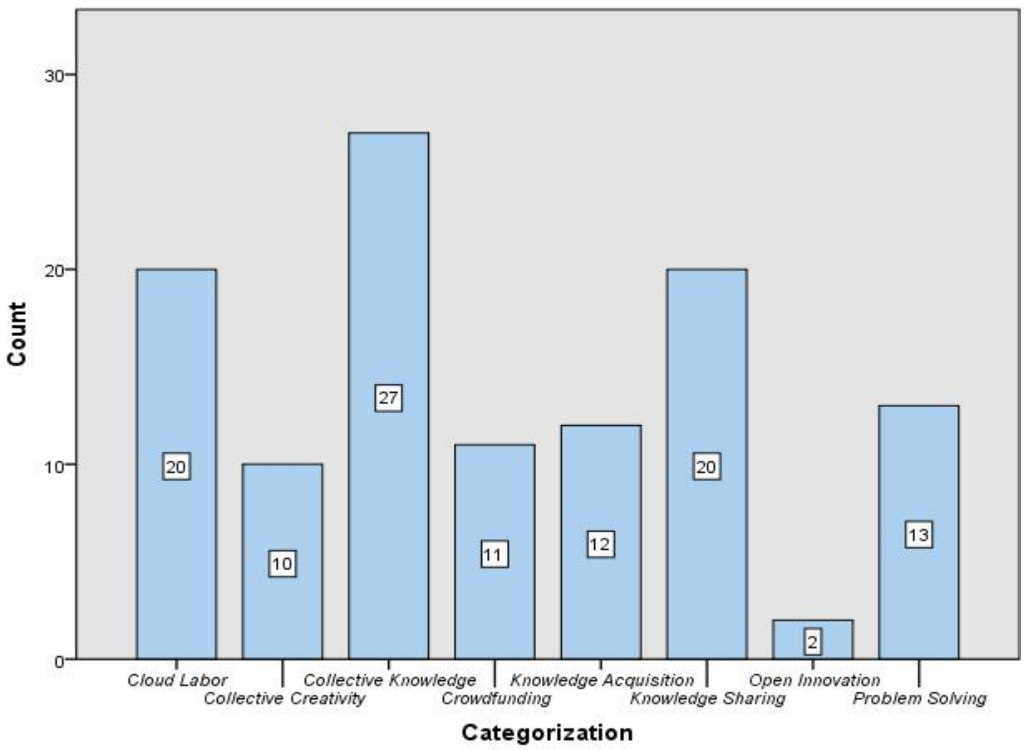

Figure 4: Categorization of elements in the empirical dataset

***Due to elements belonging to more than a category, the count add up to more than 61**

Crowdsourcing Type

Following the path traced by Schenk & Guittard (2011), this dimension distinguishes between two different situations in which the *crowdsourcing* paradigm is used. In the first case *crowdsourcing* is used to accumulate multiple and complementary information or data. We call this behavior Integrative *crowdsourcing*. Examples of this behavior are Wikipedia, Delicious[12] or Waze. In these applications user contributions are aggregated to form a collective database of information: Wikipedia keeps track of every change in its knowledge based, Waze instead keeps on collecting geographical data coming from its user-base.

In the second case only a subset of the information coming from the *crowd* is taken into account. We name this behavior Selective *crowdsourcing*. This situation happens frequently in applications that fall in the category of Cloud Labor, Problem Solving or Open Innovation, according to the previous taxonomy. In these applications the focus is not on the information itself but on the execution of a task or the solution of a problem. The way the subset is selected can vary greatly as consequence. Selective *crowdsourcing* generally implies a winner-takes-all mechanism where only the finder of the "winning" solution is rewarded and the others are discarded. This is the case of InnoCentive for instance. Counter-examples to this behavior come from Get A Slogan, where the "loosing" solutions are not discarded but kept in a virtual catalog for a possible future usage.

Table 8 summarizes the discussion concerning the Task Type through a set of operational definitions and can be used to assess this dimension for a *crowdsourcing* implementation.

[12] Delicious is a social bookmarking web service for storing and sharing web bookmarks. [http://www.delicious.com]

Table 8: The metric for the dimension Crowdsourcing Type

Metric Value	Definition
Integrative	*Crowdsourcing* is used to accumulate multiple and complementary information or data. User contributions are aggregated to form a collective database of information
Selective	Only a subset of the information coming from the *crowd* is kept. Usually this involves a set of criteria to select the best or most suitable data

Figure 5 shows the Crowdsourcing Type of the elements in the empirical dataset.

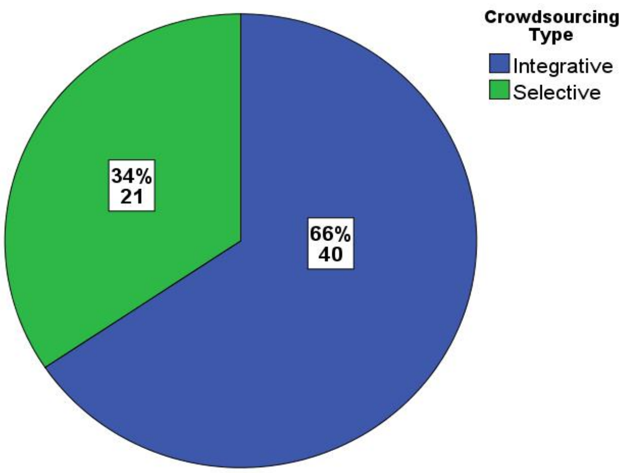

Figure 5: Crowdsourcing Type of the elements in the empirical dataset
***Due to rounding, the percentages may not add up to 100%**

Required Knowledge

Being part of the community of a *crowdsourcing* project and actively collaborating to it requires a certain amount of effort. In particular we can state that a user willing to join a *crowd* has to detain some basic skills and knowledge. These skills can be for example the ability to perform some tasks (like translating from a language to another) or to use some specific software and technology. The order of magnitude of the required knowledge varies according to the specific applications and contexts. The required knowledge represents the entry barriers that a user has to overcome in order to take part to a *crowdsourcing* project. The entry barriers can vary within a single project. For example the entry barriers in Wikipedia can be quite low if we consider a user willing to bring just little modifications to an article but become considerable if instead he would like to write a new article from scratch.

The Required Knowledge dimension models this property of the *crowdsourcing* paradigm. It can assume the following values: low, medium, high or any combination of the previous. The reason why it may assume multiple values is expressed well by the Wikipedia case: multiple values define a range for the required knowledge. So if some tasks can be accomplished with low entry barriers and others instead with high entry barriers, then the Required Knowledge dimension will assume both the low value and the high value. It's important that while modeling a *crowdsourcing* project both the lowest and the highest required knowledge find expression in this dimension. This is the criterion we followed for our empirical dataset.

For our purposes we need to specify what can be considered a low, a medium or a high requirement of knowledge. Clearly the definition will be abstract and not too stringent because the number of cases varies a lot and it's hard to bind this variability to only three precise values for the dimension. We looked at our dataset and we tried to assess the lowest and the highest knowledge requirements among all the applications in the set. Starting from these two extremes we have been able to empirically define what is a medium required knowledge. Considering the extension of our dataset and it's heterogeneity we can assume with good confidence and by induction, that our judgments on the scale of the Required Knowledge dimension are valid also for applications not included in the empirical dataset. Updating your position on Waze or donating money on Kickstarter has low entry barriers. Proposing new clothes design on Threadless has medium entry barriers (see Chapter 2 for a description). Finally as we already discussed, writing an article for Wikipedia presents high entry barriers.

Table 9 summaries the concepts exposed above and can be used as a fast reference while assessing a *crowdsourcing* system.

Table 9: The metric for the dimension Required Knowledge

Metric Value	Examples
High	Write an article for Wikipedia, design a new advertising campaign, develop a new software application
Medium	Translate short texts, test software
Low	Rate a movie or a band, provide traffic jam information, donate money

Figure 6 shows the Required Knowledge of the elements in the empirical dataset.

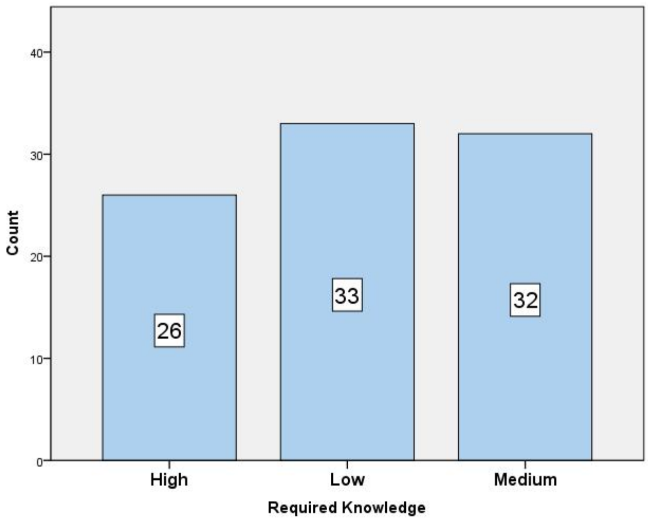

Figure 6: Required Knowledge of the elements in the empirical dataset
***Due to elements belonging to more than a category, the count add up to more than 61**

Community Size (Quantitative and Qualitative)

It should be clear from what we discussed in Chapter 2 that the *crowd* plays a vital role in *crowdsourcing* applications. Thus, it's essential for the *descriptive framework* to provide a metrics to describe the *crowd*. The Required Knowledge dimension can help in determining the user-base because it states which are the entry barriers that the people in the *crowd* must be able to meet prior of joining the community. However this characterization is not enough. The *descriptive framework* proposes other measures at this regard, namely the Community Size presented here and the User Type presented in the next section. The Community Size expresses how big is the user-base of a *crowdsourcing* application, both in a quantitative and a qualitative way. To assess how many people belongs to the *crowd* we need to specify in a rigorous way what we consider to be the *crowd* for our modeling purposes. In Chapter 2, we widely discussed the term *crowd* and we pointed out the commonly accepted definition. Here we provide more details.

The Quantitative Community Size is the average number of active contributors that belong to the community engaging with the *crowdsourcing* project. This number expresses the "size" of the *crowd*. Thus, the *crowd* is composed by active contributors. An active contributor is a user who provides his resources by mean of the *crowdsourcing* application. The resources can be information, work-time, problem solving abilities, money, etc. The users who passively benefit from the knowledge generated by the user-base are not part of the *crowd* and should not be counted while assessing the community size. In the same way, the agents who exploit the *crowdsourcing* paradigm with the intent of gaining resources should not be counted as well. As an example consider the usual Wikipedia. The majority of the users of its website are just passive readers and they don't contribute in any way to the knowledge base (Holmes, 2006), therefore they are not part of the *crowd* in this case. Only the small part of users who are also writers or editors is counted. Furthermore consider Amazon Mechanical Turk as a case of Cloud Labor. Here we have a set of agents who propose tasks to be accomplished by the *crowd*. These agents leverage the *crowdsourcing* platform but they are not part of the *crowd* according to the definition we gave before. In the majority of cases assessing the exact dimension of the *crowd* and coming out with a number can be really troublesome, especially when the community is not static but varies with the time. For this reason it's sufficient to assess an "average" number of people belonging to the user-base. In Appendix A it's possible to see that we used this approach for almost all the elements in our empirical dataset.

The size of the *crowd* is a fundamental measure in the *crowdsourcing* paradigm. Surowiecki (2004) argues that besides the requirements of "openness", "peering", "sharing" and "heterogeneity" (see Chapter 2), the size of the *crowd* is a key factor for the emerging of the *collective intelligence.* Small group of people fails in making intelligence decisions or performing complex tasks, where big groups collaborating with the right instruments, can succeed. Lorenz *et al.* (2011) argue that decreasing heterogeneity in social group can undermine the *wisdom of the crowds* effect and weaken the accuracy of the crowd-sourced data. Therefore under the hypothesis of openness to online contributions, the community size matters when it comes to heterogeneity. The Web in particular, is an ideal technology for aggregating millions of disparate, independent ideas without the dangers of 'too much communication' and compromise, in this way avoiding heterogeneity. In addition, we argue that the size of the user-base is a value for technical reasons: the technologies empowering the *crowdsourcing* paradigm require a consistent mass of users to be effective. We refer in particular to the mechanisms to validate the quality of the crowd-sourced information. Eventually different numbers of contributors influence different *crowdsourcing* paradigms where it comes to data quality mechanism, incentives and rewards. These remarks will be discussed while presenting the *prescriptive framework* because they are part of our general conclusions on the modeling of *crowdsourcing* systems. At the moment they support our thesis on the importance of the Community Size dimension. Surowiecki (2004) claims that there is a *critical mass* of users. Without the critical mass the system will not meet the expectations of quality of the outcomes. Albors *et al.* (2008) express the same concept while analyzing the sustainability of online communities. Finally Sharratt & Usoro (2003) enlarge this vision and point out that a critical mass of users can attract new users in the community; otherwise there is the risk of loosing them. Thus, the community size itself is an incentive to join the community.

The Qualitative Community Size dimension derives from the Quantitative Community Size. We introduced it mainly for allowing the clustering of different elements in our dataset in this way aiding further analysis. Moreover we were not able to collect exact size information for some elements in the set, but anyway we were able to estimate it in a qualitative way using various sources on the Web (see Appendix A for details). Thus, a qualitative measure for the size was welcome. Qualitatively the community size can assume three values: small, medium and big. As usually to define the correct meaning of these values we started from the information we had from our dataset. In particular we proceed in this way:

1. We created a set containing the quantitative community sizes of the elements in the empirical dataset
2. As we stated we couldn't retrieve this information for all the elements. We retrieved 36 cases, corresponding to 59% of the elements in the empirical dataset
3. We removed from the sample, the lowest and highest values
4. We computed the median value of the remaining cases, namely M = 566000 ≈ 500000 users
5. We took as medium size the range [M/2 ; M*2], extremes included, as small size the range (0 ; M/2) and as big size the range (M*2 ; ∞)

Figure 7 and 8 show the size distribution of the elements in the empirical dataset.

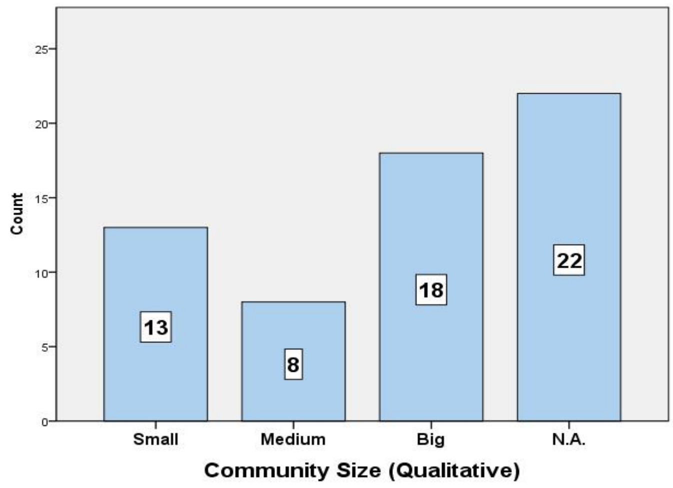

Figure 7: Qualitative Community Size of the elements in the empirical dataset

N.A. stands for not available

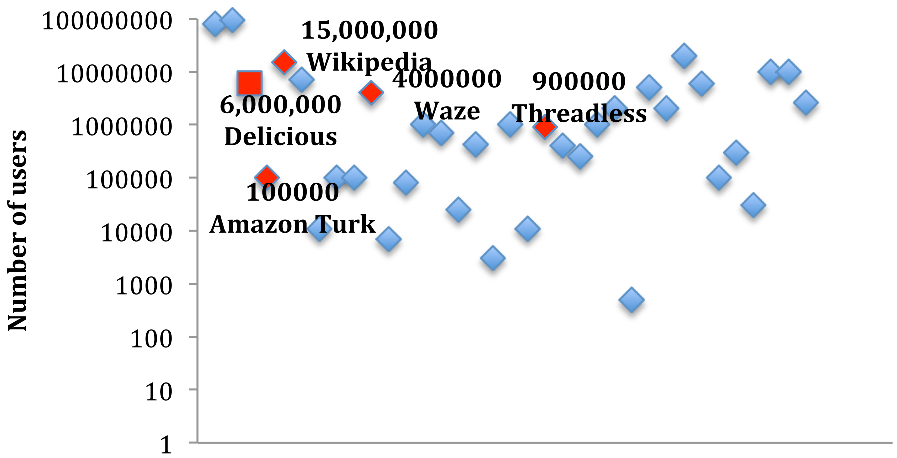

Figure 8: Quantitative Community Size of the elements in the empirical dataset

***some elements are highlighted, logarithmic scale**

User Type

Our *descriptive framework* distinguishes between two types of users belonging to the *crowd:* amateur users and professional users. In our context, an amateur user is someone who belongs to the community of a *crowdsourcing* project and performs the tasks without specific professional training or education. At the opposite side a professional user can apply a prior knowledge coming from professional education or schooling while contributing to a *crowdsourcing* community. It's important to stress that the definition doesn't commit to any financial incentives that *crowdsourcing* systems may offer.

Many authors studied the composition of the *crowds* according to this taxonomy. Howe (2008) coined the term "*crowdsourcing*" and he actively tethered the image of amateur or hobbyist to the *crowdsourcing* paradigm. Likewise, Schenk & Guittard (2011) point out that although professionals are not excluded a priori from *crowdsourcing*, by nature they are more likely to function in classic outsourcing process. Thus, they argue that the *crowd* is mostly composed by amateur participants. On the other end, Brabham questions this vision and brings a survey he made on iStockphoto[13] in which results that the 58% of the surveyed contributors had at least a year of formal schooling in art, design and photography (Brabham, 2008b, 2009).

We claim that the nature of the user type (amateur or professional) is deeply tied to the nature of the *crowdsourcing* application and to the nature of the task that has to be crowd-sourced. Amateurs and professionals can concur in composing the *crowd.* Therefore we stand aside from both Howe's vision of a *crowd* mainly composed of amateur and the Brabham's one of a predominant share of professional contributors.

The User Type dimension reflexes this position. Indeed it is not tied to assume just one value. We believe that the "openness" property of *crowdsourcing* projects (especially the ones in our dataset as we stated before) allows almost in all the cases the amateur contributors to engage in the projects. As we discussed in Chapter 2, this is part of the nature of *crowdsourcing.* However some *crowdsourcing* applications present such high entry barriers (see Required Knowledge) and complexities in the tasks that they are practically accessible just to professionals. An example is

[13] iStockphoto is an online, royalty free, international micro-stock photography provider operating with the micropayment business model. [http:///www.istockphoto.com]

InnoCentive where the *crowd* is mainly composed by Phds or Professors (Lakhani et al. 2006).

Table 10 summarizes the discussion concerning the User Type through a set of operational definitions and can be used to assess this dimension for a *crowdsourcing* implementation.

Table 10: The metric for the dimension User Type

Metric Value	Definition
Amateur	Amateur user performs the tasks without specific professional training or education
Professional	Professional user can apply a prior knowledge coming from professional education or schooling while performing the tasks

Figure 9 shows the User Type distribution of the elements in the empirical dataset.

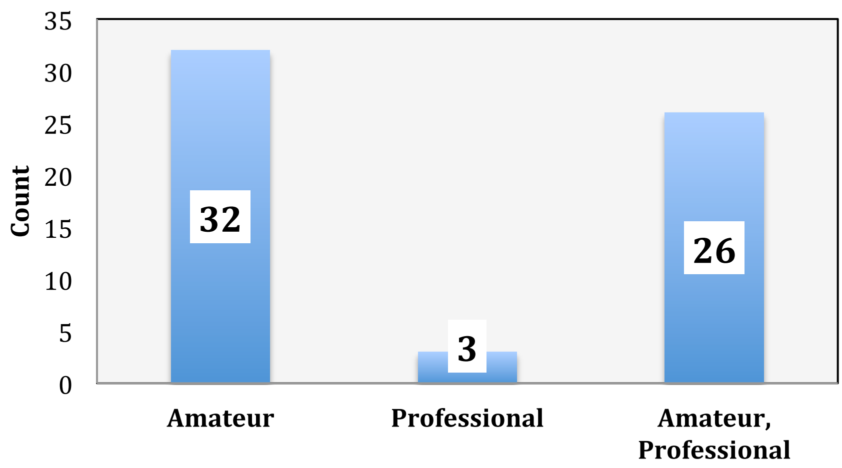

Figure 9: User Type of the elements in the empirical dataset

Task Type

Schenk & Guittard (2011) try to answer the question "What can be crowd-sourced?" proposing a three-categories taxonomy of the type of tasks than can be crowd-sourced: simple, complex and creative tasks. The examples we discussed in Chapter 2 should have already revealed that the nature of the tasks varies to a large extent. *Crowdsourcing* may be used for simple tasks such as data collection and translation of texts, or instead *crowdsourcing* can be implemented to achieve complex tasks (e.g. problem solving). Simple tasks are rather poor from a cognitive point of view and their completion requires a relatively low involvement from the individuals. At the other end complex tasks involves knowledge intensive activities and may require advanced skills such as software developing or filmmaking.

In our *descriptive framework* we recall the same taxonomy proposed by Schenk & Guittard (2011) but we discard the notion of creative task and instead we propose the notion of game task. According to Schenk and Guittard, creative tasks imply the creation of original creative contents involving arts and design production. In our opinion creative tasks can't form a category by themselves because any creative task is always either simple or complex. Thus, the category is completed overlapped by the other two and doesn't present any characteristic of originality. Game tasks were not taken into account by the work of Schenk and Guittard. However while surveying the *crowdsourcing* applications on the market, we found several examples of systems that tap gaming to collect the *wisdom of the crowds*. Gaming is a type of task that falls apart from the other two and presents traits of uniqueness that spurred us to consider it separately. Indeed a popular approach to motivating volunteers is to create a game that requires the player to perform some computation in order to get points or succeed. The idea is that since people play games online, it may be possible to divert that energy for some particular purpose. Quinn & Bederson (2009) use the term "Games with a Purpose" while discussing this concept.

A *crowdsourcing* implementation can also involve complex and simple tasks at the same time. For this reason the Task Type dimension isn't tied to assume a single

value; instead it can also assume the simple and the complex values together. For what we said before game task falls apart and it can't be mixed.
Table 11 summarizes the discussion concerning the Task Type through a set of operational definitions and can be used to assess this dimension for a *crowdsourcing* implementation.

Table 11: The metric for the dimension Task Type

Metric Value	Definition
Simple	Simple tasks are rather poor from a cognitive point of view and their completion requires a relatively low involvement from the individuals. They require few steps and short amount of time to be accomplished
Complex	Complex tasks involve knowledge intensive activities and may require many steps and/or a medium/long amount of time to be completed
Game	The task consists in playing computer games

Figure 10 shows the Task Type distribution of the elements in our empirical dataset.

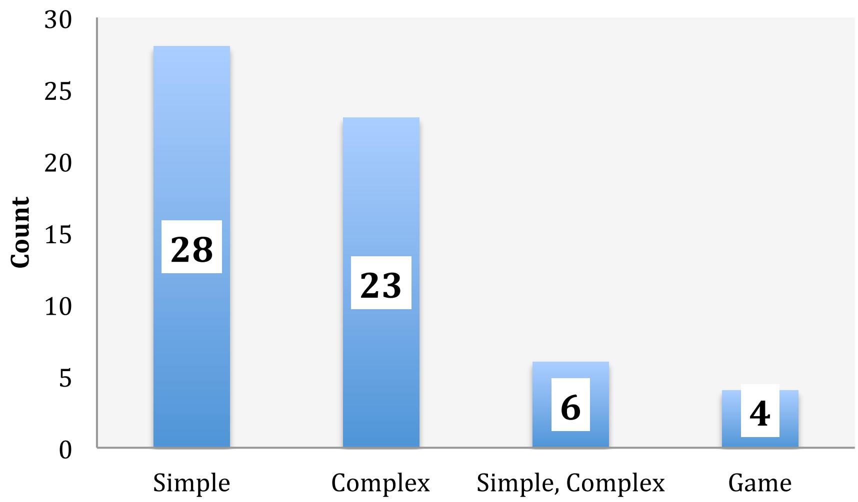

Figure 10: Task Type of the elements in the empirical dataset

Rewards (Main and Minor)

Describing the *crowdsourcing* phenomenon requires an insight of the reasons that move the users in the *crowd* to take part to collaborative projects. Assessing these reasons is a fundamental process in order to efficiently implement a *crowdsourcing* system capable of exploiting them properly. We already discussed in Chapter 2 and in the previous sections that the quantity of data collected from the *crowds* greatly influences the quality of the outcomes produced by *crowdsourcing*. Roughly speaking a greater amount of gathered information corresponds to an overall better quality of the data coming out from the *crowdsourcing* (Surowiecki, 2004). Thus, stimulating user participation and contribution is vital.

The motivations that push people to join communities have been deeply studied in many fields of research much time before the digital and Internet revolution. Questions about human motivation have been central in philosophy, literature, economics, and psychology for centuries. It is impossible to do justice of this here. For our aim is sufficient to identify which are the central drivers and use them to shape a metric for our *descriptive framework*. We name the motivations *rewards* and we propose two dimensions for describing them: the Main Reward and the Minor Reward. These dimensions range over the same set and are forced to assume just one value. The rationale behind the choice of splitting the rewards description into two dimensions lies in the fact that it allows us to better analyze how the rewards affect each other and how they are linked. In particular the Main Reward dimension expresses which is the most important driver of user participation while Minor Reward describes the secondary reason. Minor Reward may also assume a "none" value if the *crowdsourcing* system exploits only one reward mechanism.

The ways in which individuals are gratified by being part of online communities, such as the ones powering *crowds*, partly differ from individuals' reasons for participating in offline communities such as firms or work places. Historically, studies of the incentives for increasing the performance of employees working in big organization have focused on extrinsic motivations. The traditional rule was "run a firm as if it were a set of markets", meaning rewarding employees according to their marginal productivity and relying on extrinsic rather than intrinsic motivations (Prendergast, 1999). The economists have contributed the most to our understanding of how extrinsic motivations drive human behavior. The classical human behavior economic model for example, is based on incentives applied from outside the person and following the principle that "people change their actions because they are induced to do so by an external intervention" (Frey, 1997). Economic theory thus takes extrinsic motivation to be relevant for behavior and the money has been traditionally considered the main extrinsic motivation (Lakhani & Wolf, 2003). However traditional models based only on extrinsic drivers don't suffice in describing the reasons why people contribute to modern online communities (Brabham, 2009). Facing this problem, more recent studies have focused on the practice of open-source. In this production, users essentially work for free to create software (Coar, 2006), which in itself undermines the power of simple extrinsic motivators such as money. Several researches on motivations in open-source participation (Bonaccorsi & Rossi, 2004) point out that the primary motivator has to be related to the pleasure of doing hobbies. As Torvalds states, "most of the good programmers do programming not because they expect to get paid or get adulation by the public, but because it is fun to program" (Ghosh, 1998). Although similar in some aspects, open-source production is not the same as *crowdsourcing*. Indeed *crowdsourcing* participation relies both on extrinsic and intrinsic motivations. Intrinsic motivation is defined as the doing of an activity for its inherent satisfactions rather than for some separable consequence and because of external prods, pressures, or drivers (Lakhani & Wolf, 2003). Calder & Staw (1975) states that an intrinsic motivation "is valued for its own sake and appears to be self sustained". A *crowdsourcing* application relies more on extrinsic or intrinsic motivation according to its structure, goals and implementation. Cloud Labor *crowdsourcing* usually proposes a mechanism of rewarding completely based on direct monetary compensation so exploiting a model based mainly on extrinsic drivers. Instead a Knowledge Sharing community such as Wikipedia doesn't rely at all on money to stimulate user contribution. Nov (2007) conducted a survey of the reasons that drive *Wikipedians* and it results that the mains are in order of importance: Fun

("Writing/editing in Wikipedia is fun"), Values ("I feel it is important to help others"), Understanding (Writing/editing in Wikipedia allows me to gain a new perspective on things"), Enhancement ("Writing/editing in Wikipedia makes me feel needed"), Career ("I can make new contacts that might help my business or career") and Ideology ("I think information should be free"). It appears that besides merely intrinsic reasons connected to altruism and enjoyment (Values, Fun and Ideology), there are also motivations not directly connected to monetary compensation but still ascribable to it (Career), and related to the user acknowledgement in the community (Enhancement). In particular the role of career advancements as motivator for contributing to knowledge sharing communities has been studied also by Sharratt & Usoro (2003). Brabham (2009) asked to the Threadless *crowdsourcing* community "why do you participate on the site?" and collected the answers from a sample of 17 participants. It resulted that the user were driven mainly by the desire of "making money", "improving their creative skills" or "getting freelance opportunities" and by "the love of the community itself". Although the author admits that the sample is too small to allow general conclusions concerning the *crowdsourcing* phenomenon, it anyway supports a model of concurring intrinsic and extrinsic motivations.

We propose a slightly different taxonomy of the motivations that drive user participation. Our proposal is not grounded on the traditional split between extrinsic or intrinsic motivations but instead on a taxonomy composed by three main categories. These three categories encompass all the reasons reviewed before offering at the same time a more clear illustration of the drivers. Our approach derives from Malone, Laubacher & Dellarocas (2009). They argue that the main factors that spur user participation can be classified essentially in money, love and recognition ("glory"). We enlarge this vision and propose a classification in which the reasons can be opportunistic, enjoyment-based or prestige-oriented. The opportunist category covers the extrinsic motivations such as money but also intrinsic ones such as career development. The enjoyment-based category covers intrinsic motivations connected to altruism, fun, volunteerism, curiosity, etc. Finally, the prestige-oriented category follows the glory category definition of Malone *et al.* (2009).

Table 12 offers a complete vision of our taxonomy and can assist while assessing the rewards a *crowdsourcing* system.

Table 12: The Reward dimension (Main or Minor) metric in the *descriptive framework*

Main or Minor Reward	Description and declination
Opportunistic	Not Monetary: 1. Receiving a fair share of the result 2. Career related 3. Skills improvement Monetary 1. Direct: direct monetary compensation, micropayments 2. Indirect: improve future earnings (through skills and reputation improvements)
Enjoyment-based	Desire to do something different from your work Desire to express yourself Curiosity and desire to test if it works Values and Ideology 1. Volunteerism and desire to support a cause of the project 2. Reciprocity, exchange and mutual help Desire to communicate and to establish networks with other people Fun
Prestige-oriented	Desire to influence other people Increasing online reputation and recognition Desire of power and control

Figure 11 and 12 show the Main and Minor Reward distribution of the elements in our empirical dataset.

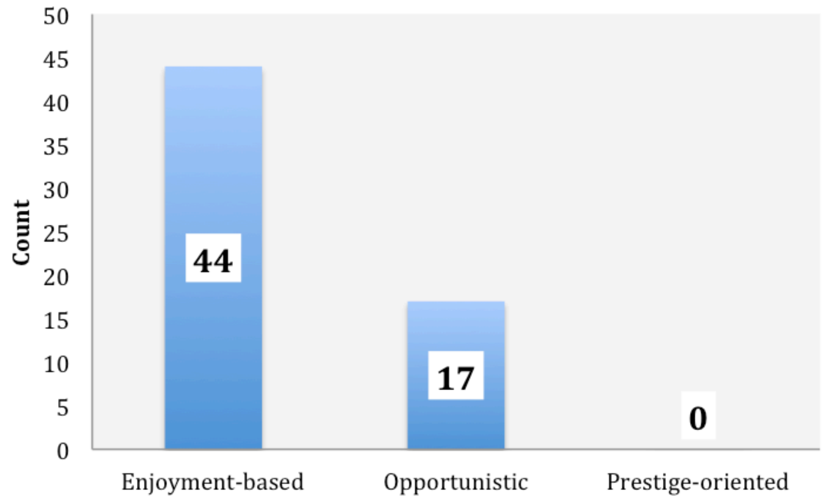

Figure 11: Main Reward of the elements in the empirical dataset

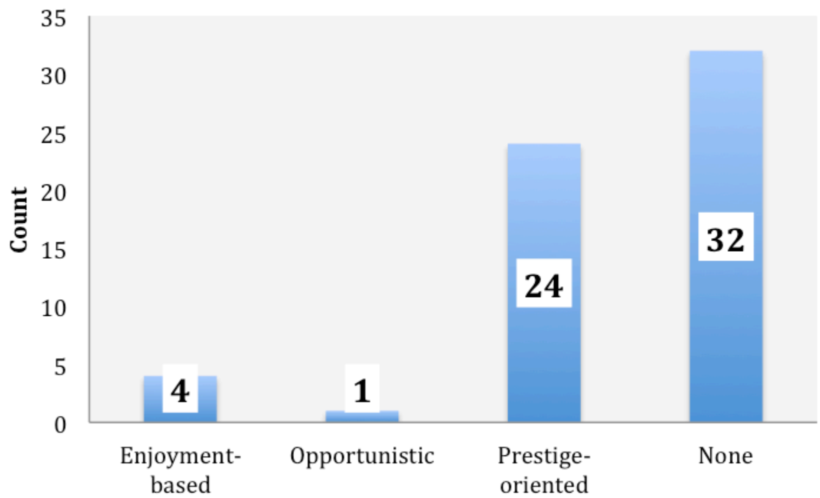

Figure 12: Minor Reward of the elements in the empirical dataset

Remuneration (Quantitative and Qualitative)

We discussed in the previous section that one of the motivations that moves people to join a community (Reward) is money and this is particularly true for the *crowdsourcing* ascribable to Cloud Labor. The Quantitative Remuneration dimension describes the amount of money that a *crowdsourcing* implementation offers to its contributors. According to a survey we conducted over the elements in our empirical dataset the remuneration can varies greatly among the various *crowdsourcing* services. It starts with the micropayments offered by Amazon Mechanical Turk and ends with the big prizes proposed by InnoCentive. Moreover Threadless uses a mechanism in which the user shares a fraction of the revenues collected by the service thanks to its creative works. Even among the same platform the direct monetary compensations can be different according to each task. For this reason the Quantitative Remuneration dimension is modeled as a range in the *descriptive framework*. For our empirical dataset we used the dollar as monetary unit but obviously the framework isn't commit to any particular unit. As we already saw for the Quantitative and Qualitative Community Size dimensions, we usually tether a qualitative dimension to a quantitative dimension. Remuneration isn't an exception to this norm. Thus, the Qualitative Remuneration derives from the Quantitative one. Qualitatively the remuneration can assume three values: low, medium and high. To define the correct meaning of these values we started from the information we had from our dataset and preceded with the same technique discussed for the community size. It resulted a range between (0 ; 100] for low remunerations (micropayment), a range between (100 ; 1000] for medium remunerations and finally a range between (1000 ; ∞) for high remunerations. Obviously as far as the remuneration of a *crowdsourcing* service can range to a large extent, the qualitative remuneration dimension can assume multiple values.

Figure 13 shows the Qualitative Remuneration of the elements in our empirical dataset.

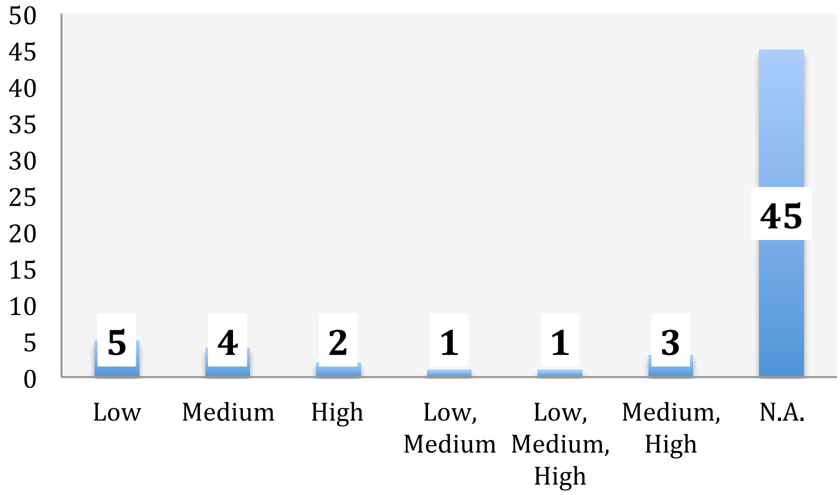

Figure 13: Qualitative Remuneration of the elements in the empirical dataset

N.A. stands for not available

Incentive

While discussing the reward dimensions we listed the motivations that move people to join a *crowdsourcing* community. It appears obvious that a *crowdsourcing* implementation should properly leverage this set of motivations in order to push as many people as possible to join its community and remaining over time. The Incentive dimension provides a description of the mechanisms that a *crowdsourcing* platform can use in order to effectively leverage the rewards discussed in the previous section. Thus, we could say that while the reward dimensions describe an implementation from the point of view of the user-base, the Incentive dimension uses instead the viewpoint of the developers.

The metric for the reward dimensions derives from an analysis of the academic literature. Thanks to this analysis we have been able to identify three main categories to which ascribe the motivators. Unfortunately we were not able to follow the same path also for the Incentive dimension because it hasn't been deeply covered by a relevant number of researches yet. Thus, we proceeded in a more experimental way

starting from the elements in the empirical dataset as did few times before and developing a set of cases depicting the metric for the dimension.
A first consideration is that every incentive is connected and exploits a subset of the reward mechanisms. The following table provides a summary of our conclusions. After listing the set of incentives we will provide a description of each of them with some references to the literature when possible. Table 13 can also be used as a set of operational descriptions while assessing a *crowdsourcing* application.

Table 13: Metric for the Incentive dimension

Incentive	Reward category	Description	Examples (cf. Appendix A)
Money	**Opportunistic**	▪ Offering money to the users in exchange of their contributions	▪ Amazon MT ▪ TopCoder ▪ Foursquare (indirect through badges)
User ranking and voting system	**Enjoyment-based, Prestige-Oriented**	▪ Implementing user-ranking systems according to the level of their contributions ▪ Implementing voting mechanisms allowing the users to express on others' contributions	▪ Threadless ▪ Get A Slogan ▪ Yahoo! Answers
Competition and Gaming	**Enjoyment-based**	▪ Introducing competition among the users for example by using gaming, contests or races	▪ Threadless ▪ gwap ▪ Cerberus game
Position inside the community and user power scaling	**Prestige-oriented**	▪ Implementing hierarchies of users with different powers and status in the community	▪ Wikipedia ▪ Waze ▪ OpenStreetMap
Sharing of the results	**Enjoyment-based**	▪ Allowing the users to access and enjoy the others' contributions	▪ Wikipedia ▪ Waze ▪ Yahoo! Answers

Sharing of the goals	**Enjoyment-based**	▪ Making the users aware of the goals of the *crowdsourcing*	▪ Wikipedia ▪ Phylo ▪ Fold.it

Money is the first incentive that we propose. It's straightforward that it is linked with the reward mechanism of direct monetary compensation. A *crowdsourcing* application that leverage on financial compensation for moving the people to take part to its community should relies on money as incentive. This is particularly true for Cloud Labor implementations. The financial incentives have been covered by Shaw *et al.* (2011), Mason & Watts (2010) and Chandler & Kapelner (2010). All of them agree that the money is the best incentive for leveraging the opportunistic expectations of the *crowd* and should be considered as the best incentive of extrinsic motivations. But Mason & Watts (2010) and Chandler & Kapelner (2010) point out that money alone is not a sufficient incentive for a *crowdsourcing* system. They conducted two experimental researches focusing on the workers of the Amazon Mechanical Turk community and found out that it's possible to consistently boost the user participation by providing other incentives alongside the money. In particular they discovered that explaining the reasons and the goals of the tasks to the users, greatly increases the level of their contributions. In other words, they linked an incentive for opportunistic rewards to incentives for enjoyment-based rewards. We model these discoveries in the Incentive dimension metric with the values Sharing the results and Sharing the goals. The later directly expresses the concept discussed above, namely to incentive the users by making them sharing the goals of our initiatives. Sharing the results instead is slightly a different case. In this situation the incentive leverage the fact that a user actively participating to the *crowdsourcing* community will be able to enjoy the result of its and others' contributions. Wikipedia it's an example of *crowdsourcing* that relies more on this incentive. The users are stimulated to write and editing articles because the platform allows them to enjoy the results of the work of the others. Wikipedia also exploits the Sharing the goals incentive because it makes the contributors aware of the importance of their work.

User ranking and voting systems are other incentives to stimulate user contributions and they usually come together thus we grouped them in one value of the dimension. They leverage both the prestige-oriented and enjoyment-based rewards mechanisms. Indeed user ranking can propel the recognition of the best participants within the *crowd* while voting systems can aid the emerging of the top users and contributions as well. Moreover these incentives can increase the “fun” in the participation exploiting the enjoyment side.
The Competition and Gaming incentive is similar to the previous User ranking and voting systems but more general. Some *crowdsourcing* systems can propose contests and races among the *crowds* in which the paradigm is winner-takes-all. Or as we already saw some can exploit competition through gaming. These systems may not use any user rankings or voting systems (for example in case the winner is selected by a single agent outside the *crowd*) and so completely lack a public recognition mechanism of the users. Thus, they don’t rely on prestige-oriented rewards but instead the competition boosts the enjoyment-based rewards, especially by increasing the fun side of the participation.
Finally the last incentive we propose is Position inside community and user power scaling. This incentive exploits the prestige-oriented reward mechanisms. Giving increasing powers to the users according to the quantity and the quality of their contributions stimulate their participation by leveraging all the three declinations of the prestige-oriented rewards: desire to influence other people, increasing online reputation and recognition and desire of power and control. Cheng & Vassileva (2005b) point out that introducing hierarchies of users inside a community increases the participation and the quantity and quality of contributions.

Figure 14 shows the Incentive distribution of the elements in our empirical dataset.

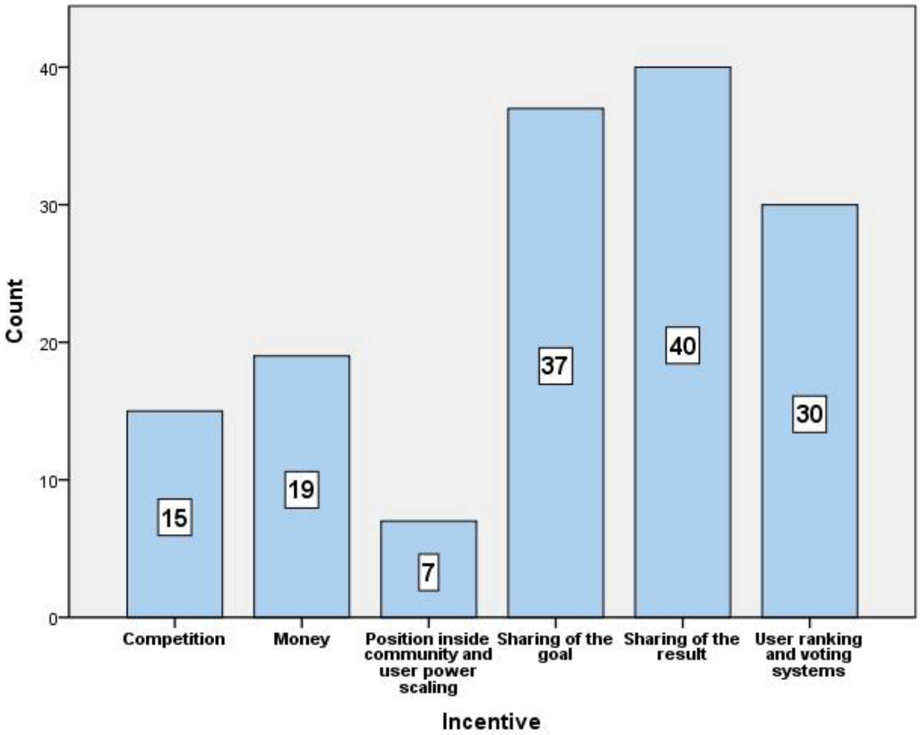

Figure 14: Incentives of the elements in the empirical dataset

***Due to elements belonging to more than a category, the count add up to more than 6**

Data Quality Mechanism

Crowdsourcing relies on individuals to accomplish its tasks and the quality of its results is greatly affected by the behaviors and the resources provided by the user-base. Thus, it's vital to provide a set of mechanisms to ensure an overall sufficient quality of the input and output data, in this way meeting the goals set for the *crowdsourcing* system. The basic concept behind *crowdsourcing* is that gathering a large amount of information from the *crowd* the result will be better than if just relying on small group people (this is extensively covered in Chapter 2). This is true even for Cloud Labor systems: the logic in this case is that exploiting the skills and the resources ("information" to a great extent) of many people it is possible to accomplish a task in a cheaper and better way. The problem is how to correct

aggregate all the data coming from the *crowd* and put it together through a reliable system.
The basic mechanism is averaging (Buecheler *et al.*, 2011; Surowiecki, 2004). According to this paradigm all the information coming from the *crowd* are weighted and mixed together to produce an output according to some formula. The averaging formula can vary by far. It may be a simple arithmetic mean of all the data rather than a much more complicated algorithm that considers several factors such as for instance, the time of contributions. Averaging is already able to offer a mechanism to ensure the quality of the crowd-sourced data. The reason relies on a mathematical truism: if you ask a large enough group of diverse, independent people to accomplish a task (whatever it is) and then average the results, the errors each of them makes in coming up with an answer or a solution, cancel themselves out (Surowiecki, 2004).
In some *crowdsourcing* implementation it's not possible or wanted to average the information produced by the *crowd.* Instead a mechanism to bring out the best data produced by the community would be a much better way to address the problem. A possible solution to this situation would be designing a system allowing the user-base itself to select and elicit the best data (or solutions) (Cheng & Vassileva, 2005a). This scenario can be implemented through voting and rating systems so that the best scoring data would automatically emerge (Buecheler *et al.*, 2011). This scenario can be described as a *double crowdsourcing system* because *crowdsourcing* is used both to produce the data and to select the best one among it. Some authors prefer to name it peer-vetted production (Brabham, 2010).
Another data quality mechanism is consensuses (Buecheler *et al.*, 2011). According to this a contribution to the knowledge base is subject to a continuous review by the community and the quality is ensured by the collective and steady process of reviewing and correcting. An example of this paradigm is offered by Wikipedia articles that can be edited by any user. It's important to underline that this paradigm can be implemented only by *crowdsourcing* applications that don't need to produce final static information but instead can rely on a dynamically changing knowledge base. Collective geographical information systems such as Waze (see Chapter 2 for a

description) or OpenStreetMap [14] can benefit the most from this mechanism. Moreover this mechanism allows the data to be kept updated by the user-base itself.

Cloud Labor and Problem Solving *crowdsourcing* systems often use a different approach to data quality. Usually in this environment only one solution among the many proposed by the *crowd* is selected (cf. Selective *crowdsourcing*) and the selection is made by a group of agents that is outside the *crowdsourcing* community. In these cases the quality is ensured by the fact that the agents will select only the solutions matching their criteria and therefore obviously satisfying their expectations. In our modeling framework we name this paradigm for data quality as Reward Accuracy. The name derives from the fact that in the majority of cases a winner-takes-all mechanism follows the selection process so that only the winning solution is rewarded. To foster even more the quality it's possible to use a broader Reward Accuracy system in which the winning solution is dynamically rewarded according to its performance and/or quality (Shaw *et al.*, 2011).

Shaw *et al.* (2011) argue that surveillance is another widely used mechanism to foster better quality in data coming from online communities. Surveillance can be implemented through automatic algorithms that check the information (Sinha & Swearingen, 2001). For example natural language processing algorithms are often used in *crowdsourcing* translation systems to detect unwanted words or expressions (Kolbitsch & Maurer, 2006). Other *crowdsourcing* applications can rely instead on manual surveillance. In this case a restrict group of people check that the crowd-sourced data complies with the system rules and guidelines. If the group of supervisors is selected among the *crowd,* manual surveillance can be combined with incentives such as position inside the community and user power scaling and a prestige-oriented reward mechanism (Cheng & Vassileva, 2005a).

The last data quality mechanism we propose for this dimension is competition, i.e. the competition among the users' productions. This paradigm by itself is not tied to any specific implementations but instead is a broader abstract concept. However we think that competition by itself is a mechanism capable of increasing the quality of

[14] OpenStreetMap is a collaborative project to create a free editable map of the world. [www.openstreetmap.org]

the data produced by *crowdsourcing* thus it should be listed among the possible values for metric of this dimension.

Finally we must observe that data quality mechanisms are important also to avoid malicious behaviors by the users. For example in *crowdsourcing* applications based on the gaming paradigm this can be a serious problem especially if they also offer some form of monetary compensation. In these situations a set of mechanism to avoid cheating is unavoidable (von Ahn & Dabbish, 2004).

Table 14 summaries the concepts presented above together with examples taken from the empirical dataset.

Table 14: Metric for the Data Quality Mechanism dimension

Data quality mechanism	***Description***	***Examples***
Group Evaluation [Voting] or Peer-vetted production	The user-base itself selects and elicits the best data (or solutions) produced by mean of *crowdsourcing*. This scenario can be implemented through voting and rating systems	▪ Threadless ▪ Delicious ▪ Digg
Group Evaluation [Averaging]	According to this paradigm all the information coming from the *crowd* are weighted and mixed together to produce an output according to some formula	▪ Foldit ▪ Gwap ▪ Phylo
Group Evaluation [Consensus]	According to this paradigm users' contributions are subject to a continuous review by the community and the quality is ensured by the collective and steady process of reviewing and correcting	▪ Wikipedia ▪ Waze ▪ Foursquare
Reward Accuracy	In this scenario only one solution among the many proposed by the *crowd* is selected and rewarded. To foster even more the quality it's possible to use a broader Reward Accuracy system in which the winning solution is dynamically rewarded according to its performance and/or quality	▪ InnoCentive ▪ Threadless ▪ Amazon MT ▪ Crowdspring
Competition	Introducing competition among the contributors independently from its implementation, can provide a	▪ Get A Slogan ▪ Cerberus game

	rough mechanism for ensuring data quality	▪ InnoCentive
Surveillance	Surveillance can be implemented through automatic algorithms that check the information or by selecting a group of agents for this purpose. In the later case a restrict group of people check that the crowd-sourced data complies with the system rules and guidelines	▪ Amazon MT ▪ Clickworker ▪ Cerberus game
None	No data quality mechanism is applied	

Figure 15 shows the Data Quality Mechanism distribution of the elements in our empirical dataset.

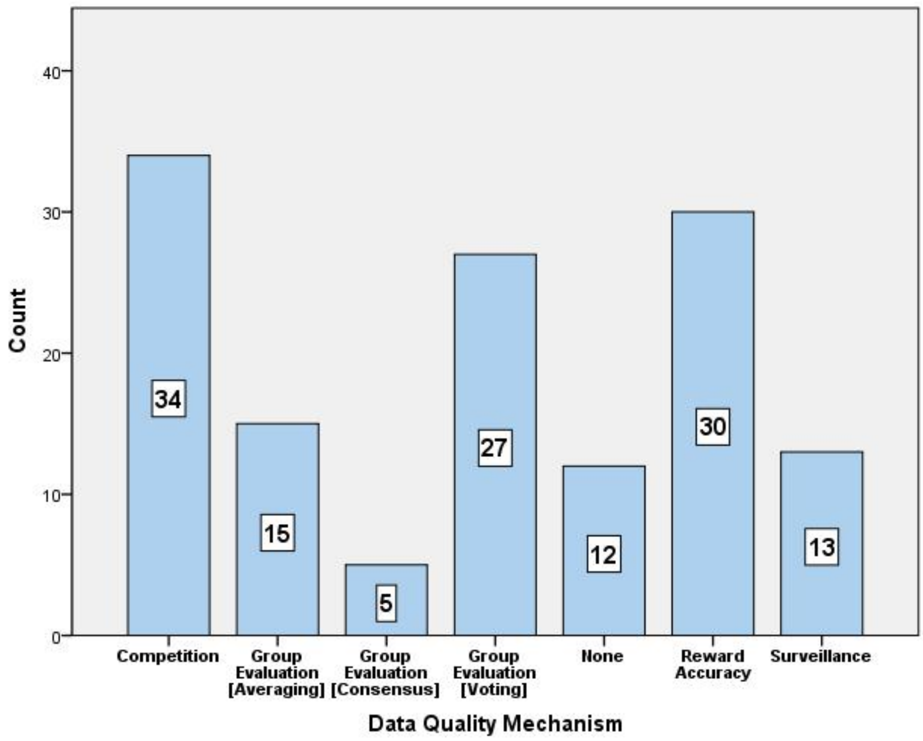

Figure 15: Data Quality Mechanism of the elements in the empirical dataset

***Due to elements belonging to more than a category, the count add up to more than 61**

3.4 Prescriptive Framework and empirical dataset analysis

3.4.1 Introduction

In the previous section we outlined our novel *descriptive framework* in a detailed way. As we already stated, the main goal of the *descriptive framework* is to provide a coherent and systematic way to model *crowdsourcing* applications. *Crowdsourcing*, especially when combined with IT technologies, is an innovative domain and we found out a substantial lack of a correct and exhaustive method to analyze and characterized all its possible facets. We think that the *descriptive framework* fulfills this lack. In this section we will apply the *descriptive framework* to a large set of *crowdsourcing* systems, the ones in our empirical dataset, and we will draw some conclusions and analyses. Therefore, this section will form the *prescriptive framework* as we outlined in the introduction to this chapter.

The *prescriptive framework* is a comprehensive example of usage of our *descriptive framework.* In this way it also assumes the form of a test case for our modeling techniques. However, the *prescriptive framework* is not just a test case but also a fundamental part of our research effort in the field of *crowdsourcing.* Indeed, in this section we will perform an analysis of the *crowdsourcing* phenomenon and try to offer a complete insight of the domain, shaping it into considerations over the elements in the dataset. The research is limited to the elements in the dataset but as, we already stated, we tried to choose an as much as possible representative set of all the various facets that *crowdsourcing* can assume. Moreover, we think that the *prescriptive framework* can also assume the form of guidelines aiding in the design phase of new *crowdsourcing* applications. We will analyze the links between the dimensions in the *descriptive framework* looking to how the values of each dimension are mapped in the values of the other dimensions, using the dataset as source of information. In this way we will provide a characterization of the most common practices implemented by real and successful *crowdsourcing* systems on the market. This set of information could be a precious aid for setting up a coherent methodology when developing new applications exploiting *crowdsourcing* and in particular, for developing the experimental part of our research (see next chapter).

While presenting the *descriptive framework* we already partly applied it to the elements in the empirical dataset. In fact, each dimension description was followed by the dataset distribution chart according to the dimension itself. We don't repeat this passage here. Instead, now we move our focus on the connections between dimensions. Thus, we identified a set of relationships amid the dimensions and we analyzed it by presenting charts followed by our considerations. These relationships can be either binary, thus involving just two dimensions, or ternary, thus involving three dimensions at the same time. The following paragraphs will cover this topic.

The *prescriptive framework* is not just an end in itself, but a starting point for the next step of our research methodology. Indeed the modeling tools (the frameworks) were meant as the foundations from which studying experimentally another important factor: the user satisfaction. Our final aim was establishing which are the best practices to implement a *crowdsourcing* methodology to populate and keep updated the knowledge domain of a sentiment analysis engine, maximizing the user satisfaction variable (see Introduction for more details). Thanks to the analysis performed while developing the *prescriptive framework,* we have been able to identify and design the four main different *crowdsourcing* scenarios involving the sentiment analysis engine. These four scenarios were then experimentally tested and the results form the core of our efforts. The next chapter will discuss these results in details.

3.4.2 Relationships among dimensions

While analyzing the connections among the ten dimensions in the *descriptive framework,* we decided to split the set of all the dimensions in two subsets. We named them: fixed dimensions set and variable dimensions set.

The values of the dimensions in the fixed set cannot vary just according to the wishes of the *crowdsourcing* system designers but are deeply linked to external factors. For example, Task Type and Required Knowledge are connected to the type of job that the users in the community will have to perform. The type of job is chosen before building the community. Actually, we can say that the *crowdsourcing* community is built for performing a specific job (or set of jobs) and around this (these) job (jobs). The designers can propose different mechanisms or interfaces to aid the users while performing these tasks but they can't change their intrinsic nature and all the consequences that it brings in term of required knowledge and complexity. Moreover, consider Community Size as another example: the developers of a *crowdsourcing* application can't actually control this dimension. They might work on the factors that at the end affect the size of community, but they simply can't decide which size their community will have to assume (actually, it would be better to say that they can't force a community to grow to any desired size but they can limit the grow to a particular value). Thus, Community Size is merely a descriptive dimension of the *descriptive framework* and its value is constrained by reality and not by the designers' choices. Categorization, User Type and Crowdsourcing Type are part of the fixed dimensions set as well because they are connected to the nature of the crowd-sourced task itself as seen previously for Task Type and Required Knowledge. For instance, consider the case in which we want to build a community to perform online translations, we will use a Cloud Labor system and as consequence the Categorization dimension will have to reflex this choice.
On the other hand, Reward (Main and Minor), Incentive and Data Quality Mechanism form the variable dimensions set. Remuneration is part of the variable dimensions set as well but it is a different case because it is linked just to the monetary reward and thus we will not perform any analysis on this dimension. These dimensions are different from the ones in the fixed set because the designers while implementing a *crowdsourcing* application, can choose among different values for this group of dimensions and these choices lead to several different ways of exploiting the *crowdsourcing* to accomplish a collection of tasks. The range of options from which the developers can pick the values is not constrained just by the

dimension metrics. Indeed, even for the variable dimensions set, the external factors can limit the range of possible options. Anyway, a certain amount of freedom is always given to the developers and the choice of the correct values for these dimensions is vital to ensure the success of a *crowdsourcing* application. We can see them as *critical success factors* (CSF) necessary for a *crowdsourcing* project to achieve its mission (Rockart, 1979). Previously, we stated that Community Size cannot be set by the developers but anyway it is affected by the designers' choices. As proof of this we can see that the variable dimensions deeply affect how much a *crowdsourcing* project will boost user participation and involvement: correctly rewarding and incentivizing users increase the chances that they will engage with the community and leveraging on an effective data quality mechanism ensure a high quality output from the *crowdsourcing.* These examples should clarify why we consider the variable dimensions so crucial for a *crowdsourcing* system. In the wake of these considerations, the *prescriptive framework* will be focused on the variable dimensions. We will try to answer to the question: "Given some values for the dimensions in the fixed set, how to best choose the values for the variable dimensions?". Part of the answer is just "according to the situations". Indeed, as we already stated, external factors may limit our choices for the variable dimensions. For instance, if we don't leverage monetary rewards we can't use the monetary incentive. But in all the situations in which we have different options we will try to identify which are the best practices studying the elements in our empirical dataset thus looking to the consolidated experience of real existing systems on the market.

The following table summarizes which dimensions are in the fixed subset and which ones are in the variable subset. It must me pointed out that for the Community Size dimension only the qualitative metric has been taken into account in order to make the analysis easier and more understandable. We already stated that this was one of the reasons why we introduced the qualitative metrics.

Table 15: Dimension subsets and the belonging elements
(*) not considered here

Dimension Subset	Dimensions in the subset
Fixed Dimensions Set	▪ Categorization ▪ Crowdsourcing Type ▪ Required Knowledge ▪ Community Size (Qualitative) ▪ User Type ▪ Task Type
Variable Dimensions Set	▪ Main Reward ▪ Minor Reward ▪ Remuneration (*) ▪ Incentive ▪ Data Quality Mechanism

We already pointed out in the introduction to this section that we considered two types of relationships: binary and ternary relationships. Moreover, in this paragraph we underlined our choice of focusing on the dimensions coming from the variable set. Thus, any relationship that we analyzed has an element coming from the variable dimensions set. The other element (or elements in ternary relationships) can either come from the fixed set or belong to the variable set as well. In particular, every dimension in the variable set has been analyzed according to some dimensions in the fixed set, but we also studied relationships over dimensions all belonging to the variable set and some other mixed cases. Even with these constraints the number of possible relationships among the dimensions in the *descriptive framework* was still huge. A complete analysis of all of them was unneeded and outside the scope of our research. Thus, we focused only on the relationships we believed worthy to present according to the quality of the considerations that was possible to carry on them. In particular, we elicited all the analyses bringing to obvious conclusions such as, for instance, the Remuneration-Task Type relationship: it comes straightforward that an increasing complexity of the task corresponds to a proportional increasing remuneration. The usefulness of each analysis will be clarified with the analysis itself.

We have to spend few more words on the Categorization dimension. Indeed, it falls outside the previous considerations because it belongs to the fixed set. Nonetheless, we analyzed it in relation with the Reward and Incentive dimensions. The reason is that we think this to be a fundamental passage in order to provide a complete characterization of the *crowdsourcing* phenomenon. We will spend few more words on this in the following Categorization paragraphs.

The next table provides a detailed summary of all the relationships that have been taken into account for shaping the *prescriptive framework*. It must be read in the following way: the first column specifies the main variable; the second column specifies the second element of the relationship; if in a row it is present even an element in the third column it means that the relationship is ternary.

Table 16: The relationships analyzed in the *prescriptive framework*

Dimension	Binary Relationship	Ternary Relationship
Main Reward	Required Knowledge User Type Task Type	
Minor Reward	Main Reward	
Incentive	Main Reward Required Knowledge Community Size (Qualitative) User Type	Minor Reward
Data Quality Mechanism	Community Size (Qualitative) Task Type	
Categorization	Main Reward Incentive	Minor Reward

3.4.3 Main Reward

<u>*Main Reward – Required Knowledge*</u>

Figure 16 shows a chart depicting the relationship between the Main Reward dimension and the Required Knowledge dimension. The values on the chart bars represent each time the number of elements in the empirical dataset ("count"). First, it's possible to notice that no *crowdsourcing* system in the dataset has a main reward of type Prestige-Oriented and, as consequence, only Enjoyment-based Motivation and Opportunistic appear as possible values for the Main Reward dimension. From this first observation we can argue that a reward of type Prestige-Oriented doesn't usually suffice in encouraging user participation and involvement and thereby the projects in the dataset usually prefer to exploit other types of reward at first stage.
Main rewards of type Enjoyment-based are mainly used when the required knowledge is low whereas medium or high entry barriers (see previous section) usually bring to opportunistic rewards. We can deduce that enjoyment-based rewards don't suffice in motivating users to join *crowdsourcing* communities that require an elevated effort and thus these communities favor extrinsic motivations perhaps coupled with enjoyment-based minor rewards.

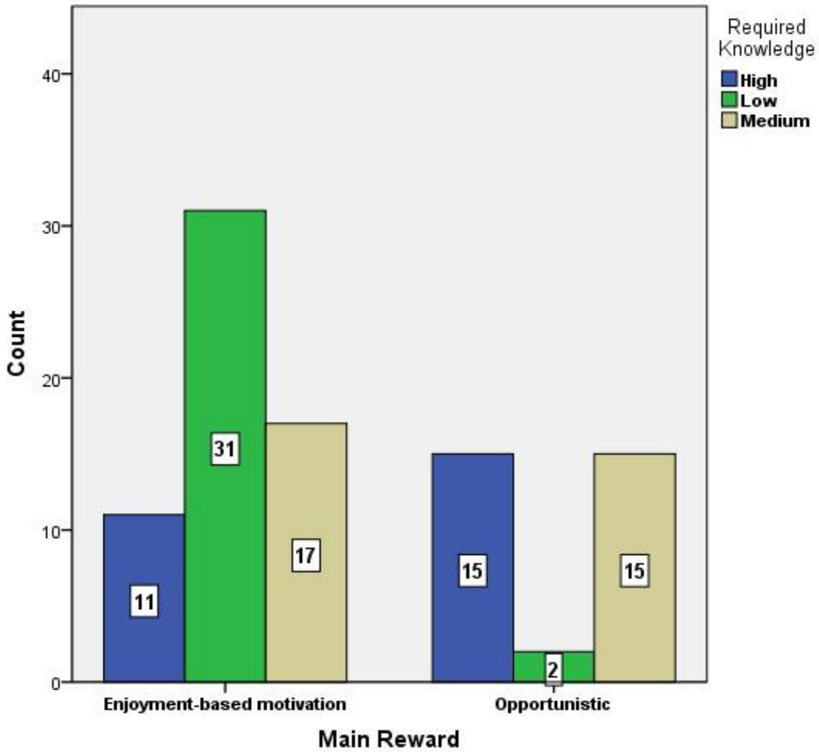

Figure 16: Main Reward – Required Knowledge relationship chart
***Due to elements belonging to more than a category, the count add up to more than 61**

Main Reward – User Type

Figure 17 shows a chart depicting the relationship between the Main Reward dimension and the User Type dimension. The trend is similar to the one seen for the previous relationship between Main Reward and Required Knowledge. Indeed, the *crowdsourcing* systems that rely more on main rewards of type Enjoyment-based have proportionally more chances of having amateur users in their communities. On the other hand, services that rely on professional users seem to prefer opportunistic rewards. Moreover, we can deduce that the opportunistic rewards stimulate more the professional users participation than the amateur one. We already stated while presenting the *descriptive framework* that usually all the *crowdsourcing* communities are open to amateur users. Nonetheless, before designing a new application, developer should identify which is their target user type. Professional users are stimulated to participate by extrinsic rewards. Instead, not only enjoyment-based motivations suffice in stimulating amateur users engagement but also amateur users seem much less affected by opportunistic rewards than professionals.

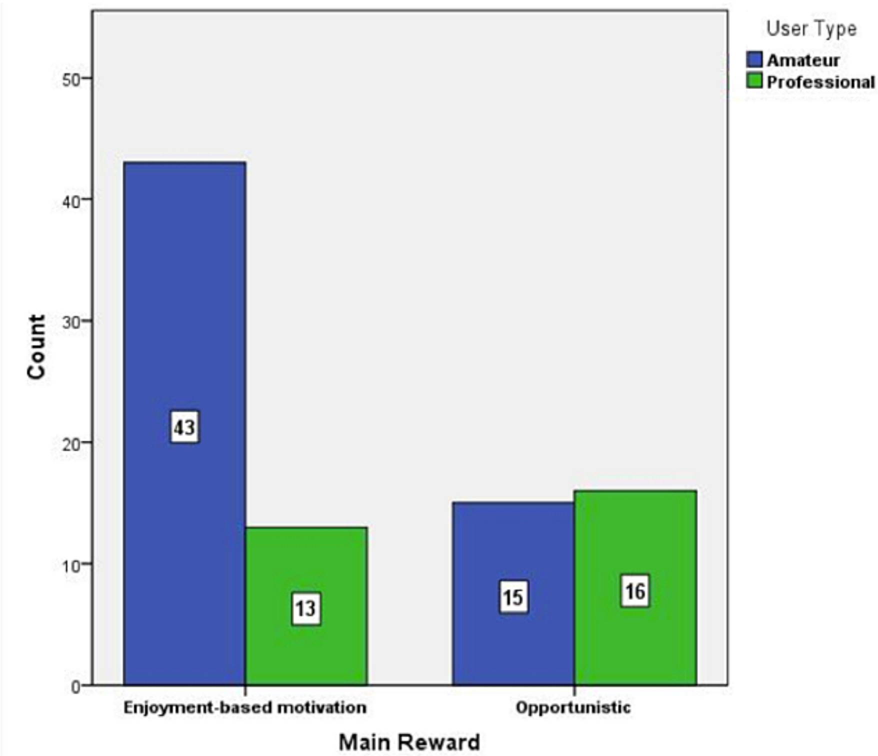

Figure 17: Main Reward – User Type relationship chart

***Due to elements belonging to more than a category, the count add up to more than 61**

Main Reward – Task Type

Figure 18 shows a chart depicting the relationship between the Main Reward dimension and the Task Type dimension. As we could aspect simple tasks are usually coupled to enjoyment-based rewards while complex tasks usually involve more the opportunistic side. Anyway, it's possible to notice that the number of *crowdsourcing* services with complex tasks and that rely on opportunistic rewards is just slightly greater than the ones relying on enjoyment-based rewards. Thus, users are willing to perform even complex tasks if motivated with enjoyment-based rewards and opportunistic motivation is not mandatory in every case. Moreover, we can observe that 4 *crowdsourcing* applications belonging to our empirical dataset, exploit game tasks and enjoyment-based rewards, while no service relies on game tasks and opportunistic rewards. At first glance this could appear straightforward but it should be observed that at the moment of writing more *crowdsourcing* services exploiting game and extrinsic-monetary motivations (for instance crowd-sourced social betting platforms) are reaching the market.

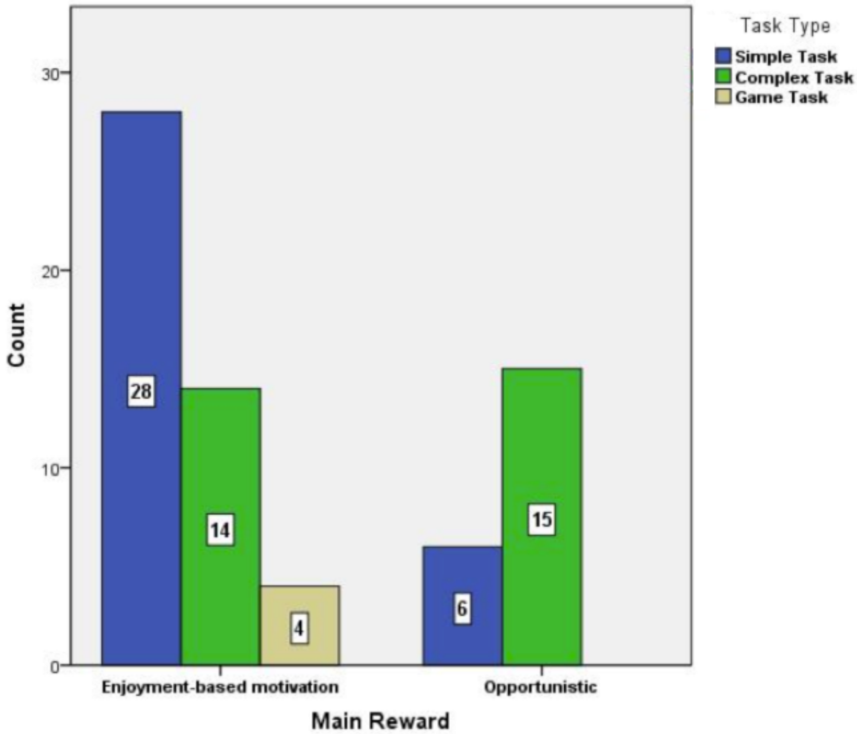

Figure 18: Main Reward – Task Type relationship chart

***Due to elements belonging to more than a category, the count add up to more than 61**

3.4.4 Minor Reward

Minor Reward – Main Reward

Figure 19 shows a chart depicting the relationship between the Minor Reward dimension and the Main Reward dimension. The analysis of this relationship is interesting because it shows how usually the *crowdsourcing* applications couple the

Minor and Main reward dimension. First, it’s possible to notice that unlike what was for the Main Reward, several *crowdsourcing* systems rely on Prestige-Oriented motivations as Minor Reward. Actually, it looks that Prestige-Oriented is the foremost minor reward. Secondly, several elements in the dataset don’t rely on any Minor Reward. These systems usually seem to choose enjoyment-based main rewards. The reason why they lack minor rewards is mostly connected to their own intrinsic nature: often they can’t exploit monetary or other extrinsic rewards because it’s not part of their mission, neither they can’t rely on rewards of type prestige-oriented because this is not supported by the nature of the crowd-sourced task or by the platform.

Finally, it looks that the rewards of type opportunistic are hardly used as Minor Reward. We can argue that the reason is that the opportunistic rewards easily become the Main Reward when used.

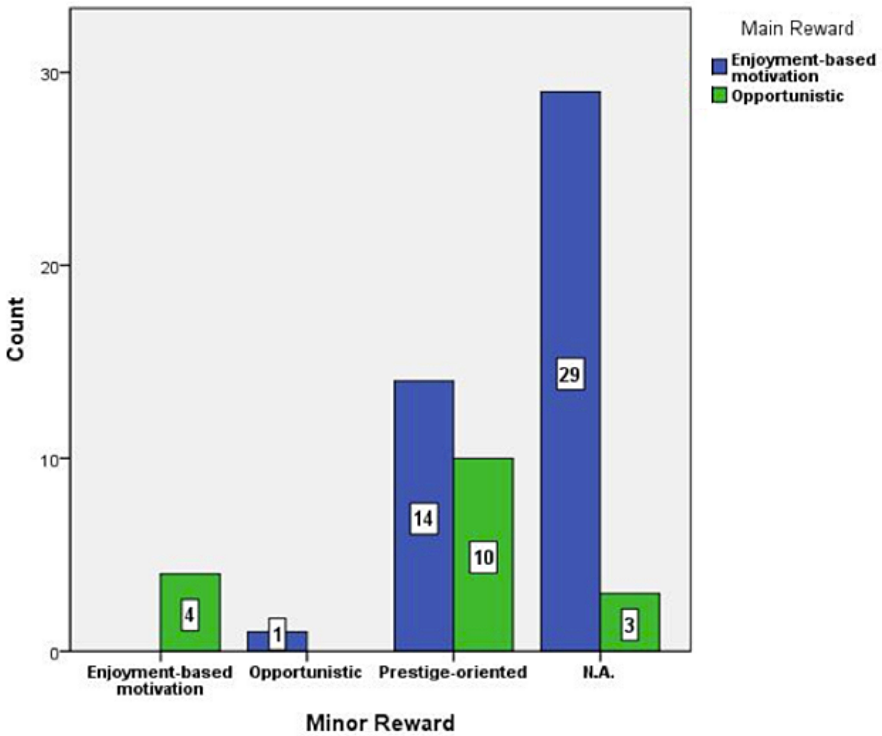

Figure 19: Minor Reward – Main Reward relationship chart

***Due to elements belonging to more than a category, the count add up to more than 61**

3.4.5 Incentive

<u>*Incentive – Main Reward – Minor Reward*</u>

Figure 20 shows a chart depicting the ternary relationship between the Incentive, Main Reward and Minor Reward dimensions. While presenting the *descriptive framework* we repeatedly stated that the Incentive dimension is linked to the Reward dimensions (both). In particular we showed a table summarizing this fact (see the *descriptive framework* for more details). In this paragraph we will support our statement through the analysis of the elements in the empirical dataset and we will show how we experimentally built the links between the incentives and the rewards. Conversely, in this paragraph we will not show any correlations between the Main and Minor Reward dimensions because we already discussed them previously in this section. We must underline the fact that it's possible for projects in our empirical dataset to exploit more incentives at the same time, but they can have just one Main Reward at the time. They might also have one Minor Reward, which must be different from the main one, but they are not forced to have it.

If we look to the Money incentive we can see that it is linked to the Opportunistic Main Reward. This is perfectly coherent from what we presented in the *descriptive framework.*

Instead, if we move to the User ranking and voting systems incentive, we can notice that the Main Reward is of type Enjoyment-based motivation while the Minor Reward is either Prestige-oriented or not available. This is perfectly summarized by Table 13 as well.

The incentive Competition (and Gaming) has almost an equal count of elements with an Enjoyment-based Main Reward and an Opportunistic one. The Enjoyment-based count is slightly greater and therefore we put this incentive in relation with the Enjoyment-based motivations in our *descriptive framework*. Moreover, we concluded that the opportunistic component comes from the fact that competition can aid the emerging of the best users in online communities. The greater visibility of the best users is an important factor for many communities, especially for the ones with professional users, because the people in the *crowd* can use it to improve their reputation and get better compensation and jobs, in other words as a form of indirect monetary compensation. We will see several examples of this behavior throughout all the analysis of the dataset and it will become clearer.

Position inside community and user power scaling are referred in the *descriptive framework,* as linked to a Prestige-Oriented reward mechanism. This doesn't emerge from the chart but there is an explanation of this trend. Indeed, we can see that the link emerges when looking to the Incentive – Minor Reward relationship and we already said that main rewards of type Prestige-Oriented never appear in our empirical dataset concluding that this type of reward alone doesn't suffice in stimulating users involvement. Thus, we need to move our focus to the Minor Reward dimension to find the link and this is what it exactly appears to happen.

Finally, Sharing of the result and Sharing of the goal have always a dominant connection with Enjoyment-based main rewards as we would aspect.

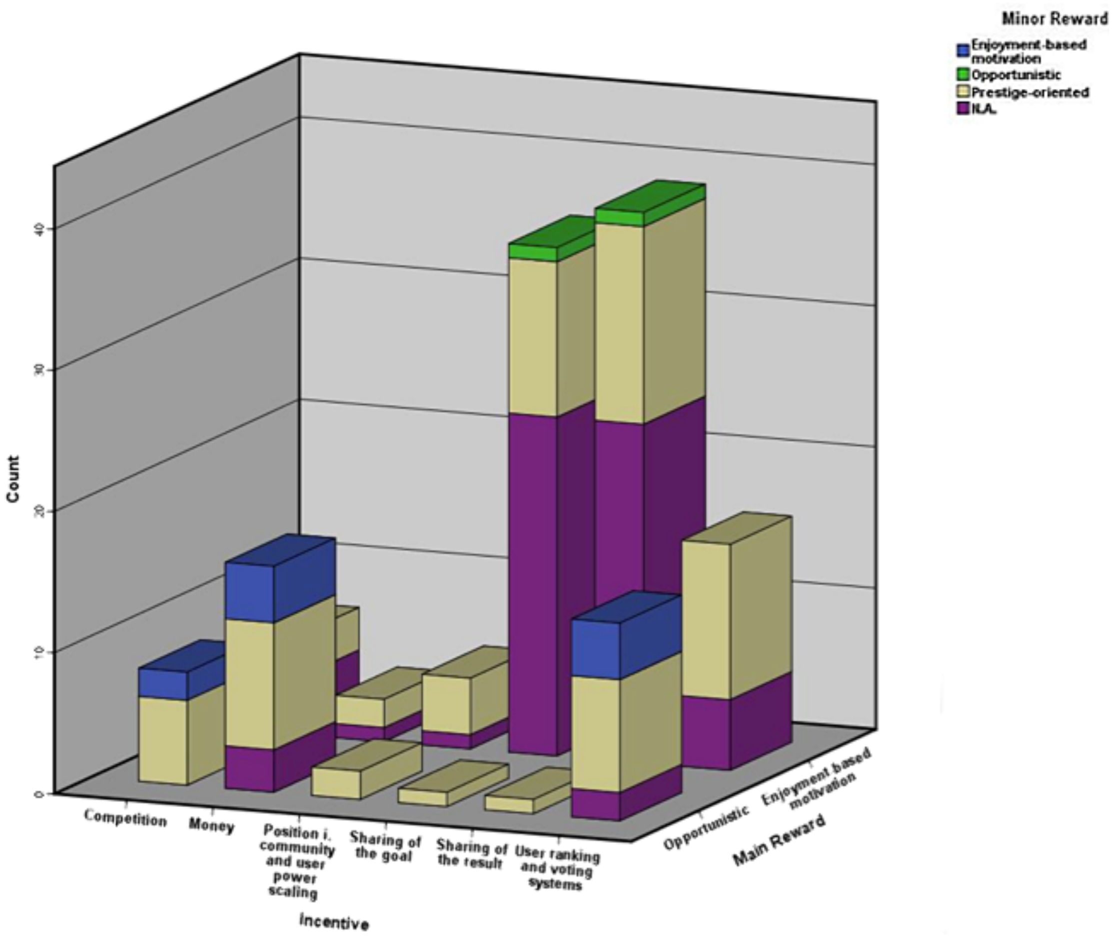

Figure 20: Incentive – Main Reward – Minor Reward ternary relationship chart

***Due to elements belonging to more than a category, the count add up to more than 61**

Incentive – Required Knowledge

Figure 21 shows a chart depicting the relationship between the Incentive dimension and the Required Knowledge dimension. First, it must be observed that incentives are mainly linked to the reward mechanisms adopted by the *crowdsourcing* applications. We already pointed out this fact previously in the *descriptive framework* and while describing the Incentive – Main Reward – Minor Reward ternary relationship. However, it's still desirable to analyze how the other dimensions interact with the Incentive dimension looking for possible influences and connections.

At first sight, it's possible to notice that *crowdsourcing* systems with low entry barriers and required knowledge (see the *descriptive framework* for definitions) usually exploit more the incentives Sharing of the goal and Sharing of the result. We can argue that the main reason is that these two incentives are relatively the easiest and cheapest ones to deploy by an online community and thereby, when possible, they are chosen as first solution. Sharing the goal theoretically doesn't require any investment from the *crowdsourcing* applications developers if we exclude the actions, not mandatory, focused on brand and image developing. Indeed, it requires the users in the community to share the reasons and the aims of the tasks they are invited to accomplish. Sharing the result requires the *crowdsourcing* applications to set up an infrastructure allowing the user-base to enjoy the results produced by mean of *crowdsourcing*. For online communities using Internet technologies this can be achieved at relatively low expenses. Thus, Sharing of the results and Sharing of the goal are two incentives that are vastly used for boosting user participation in *crowdsourcing* projects with low entry barriers and required knowledge, and in which low-cost and easy incentives suffice. It is not always possible to exploit these two incentives. Many services (for instance Cloud Labor), don't allow the user to share the results produced by mean of *crowdsourcing* because of legal reasons and often *crowdsourcing* platforms don't even hold the intellectual rights on the *crowdsourcing* output because they work as intermediate agents between the demand of *crowdsourcing* and the offer. Moreover, it is not always possible to incentive users engagement by mean of Sharing of the goal. Usually, only not-for-profit online communities can hope to exploit this incentive. Finally, it must be observed that when the required knowledge gets higher, these two incentives may reveal not effective enough to stimulate user participation. This is the reason why Money, Competition and User Ranking and Voting System seem to be preferred in *crowdsourcing* projects that have medium or high entry barriers. Moreover, the trend looks similar both for a medium required knowledge and for a high one. Analyzing the chart it also appears that money is seldom exploited as incentive for tasks with a low required knowledge. As final conclusion it must be observed that one of the incentives, namely Position inside community and user power scaling, is rarely used by all the systems in the empirical dataset independently from the required knowledge. The reason is that it is not always possible to build online *crowdsourcing* communities in which users have different roles and powers either because there is no need of such powers and roles or because the designers don't want to divide the users in the community.

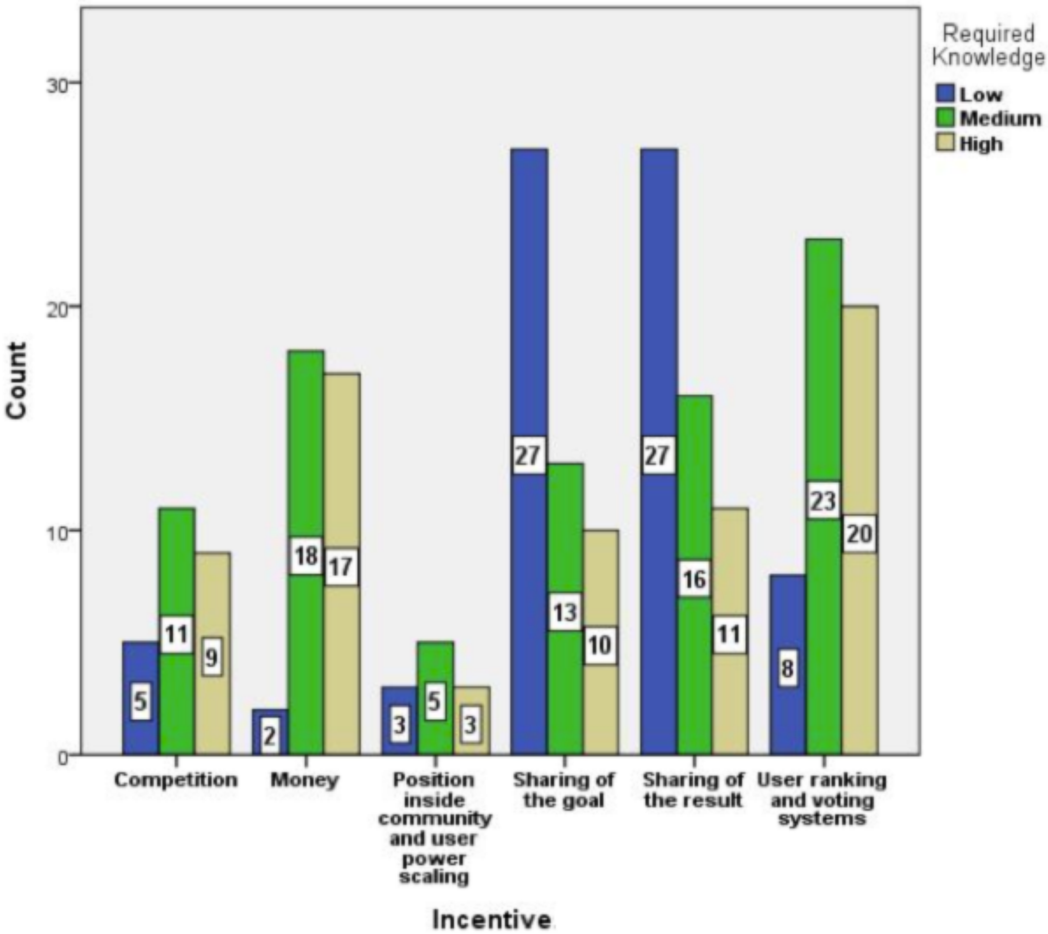

Figure 21: Incentive – Required Knowledge relationship chart
***Due to elements belonging to more than a category, the count add up to more than 61**

Incentive – User Type

Figure 22 shows a chart depicting the relationship between the Incentive dimension and the User Type dimension. The analysis of this relationship is interesting because it can shed light on which are the most effective incentives for the amateur and professional users. In particular our attention will be focused on the professional users because, generally, it is harder to get them involved in online communities with respect to the amateurs ones.

As we could aspect the main incentive for professional users appears to be the financial incentive, namely the money. Indeed while analyzing the Main Reward – User Type relationship, we found out that opportunistic rewards were important to get the involvement of professional users in the *crowdsourcing* community and thus it comes straightforward to support opportunistic rewards with the monetary incentive. When it comes to amateur users the main incentives are Sharing of the goal and Sharing of the result. This trend is similar to the one we have seen for the previous relationship between the Incentive and the Task Type dimensions. The similarity can be explained considering that communities with a low required knowledge usually attract amateur users while the ones with high or medium entry barriers attracted more professional users so the trends for the two relationships should follow the same path. Moreover, Sharing of the goal and Sharing of the result are linked to enjoyment-based rewards and therefore work better with amateurs.

It’s interesting to notice that the Competition and User ranking and voting systems are widely used as incentives even in *crowdsourcing* communities with a professional user-base. The reason can be traced in the fact that user rankings and voting systems are linked to Prestige-Oriented reward mechanisms. Professional users usually pay attention to their reputation because it is a fundamental value in order to get more earnings and jobs. Thus, they would actively contribute to *crowdsourcing* projects in order to have a showcase of their abilities and skills. The *crowdsourcing* applications can support this behavior by setting up users ranking, voting systems and contests that reward the most talented users. Finally User ranking and voting systems and Competition seem to encourage also amateurs participation. In this case we argue that the reason should be found in the enjoyment side of these systems.

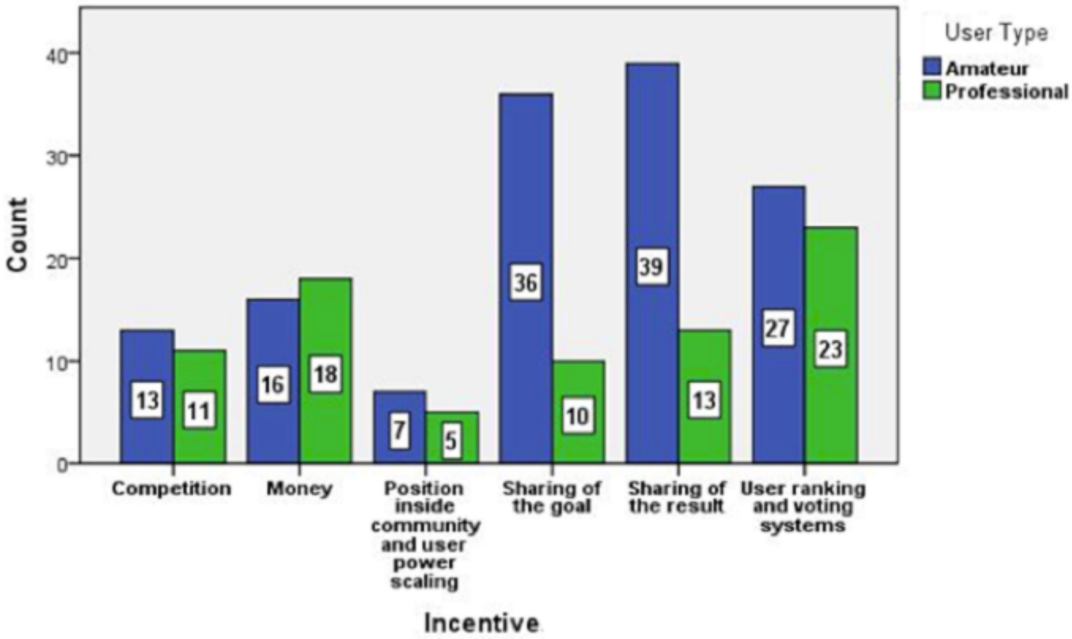

Figure 22: Incentive – User Type relationship chart

***Due to elements belonging to more than a category, the count add up to more than 61**

3.4.6 Data Quality Mechanism

Data Quality Mechanism – Community Size

Figure 23 shows a chart depicting the relationship between the Data Quality Mechanism dimension and the Community Size dimension. This relationship is interesting because, as we already stated while describing the Community Size dimension in the *descriptive framework*, the size of the community is fundamental in order to effectively implement systems that would ensure a high quality of the *crowdsourcing* output.

We start looking to the Surveillance mechanism, i.e. a mechanism in which there is a manual control of the output quality. This type of data quality mechanism can be used only by small *crowdsourcing* communities. Indeed, in bigger communities where the amount of exchanged data is huge, this type of mechanism would require a big effort and a great quantity of resources that only few projects can provide. This

emerges from the relationship chart where we can see that the communities exploiting Surveillance are rather small.
Bigger communities, generally, prefer to rely on two of the Group Evaluation mechanisms (Averaging and Voting), on Competition and on Reward Accuracy. The same mechanisms are exploited even by smaller communities as proof of the fact that they work well independently from the community size.
Competition and Reward Accuracy have a similar trend. This is a consequence of the fact that these two mechanisms are often coupled in several *crowdsourcing* projects and they are implemented together (see *descriptive framework*).
Group Evaluation – Consensus mechanism doesn't seem to be really used by many elements in the empirical dataset. We can argue that the reason is that it requires a *crowdsourcing* system based specifically on collaborative tasks in which multiple people from the *crowd* can work on the same job at the same time and make contributions like, for instance, what happens for Wikipedia articles.
It's interesting to notice that all the communities but two, of which we could retrieve information about the community size, use one or more Data Quality Mechanisms. This is another evidence of the importance of this dimension in the *descriptive framework.*
Finally, as we could aspect, the trend of medium sized communities is something in between the other two cases. Medium size communities seem to exploit all the Data Quality Mechanisms included in our *descriptive framework* without expressing any clear preference for any of them. This is coherent with the way we choice to shape the qualitative metrics for the Community Size dimension.

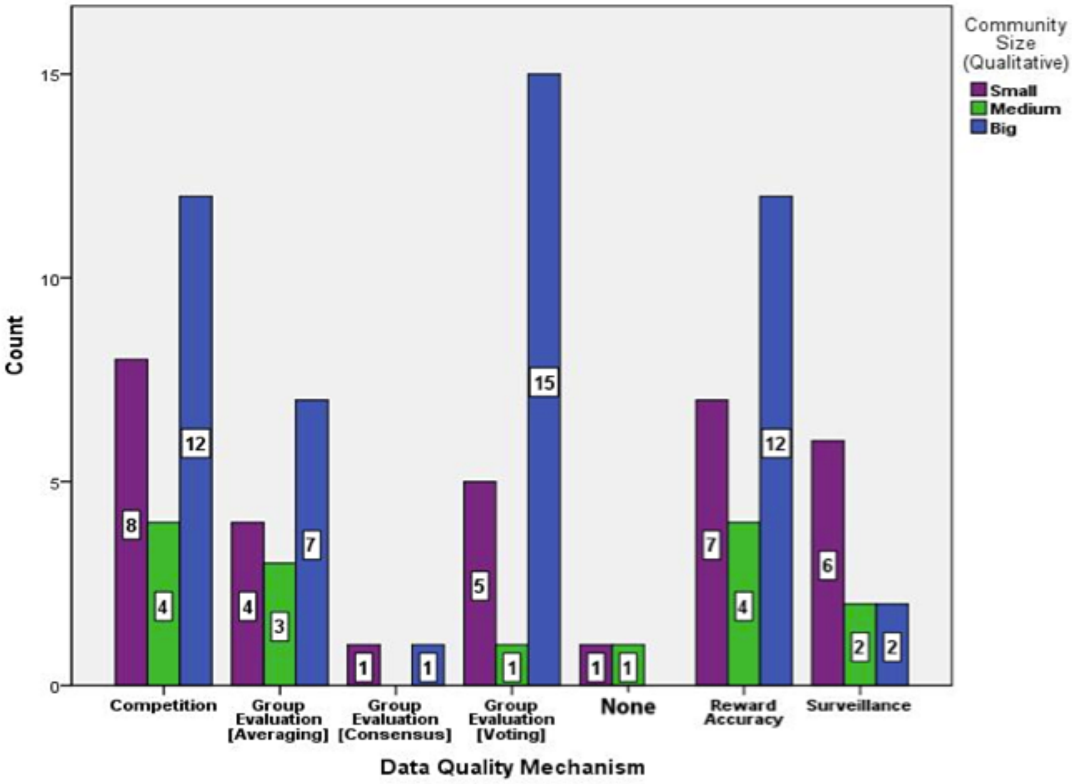

Figure 23: Data Quality Mechanism - Community Size relationship chart
***Due to elements belonging to more than a category, the count add up to more than 61**

<u>*Data Quality Mechanism – Task Type*</u>

Figure 24 shows a chart depicting the relationship between the Data Quality Mechanism dimension and the Task Type dimension. At first glance, it’s possible to notice that complex tasks require different data quality mechanism with respect to simple tasks. This is coherent with the fact that complex and simple tasks need, generally, a different amount of effort to be accomplished.

Reward Accuracy and Competition seem to be the preferred mechanisms by communities with complex tasks while Group Evaluation – Averaging and Group Evaluation – Voting are preferred by communities based on simple tasks. Moreover, Surveillance is used more when the tasks are harder.

It’s interesting the case of game tasks. The systems that exploit gaming as base for *crowdsourcing* use, as we could aspect, Competition as Data Quality Mechanism.

But they also exploit in the same percentage the Group Evaluation – Consensus and Group Evaluation – Voting solutions. Indeed, these games are usually designed to produce a crowd-sourced evolutionary solution to a problem, for instance genome mapping (see the Appendix A), that the gamers contribute to generate while playing. Moreover, some games allow the players to vote the other players' performances and achievements in order to produce better crowd-sourced results. These observations are well showed by the chart.

Finally we must observe that the *crowdsourcing* systems in our empirical dataset can ask to the *crowd* to perform different kind of tasks with different levels of complexity. Thus, they usually rely on more than one data quality mechanism at the same time.

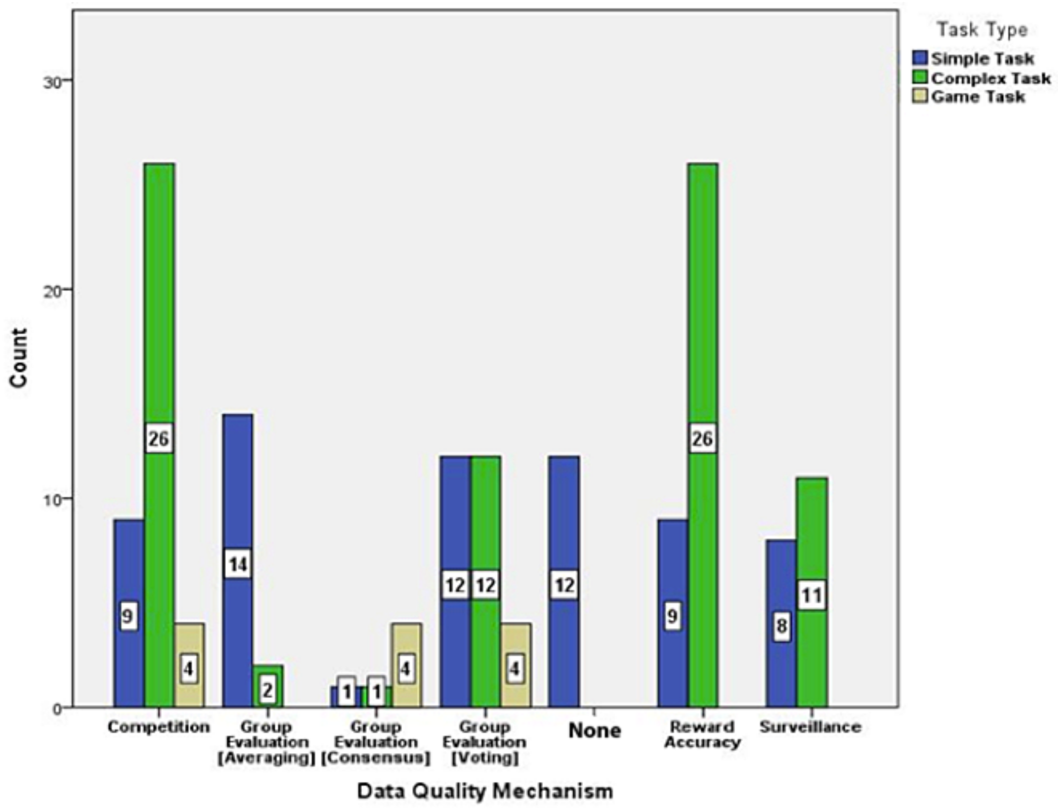

Figure 24: Data Quality Mechanism – Task Type relationship chart

***Due to elements belonging to more than a category, the count add up to more than 61**

3.4.7 Categorization

<u>*Categorization – Incentive*</u>

We will discuss the Categorization dimension in binary relation with the Incentive Dimension and in ternary relation with Main and Minor Reward dimensions. As we stated previously Categorization is not a dimension belonging to the variable dimensions set as we defined it in the introduction to the *prescriptive framework,* but we decided to analyze it anyway because we thought this could aid drawing a better description of the *crowdsourcing* phenomenon.

Figure 25 shows a chart depicting the relationship between the Categorization dimension and the Incentive dimension. It appears that the systems falling in the Collective Knowledge and Knowledge Sharing categories exploit the Sharing the goal and Sharing the result incentives. The reason should be clear: these systems usually have in their mission statements, the goal of sharing the results they produce throughout *crowdsourcing* and therefore they obviously exploit the Sharing the result incentive. Moreover, people see a mission of knowledge sharing as something valuable by itself. Therefore they usually share the goal and the user involvement results encouraged by fact.

Monetary incentive is instead the foremost incentive of Cloud Labor and Problem Solving *crowdsourcing* systems. We can argue that the main reward of these systems is opportunistic and in particular is the direct monetary compensation (see next paragraph). Thus, the incentive reflexes this fact. Moreover, usually the tasks proposed by these systems are either more complex than the ones of projects in other categories, or they don't have an enjoyment side. Therefore, only a substantial extrinsic reward can move people to accomplish them. User ranking and voting systems are very important in these systems as well. We discussed in part the reasons of this tend previously while presenting the Incentive – User Type relationship. We stated that professional users are usually stimulated by systems that allow them to increase they reputation online and we see here another proof of this conclusion. Indeed Cloud Labor and Problem Solving systems usually have a significant percentage of their user-base composed by professional and therefore they exploit User ranking and voting mechanisms for this reason.

Collective Creativity and Knowledge Acquisition applications don't seem to focus on any particular incentive but they tend to use all of them with a little preference for User ranking and voting systems.

Open innovation and Crowdfunding have a different behavior with respect to the other case presented until here. The only incentive used by the Open Innovation systems in our empirical dataset is the monetary one. These *crowdsourcing* projects usually require a great effort from the users involved and thus they have to stimulate user engagement throughout big monetary prizes. Moreover, we have just few examples in our empirical dataset of these systems.

Crowdfunding applications use just two incentives: Sharing the goal and Sharing the result. The main reason of this behavior stays in the particular nature of the tasks performed by the *crowd,* i.e. financing projects and activities. This is a form of integrative *crowdsourcing* and without any kind of competition thus it would have been hard to use other type of incentives

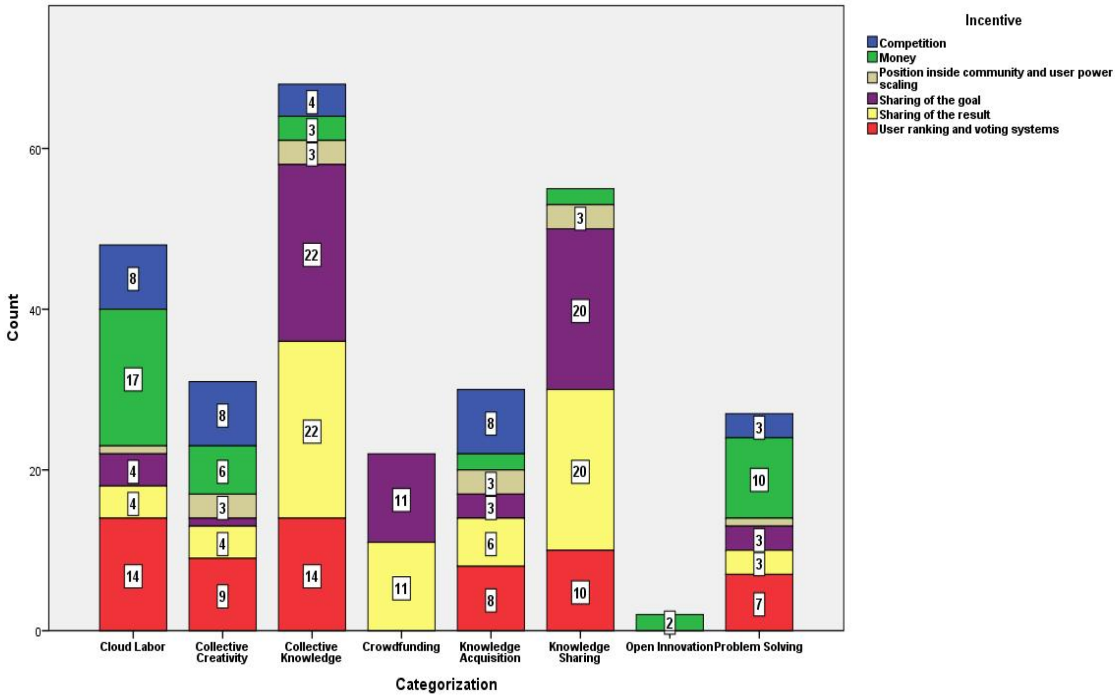

Figure 25: Categorization – Incentive relationship chart

***Due to elements belonging to more than a category, the count add up to more than 61**

<u>*Categorization – Main Reward – Minor Reward*</u>

Figure 26 shows a chart depicting the ternary relationship between the Categorization, Main Reward and Minor Reward Dimensions. The analysis of this relationship is very interesting because it allows us to check how *crowdsourcing* systems belonging to different categories exploit different kind of rewards mechanisms. In this paragraph we will cover just the ternary relationship between the three dimensions and not the link between the Main Reward and the Minor Reward dimensions that has already been discussed previously in this section.

First, we find another proof of our statement that the rewards of type opportunistic are almost never minor rewards. Indeed, we can see just two small green areas in the chart of Figure 25. These areas represent the number of systems in our empirical dataset that exploit extrinsic-opportunistic, namely monetary, rewards for the Minor Reward dimension and they are rather small compared with the rest of the chart. Therefore we can again conclude that extrinsic rewards are the main and foremost

stimulators of users participation and involvement in online *crowdsourcing* activities (but often they aren't sufficient when taken alone, see *descriptive framework*).
Cloud Labor *crowdsourcing* applications mainly propose opportunist rewards for the Main Rewards dimension and prestige-oriented ones for the Minor Reward dimension. This is the behavior we could aspect, being them a particular kind of labor systems. Problem Solving category presents the same trend of Cloud Labor. Indeed, as we already stated, they are usually coupled and the systems often belongs to both the categories at the same time (see the *descriptive framework*). Collective Creativity systems also present a similar trend but here the preference for opportunistic main rewards is not as marked as in Cloud Labor category. Crowdfunding systems rely mainly on enjoyment-based main rewards and they usually don't have a minor reward. This behavior is the same that we have previously seen while discussing the relationship between crowdfunding systems and incentives. Also the category Knowledge Acquisition has a behavior similar to the one it had with regard to the Incentive dimensions: they rely on enjoyment-based main rewards and prestige-oriented minor reward in the majority of the cases. Open Innovation systems exploit only opportunistic main rewards and prestige-oriented minor rewards for the same reasons why they use only the monetary incentive: the complexity of their tasks require really tempting prizes and thus only a monetary compensation fits here.
Finally, we can observe how clearly it again appears the link between the Incentive dimension and the Reward ones (both) as we outlined it in the *descriptive framework.*

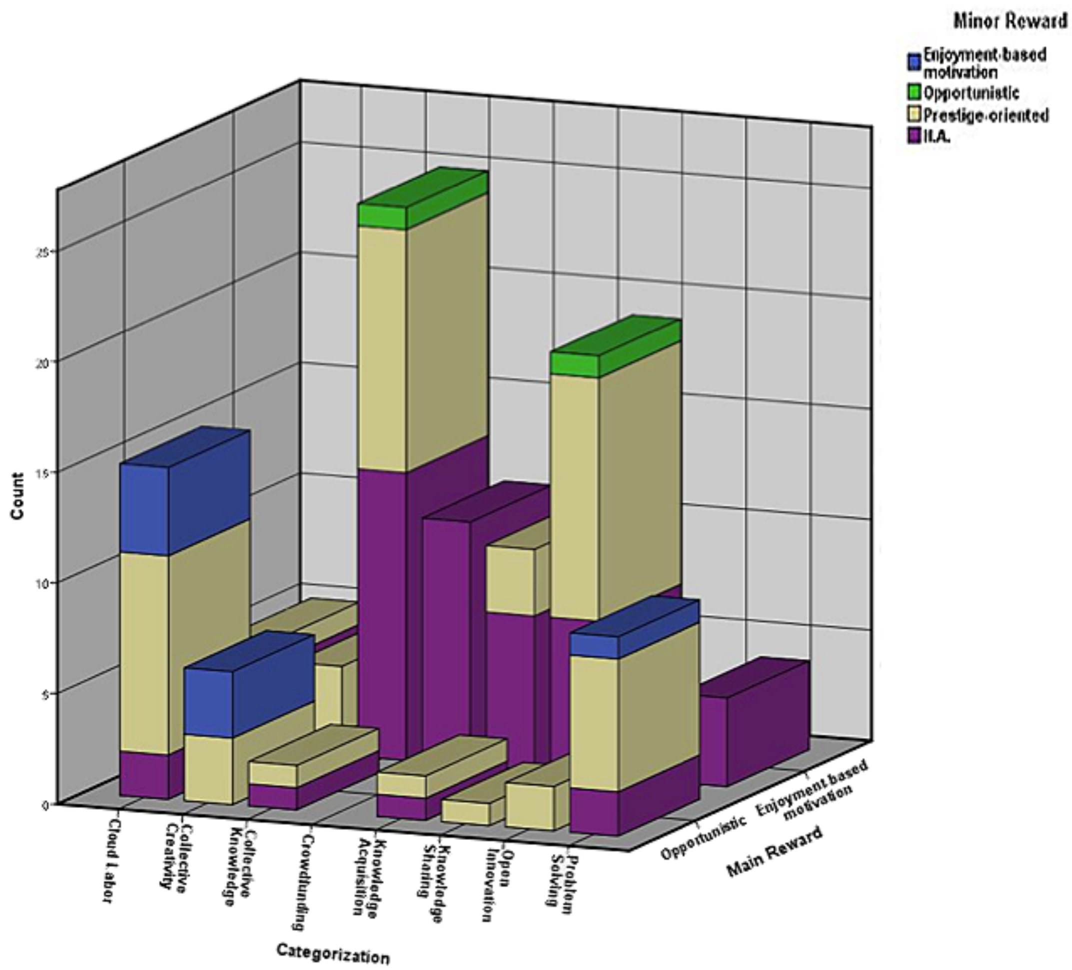

Figure 26: Categorization – Main Reward – Minor Reward ternary relationship chart

***Due to elements belonging to more than a category, the count add up to more than 61**

3.4.8 Final Conclusions

In the introduction to this section we said that the *prescriptive framework* had three main roles in our research: being a test of the modeling techniques and hypotheses developed in the contest of the *descriptive framework,* providing a set of guidelines (or best practices) derived from the analysis of the real *crowdsourcing* applications belonging to our empirical dataset and finally, providing an usage example of the *descriptive framework.* To this we can add now that the *prescriptive framework* is also a survey of the *crowdsourcing* phenomenon. We think that the *prescriptive* *framework* as outlined in the previous paragraphs, fulfills all the goals we set at the beginning of its development.

On that bases, we can conclude that the *descriptive framework* is a valid aid to analyze the *crowdsourcing* domain and its applications. It can cover all the possible facets that *crowdsourcing* assumes thanks to its inclusive list of dimensions and metrics. Moreover, the *framework* helps to understand the design logic behind a *crowdsourcing* application: which are the incentives for the users and why, how these incentives are connected to nature of the tasks to be performed on the platform, how and why the users are encouraged to join the community and finally how all these features are put together in an organic IT system, i.e. a virtual online *crowdsourcing* community.

While describing the Incentive dimension we built a connection between them and the reward mechanisms. The incentives were identified experimentally looking to the elements in the empirical dataset. Afterwards, we had a look to the academic literature in order to find some theoretical grounding supporting our choice (see the dimension description in the *descriptive framework*) and, at this point, we drew the link between incentives and rewards. We must underline that, at this stage, the connection was a hypothesis, although mediated by our experience on the field. The analysis of the ternary relationship among the Incentive – Main Reward – Minor Reward dimensions performed in the contest of the *prescriptive framework*, gave us empirical evidence of the connection we had built. Moreover, it offered another proof of the validity of the *descriptive framework.* On the same wake, although not as relevant, we proved the statement we made about the relation occurring between the Community Size and Data Quality Mechanism dimensions. The analysis of this relationship in the contest of the *prescriptive framework* gave us evidence of the link.

We can conclude that the *descriptive framework* we described in this chapter is the synthesis of theoretical research over the academic literature, and experimental analysis of real *crowdsourcing* applications. The *descriptive framework* coupled with the *prescriptive framework* forms a complete survey of the *crowdsourcing* landscape.

We have to spend some words on the best practices. At the beginning of this paragraph we said that one of the roles of the *prescriptive framework* was indeed to provide a set of guidelines to be useful when developing new *crowdsourcing* applications. Moreover, in the introduction we said that one of our aims was to answer to the question “Given some values for the dimensions in the fixed set, how to best choose the values for the variable dimensions?”. We think that the work done until here can answer to this question. In our analysis of the relationships existing among the dimensions in the *descriptive framework,* we have outlined how real *crowdsourcing* systems are built, showing how to combine the various facets that the dimensions can assume. We showed how the dimensions in the variable set assume their values according to values taken by the dimensions in the fixed set. In this way, we have offered various hints to design better *crowdsourcing* applications capable of effectively exploiting the several techniques exposed here.

We have used the knowledge presented until now to build the next phase of our research that will be covered in detail in the next chapter. As we already stated several times throughout this thesis, our biggest effort was directed at understanding which are the best mechanisms to exploit the *crowdsourcing* to build a factual knowledge base for a sentiment analysis engine (see Introduction and Chapter 2 for further details and definitions). We developed four different scenarios. They correspond to an equal number of online *crowdsourcing* platforms that exploit different declinations of *crowdsourcing* mechanisms. The importance of our *descriptive* and *prescriptive frameworks* relies in this passage: they were the key to understand which are all the possible ways to build a *crowdsourcing* application and to learn which constrains set up to make it more effective. To make clear this concept a paragraph of the next chapter will briefly show how the design choices for the four scenarios are linked to the analysis in the *prescriptive framework*.

Chapter 4

Experimental study and analysis of the results

4.1 Introduction

In this chapter we will describe the experiment we conducted as last part of our research. As we already discussed in the Introduction, we linked the theoretical work to an experimental effort, consisting in the analysis of the empirical dataset of *crowdsourcing* applications in the *prescriptive framework*, and in the experiment presented here. This is the last step of our research methodology as we outlined it in the first chapter.

The experimental study presented here has several goals: first we wanted to build an effective *crowdsourcing* methodology for a sentiment analysis engine capable of maximizing the user satisfaction variable. This is the main problem of our research, i.e. developing a novel methodology for automatically populating the knowledge domain of a semantic analysis tool. Second, we wanted to test our modeling tools (the *descriptive* and *prescriptive frameworks*) in the context of a new *crowdsourcing* application development. In this way, we could test on the field the modeling techniques that we showed in the previous chapter and moreover, apply the knowledge coming from the *prescriptive framework*. In this way we could further assess the validity of our research methodology. Finally, we had another goal that comes as consequence of the other two. While presenting the *descriptive framework*

and, even more the *prescriptive framework*, we spent many words on describing which are the best ways for stimulating user participation in *crowdsourcing* communities. It clearly appears that this is a fundamental topic: the most effective ways of harnessing the *crowdsourcing* involve not just the mere technological side, but also the social side. Actually, we could state that the implementation technologies must address and take into account this issue. Indeed, there is no *crowdsourcing* methodology without a vital and vast community of users as far as it declines in the problem of data quality and data production (see *Wisdom of the crowds,* Chapter 2). Thus, our experimental effort was also much oriented towards this issue. Practically, we linked this issue to the user satisfaction. As we said in Chapter 1, our research hypothesis is that user satisfaction is a good indicator of the quality of a *crowdsourcing* system. Moreover, we derived the conclusion that the level of user participation is directly linked to user satisfaction. Thus, we wanted to understand, from real test users, which are their opinions and feelings on various *crowdsourcing* implementations and methodologies that we set up, in order to understand in which directions we will have to move while building the future real *crowdsourcing* implementation for a specific sentiment analysis tool.
The experiment has been built thanks to the profound studies of the *crowdsourcing* phenomenon. As said in the Introduction, *crowdsourcing* is a novel domain and before starting this research our knowledge of the field didn’t allow us to directly start implementing a *crowdsourcing* methodology for the semantic *sentiment analysis*. Therefore, we came with such a complete approach.
The *descriptive framework* has helped us understand which are all the factors to take into account while developing a new *crowdsourcing* application and has given us a coherent and complete methodology to describe it in a scientific fashion. The *prescriptive framework* told us how to move in order to effectively build the application: which are the best practices, which are the fundamental issue and how real *crowdsourcing* applications on the market face them.
The test cases that we developed are four. Each of them is different in several design choices but they all share the same background architecture. The test scenarios are developed starting from this base architecture of a *crowdsourcing* web application.

The web application involves a community of users that have to accomplish tasks. It is a *crowdsourcing* system according to our operational definition (see Chapter 2). Then, each scenario applies a set of different techniques derived from the modeling frameworks that we developed. In particular the first, second and fourth scenarios use different set of incentives and rewards systems but they are equal according to the other dimensions of analysis in the *descriptive framework*. The third scenario tries also to exploit a different type of task (a game task), thus changing the rewards and incentives system as well. The base architecture and all the other details concerning the scenarios will be widely covered in the chapter.

In the next section we will describe the experiment in details: in what it consists and which is the overall methodology. Then, we will describe each scenario using the *descriptive framework* as guideline and showing also some useful screenshots. In paragraph 4.5 we will show how we used the *prescriptive framework* in this experimental phase. Then, we will present the survey that was used to collect the data from the test users and we will describe the group of test users. Finally, the last part of the chapter is devoted to the analysis of the results. It consists of two paragraphs: in the first we will analyze the specific answers of each question in the survey; in the second, we will draw our general conclusions and conclude the chapter.

4.2 Description of the experiment

In this section we will describe the experiment in details. The experiment consists in the analysis of four different *crowdsourcing* methodologies exposed in four different test cases (*crowdsourcing* scenarios). The scenarios are *crowdsourcing* web application, i.e. web applications that exploit *crowdsourcing.* Thus, according to our operational definition of *crowdsourcing,* they are platforms in which a group of individuals (the *crowd)* perform a collection of tasks according to a model (see Chapter 2).

As we already discussed in the introduction to this chapter, in our experiment we addressed a *crowd* in order to populate the knowledge domain of a sentiment analysis tool. The core of the empirical experiment has been choosing a *crowdsourcing* methodology to effectively allow the people to perform this task and test the outcomes to identify an overall final model that maximizes the user satisfaction, i.e. our quality metric.

Part of the problem was designing a *crowdsourcing* system and architecture in all its features: the design of the community, of the rewards system, etc. We also had to design a visual interface for the users. We chose to address the problem by exploiting the *Prediction Markets*. Prediction markets (also known as predictive markets, information markets or decision markets) are virtual markets created for the purpose of making predictions. The user of these markets makes prediction concerning events in the future by answering to questions. We decided to build a base architecture and interface of a *crowdsourcing* web application of *Prediction Market* and then decline this base architecture in four different scenarios. Each scenario moves from the base architecture harnessing a particular rewards and incentive system but they all share the same way of collecting the knowledge for the sentiment analysis tool from the *crowd*.

The base architecture consists of an online community in which the users can signup and login. After logging in, they can see a list of open questions concerning future predictions. For instance, a possible question can be “Will Milan A.C. win the next Italian football championship in 2011/2012?”. Then, the users can answer to these questions with a Yes or No answer. Thus, they are closed questions. The users can also see the list of answers given by other users in the *crowdsourcing* community as percentage of Yes and No with respect to the total of answers given for each question.

The users can also post new questions. The type of questions that the users can propose on the web application is not unbounded. Indeed, we restricted the field to just two domains: the *football* and the *fashion trend*. We chose these two domains because they are easily manageable by us and they are well known by a large share of the people in the test group. Indeed, we wanted to free ourselves from experimental biases coming from the fact that the questions' domains were disliked by the majority of testers. As it would be clear at the end of our analysis, this would have undermined a great part of our conclusions, in particular the ones concerning the rewards and incentive systems.

The questions must follow a scheme that is fixed by the web application. As we will further describe, the reasons why we chose this approach are two: first we were concerned to develop a limited experiment, second because, as we will see, posting questions is the way how the users provide their knowledge to our system in order to enhance the knowledge domain. In particular, our focus was more on understanding which are the best practices to stimulate user participation to the *crowdsourcing* community and thus we focused our attention on the rewards and incentives and not on the specific knowledge that we could retrieve from the users. As we will discuss in the conclusions, this is the next step of our research and will be addressed in future work.

The next table provides a list of possible question schemes for our experiment.

Table 17: Question schemes for the *crowdsourcing* web application

Domain	Question schemes
Football	▪ Will [Team X] win in [Match Y]? ▪ Will [Team X] buy [Player Y]? ▪ Will [Team X] sell [Player Y]? ▪ Will [Player Y] play in [Match Z]? ▪ Will [Team X] win against [Team Y] in [Match Z]?
Fashion Trend	▪ Will [fashion item X] by [brand Y] be trendy the next (Winter/Summer/Autumn/Spring)? ▪ Will [brand X] be trendy the next (Winter/Summer/Autumn/Spring)?

It's possible to see that the schemes bound the form of the questions. The terms that appear within square brackets are free variables that can be filled by the user input. The terms within round brackets instead are list of values from which a user can pick an option when creating a new question.

We decided to focus also on the social aspects. Our experimental model, while being limited in many aspects and being just a sketch test case, follows a common practice of real social communities. Thus, in the base architecture every user has a personal page. He can access to his personal page and see the list of questions he had answered and posted before.

Every question has a closing date. This closing date is chosen by the poster of the question. After the closing date, the question is closed and the users cannot answer to it anymore. As far as the web application is a declination of a *Prediciton Market* model, every question is a prediction about a future event. Thus, every query must have a closing date, after which either the poster decide to not accept anymore answers or he knows that the real answer will be out, allowing him to compare it with the prediction made by the community. Anyway all the posted questions, either closed or open, are archived and the users can browse the question database and query it.

Figure 27: The interface of the *crowdsourcing* web application for our experiment. Here its declination in the first scenario

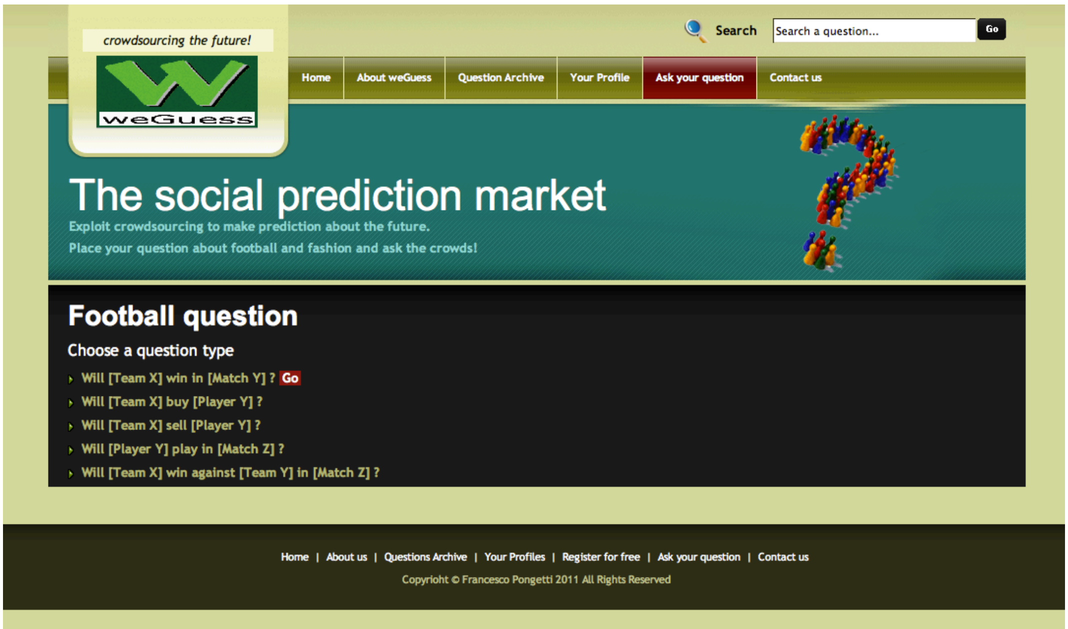

Figure 28: Interface for choosing a question scheme in the case of football question. This part is equal in all the four scenarios

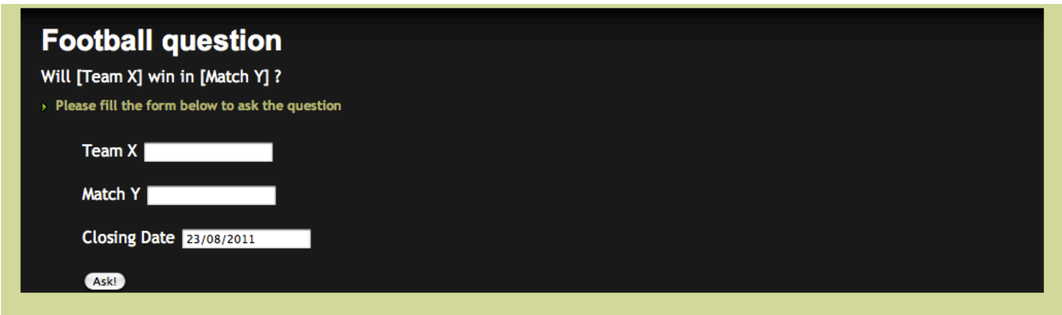

Figure 29: Interface for filling the values of the question scheme when posting a new question, in this case a football question.
This part is equal in all the four scenarios

We developed the *descriptive framework* in order to describe in a coherent and rigorous way a *crowdsourcing* application. Therefore, we can model our experimental base web application by using the *descriptive framework*. As we already stated, each scenarios differs from the others according to the rewards and incentives system. Thus, according to our *descriptive framework* they have different description for their Rewards (Main and Minor), Remuneration and Incentive dimensions. Moreover, the third scenario exploits a slightly more different architecture because it uses a task of type game (see Chapter 3, *descriptive framework*). In the next paragraph we will describe the Rewards (Main and Minor), Incentive, Remuneration, and Task Type dimensions of each scenario according to the *descriptive framework*. Here instead we will describe the other common dimensions that are equals for all the four scenarios and that descend directly from the base architecture.

We have to spend more words concerning the Data Quality Mechanism dimension. It clearly appears that the *crowdsourcing* web application of our experiment can be described according to two parameters. Indeed, we must point out that the tasks performed by the users of the *crowdsourcing* application are two. We already discussed that the main task is providing data for the a sentiment analysis engine. We will describe the methodology for this task further in this section. However the users of the community also perform the task of making predictions while answering to the questions. This comes from our choice of exploiting the *Prediction Markets* as model for a possible online community. It doesn't really matter for the outcome of the experiment but it reflexes in the description of the *crowdsourcing* application according to the *descriptive framework*. Thus, we will fill the dimensions' descriptions taking into account that the tasks performed by the users are actually two.

Coming back to the Data Quality Dimension, we can state that for what concerns the task of populating the domain knowledge, our *crowdsourcing* platform doesn't provide any quality mechanism. For the task of making predictions instead, the Data Quality Dimension assumes a value according to the scenario taken into consideration. Thus, as we already saw for the Task Type, Rewards, Incentive and Remuneration dimensions, the Data Quality Dimension will be also separately discussed for each scenario in the following paragraph.

The next table summarizes the description of the common dimensions and provides our remarks when needed.

Table 18: Part of the dimensions of the *descriptive framework* with their values for the *crowdsourcing* system of our experiment

Dimension	Value(s)	Remarks
Categorization	**Collective Knowledge, Knowledge Sharing, Knowledge Acquisition**	The *crowdsourcing* web application (in all the four test case declinations) is a system to tap the collective knowledge of the people in the *crowds*. The user shares their predictions thus it belongs in the category of Collective Knowledge and Knowledge Sharing. Moreover, the system acquires information for the sentiment analysis engine without sharing them with the *crowd* thus the system is also a platform of Knowledge Acquisition
Crowdsourcing Type	**Integrative**	The type of *crowdsourcing* is integrative. Indeed, all the information coming from the crowds are collected, used and aggregated, both for making predictions or for populating the knowledge domain
Required Knowledge	**Low**	The entry barriers for joining the community are low. No specific skills or education are needed to take part to the *crowdsourcing* community
User Type	**Any**	Any type of users can join the community
Community Size (Qualitative)	**Any**	The pilot community is composed of 51 people but the system is just an experiment. A real community of this type can have any size. In the context of the experiment we cannot describe the application according to this dimensions

Data Quality Mechanism	**None**	The system doesn't provide any data quality mechanism for the task of collecting the information for the knowledge domain. For the task of making predictions see the discussion of each scenarios

We have now to discuss which is the methodology that we exploit in our experiment to populate the knowledge domain. While posting new questions the user have to follow a wizard. This wizard declines in several ways according to the scenario to which it refers but the part in which we collect the information for sentiment analysis engine is common to all the test cases. While posting a new question the user has to input the values for free variables according to the schemes showed previously in this section. These values are the real input for the domain knowledge. When the user inserts a new question concerning football it has to insert the name of a team or of a players or of a match or a combinations of this information. Then, the *crowdsourcing* application, with a design equal in all the four scenarios, takes this input and queries the domain knowledge of the sentiment analysis tool, checking if these keywords match any items already in the knowledge database. In particular, it checks the list of brands and sub-brands (see Chapter 2). The domain knowledge for football is composed of brands, sub-brands (and categories but we will not consider it for our experiment). Teams and players are brands, while football matches are sub-brands of the teams that are playing the competition. Moreover, players are also sub-brands of the team in which they are playing and the championships are sub-brands of the teams that belong to them. If in a new question posted by a user appears an unknown brand or sub-brand, i.e. a brand or sub-brand that is not already in domain knowledge, the web application asks to the users to provide some information concerning these items in order to populate the semantic knowledge database. Thus, in this way we designed a methodology to directly ask to the users information about new brands, sub-brands and the links between them, in order to build the football (or fashion trend in the other case) specific domain knowledge. The systems asks this information about the objects that appear in new football questions:

- For an unknown team the system asks the user to specify the nation and the championship in which the team plays
- For an unknown player the system asks the user to specify the team in which the player plays
- For an unknown match the system asks the user to specify the nation where the competition is taking place, the championship to which it belongs and the teams playing

Thus the system can crawler this data and populate the sentiment analysis tool's knowledge domain accordingly with new brands, sub-brands and it can make the proper connections between them, following the questions' schemes. Moreover, the system also asks to provide information concerning the nation of origin of teams, players and matches. This data is not used in the experiment or in our research but will be used for future researches on the other component of a typical domain knowledge representation, the categories.
The next table summarizes the concept of the football domain.

Table 19: Brands and sub-brands created by the *crowdsourcing* methodology for the football domain

New Brand	New Sub-brand(s)
Team	Match, Player, Championship
Player	Match

What we explained until now for the football field can be repeated for the fashion trend domain. Indeed, it follows exactly the same design, just the questions are different and as consequence also the brands, the sub-brands and the links between these objects. Fashion brands that appear in new questions are new brands for the domain knowledge on the fashion trends. Fashion items instead will be sub-brands of the fashion brands producing it. The systems ask this information about the objects that appear in new fashion questions:

- For an unknown brand the system asks the user to specify the nation and the type of fashion goods it produces (shoes, pants, jackets, etc.)
- For an unknown fashion item the system asks the users to specify the type (shoes, pants, jackets, etc.) and at the same time is able to automatically link it to the brand thanks to the questions' schemes

As we saw for the football domain, the system also asks information that are not directly used in the experiment for creating new brands or sub-brands, namely the fashion goods that a fashion brand produces and the type of a fashion item. Even this data will be used for future researches on the other component of the domain knowledge, the categories.
The next table summarizes the concepts of the fashion trends domain.

Table 20: Brands and sub-brands created by the *crowdsourcing* methodology for the fashion trend domain

New Brand	New Sub-brand(s)
Fashion Brand	Fashion Item

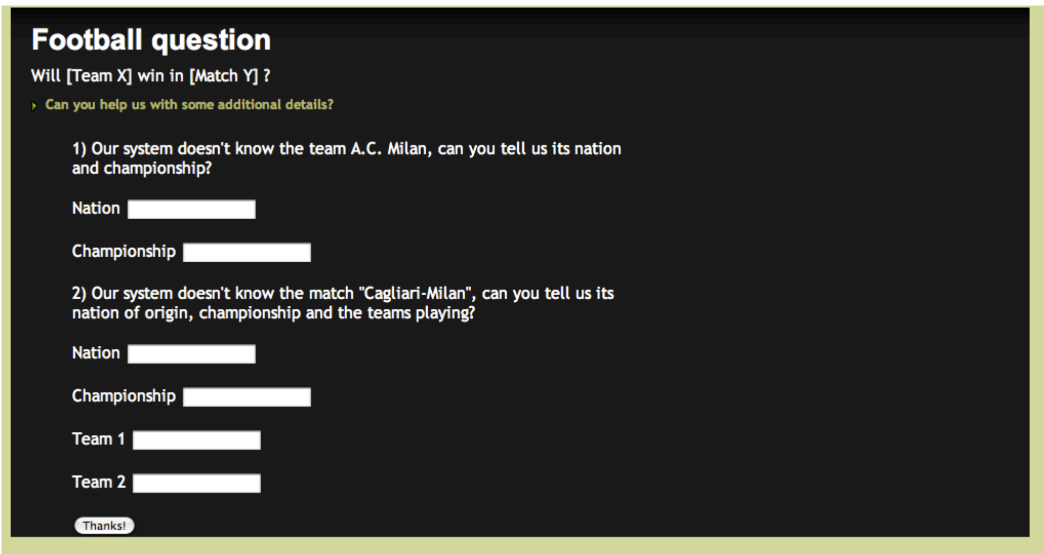

Figure 30: Interface for collecting the information for the sentiment analysis tool in the case of a football question.
This part is equal in all the four scenarios

We will know present the four scenarios descending from the base web application that we described until now. The next table provides for each scenario a description of its main features and design. Then in the next paragraph we will describe them according to the *descriptive framework*.

Table 21: Description of the four test scenarios

Test Scenario	Description
Scenario 1	This scenario is the base web application itself without any changes from the description that we offered previously.
Scenario 2	In this scenario we modify the base architecture adding the possibility for the users of the application to see the ranking of the people in the community according to the number of correct predictions they made. The users can also see correctness statistics of any other users in the community by browsing their personal pages. In this way we tried to exploit a prestige-oriented reward mechanism and the incentives coming with it (see *descriptive framework* and next paragraph).
Scenario 3	This scenario is the most different from the base web application. In this scenario the users plays against their friends. Each user has a list of friends on the website and can challenge them at making new correct predictions posting new questions. The friends can accept the challenge and answer to the question. In this scenario the user can see the personal pages of his friends in addition to his own page. Each user has correctness statistics in his page and his friends can see them. Moreover, the users can see the ranking of his friends and himself according to the number of correct predictions they made. In this scenario we wanted to exploit both a prestige-oriented rewards mechanism with the linked incentives, and a task of type game (see *descriptive framework* and next paragraph).
Scenario 4	This scenario is similar to the first one but we add the

	possibility for the users of betting with money on the correctness of their answers while answering to the questions. Thus they can earn and lost money while answering to the questions. Of course, this is just an experiment, and no real money was involved. However, we asked to the testers to act and think like the money they were betting was real. Moreover, the users get a monetary bonus every ten posted questions that provide new useful information for the sentiment analysis tool, i.e. that contains new brands or sub-brands. In this scenario we wanted to exploit a monetary reward mechanism and the linked incentives (see *descriptive framework* and next paragraph).

As concluding remarks we have to describe the procedure we used to perform the experiment itself. We made these steps:

- We interviewed one tester at the time
- The active tester had first to answer to the first five questions in the survey
- Then, we gave him a laptop asking him to try for a certain amount of time the four scenarios
- When the tester was satisfied with the time spent trying the four scenarios we closed the laptop and asked him to answer to the remaining part of the survey
- We collected all the data from 51 testers and analyzed it

4.3 Description of the scenarios

In this paragraph we will apply the *descriptive framework* to the four test cases. We already provided a description of them in the previous section and we discussed the fact that the majority of the dimensions are in common, deriving directly from the base architecture. Thus, in this section we will comment just the remaining dimensions, i.e. the Rewards (Main and Minor), Incentive, Remuneration, Task Type and Data Quality Mechanism (just for what concerns the task of predicting future events) dimensions. The following tables show the descriptions for each scenario. While reading the tables, it may be useful to keep in mind the *descriptive framework* in order to understand the terms and the values used.

Scenario 1

Table 22: Dimensions of the *descriptive framework* for the first scenario

Dimension	Value and comments

Rewards (Main and Minor)	Main Reward
	→ *Enjoyment-based*
	o Fun
	o Curiosity and desire to test if it work
	o Desire to do something different from your work
	o Desire to express yourself
	o Values and Ideology
	  o Volunteerism and desire to support a cause of the project
	  o Reciprocity, exchange and mutual help
	Minor Reward: None

	This scenario exploits just one reward of type enjoyment based. Indeed, the scenario provides no other reward mechanism besides it. Thus, the users are stimulate to join the community because they find it funny, because they are curious to see how it works, because they simple want to express their predictions or finally because they might share the goals of the community, i.e. helping the enhancing of the sentiment analysis engine's domain knowledge. The scenario doesn't even have any kind of minor reward system
Incentive	→ *Connected to enjoyment-based reward* o Sharing of the result because the user can access to all the predictions coming from the community

	The system has just a reward system of type enjoyment-based, thus the Incentive dimension must reflex this (see *descriptive framework* at this regard). The only incentive provided by the system is the sharing of the result. Indeed, the users can access to the predictions made by other users and see the closed and open questions browsing by querying the database
Remuneration (Qualitative)	None
Task Type	Simple
	The task type is simple because the users can just answer with yes or no statement to any question. This doesn't require a huge effort. Moreover, posting a new question is a simple task as well. Indeed, the users just have to follow an easy wizard with few steps. Providing data for the sentiment analysis tool, is a step in the process of posting new question and doesn't require any effort
Data Quality Mechanism	Group Evaluation [Averaging]
	The quality of predictions coming from the community is ensured by the averaging of the single answer given by the users. Thus, it is a group evaluation mechanism according to our *descriptive framework*. Moreover, it's a classic averaging approach, the same one that we discussed both while describing this dimension in Chapter 3 and while talking of the *Wisdom of the Crowds* in Chapter 2

Scenario 2

Table 23: Dimensions of the *descriptive framework* for the second scenario

Dimension	Value and comments
Rewards (Main and Minor)	Main Reward → <u>*Enjoyment-based*</u> ○ Fun ○ Curiosity and desire to test if it work ○ Desire to do something different from your work ○ Desire to express yourself ○ Values and Ideology ▪ Volunteerism and desire to support a

	cause of the project ▪ Reciprocity, exchange and mutual help Minor Reward → <u>*Prestige-oriented mechanism*</u> ○ Increasing online reputation and recognition
	For what concerns the Main Reward this scenario has the same characteristics of the previous one. Thus, our comments for the first scenario are valid also here. Moreover, in this scenario we exploit also a Minor Reward of type prestige-oriented. Indeed, we introduced in the web application a ranking system of the users in the community and the possibility for every user to see the correctness statistics of the predictions made by the other individuals in the *crowd*
Incentive	→ Connected to enjoyment-based reward ○ Sharing of the result because the user can access to all the predictions coming from the community ○ Competition → Connected to prestige-oriented reward ○ User ranking and statistics of the best users ○ Position inside the community
	The second scenario has all the same enjoyment-based reward system of the first one. Thus, it also exploits the same incentives. Moreover, it has some more incentives connected to the fact that it exploits a prestige-oriented minor reward. Indeed, the prestige-oriented minor reward relies on the user ranking and statistics of the best users system. The Position inside the community is another incentive deriving from the fact that the users with the best performances will be followed in their predictions by the majority of the community. Finally, Competition is a new incentive of this scenario linked to the enjoyment-based. Indeed, as we discussed in the *descriptive* scenario, competition can be fun factor for many users and it is introduced in this test case by the user ranking that move the users at competing between themselves
Remuneration (Qualitative)	None
Task Type	Simple
	The second scenario has the same value of the first one for what regard

	the Task Type dimension, thus it is valid all what we discussed before
Data Quality Mechanism	Group Evaluation [Averaging] Competition
	The second scenario again has the Group Evaluation [Averaging] as value for the Data Quality Mechanism dimension. Thus, what we discussed before for the first scenario is valid also here. Moreover in the second scenario, the data quality is also ensured by the competition among the users that is stimulated by the user ranking and statistics. We already discussed in the *descriptive framework*, how the competition among users can help in achieving an overall higher quality of the outcome of *crowdsourcing*.

Scenario 3

Table 24: Dimensions of the *descriptive framework* for the third scenario

Dimension	Value and comments
Rewards (Main and Minor)	Main Reward → <u>*Enjoyment-based*</u> o Fun (through gaming) o Curiosity and desire to test if it work o Desire to do something different from your work o Desire to express yourself o Values and Ideology o Volunteerism and desire to support a cause of the project o Reciprocity, exchange and mutual help Minor Reward → <u>*Prestige-oriented mechanism*</u> o Increasing recognition (among friends)
	For what concerns the Main Reward this scenario shares many commonalities with the previous one and the first. Thus, our comments for the first and second scenario are valid also here. The difference is that in this case the fun part of the enjoyment-base reward system takes a bigger role because the third scenario is a *crowdsourcing* gaming

	platform (see *descriptive framework* for this definition). We will see while analyzing the result of the survey if increasing the fun reward through gaming is an effective system for stimulating user participation. For what concerns the minor reward, we have even here a prestige-oriented mechanism. Indeed, the competition with the friends, while gaming on making predictions, brings an increasing recognition to the winner users
Incentive	→ *Connected to enjoyment-based reward* ○ Competition and Gaming → *Connected to prestige-oriented reward* ○ User ranking and statistics of the best users (between group of friends)
	The third scenario has incentives connected both to the enjoyment-based reward system and to the prestige-oriented one. For what concerns the first reward we have the incentive Competition and Gaming. The explanation is obvious and comes from the gaming system. For what concerns the prestige-oriented reward we have as in the second scenario user ranking and statistics of the best users. Indeed, also in this case we have ranking and statistics of the best users but they are shared only among friends
Remuneration (Qualitative)	None
Task Type	Game
	The main difference between this scenario and the other three relies on the task type. As we discussed in the previous paragraph the third scenario is a *crowdsourcing* game, following the definition that we proposed in the *descriptive framework*. Thus, the task type is obviously Game.
Data Quality Mechanism	Group Evaluation [Averaging] Competition
	The third scenario again has the Group Evaluation [Averaging] as value for the Data Quality Mechanism dimension. Thus, what we discussed before for the first scenario is valid also here. Moreover, as in the second scenario, we have the Competition as data quality mechanism even here. Thus, what we discussed while presenting the second case is valid also here.

Scenario 4

Table 25: Dimensions of the *descriptive framework* for the fourth scenario

<table>
<tr><th>Dimension</th><th>Value and comments</th></tr>
<tr><td rowspan="2">Rewards (Main and Minor)</td><td>Main Reward
→ <u>Opportunistic</u>
o Direct monetary compensation both for winning a bet/prediction and for providing useful new information for the domain knowledge by posting new questions.
Minor Reward
→ <u>Enjoyment-based</u>
o Fun
o Curiosity and desire to test if it work
o Desire to do something different from your work
o Desire to express yourself
o Values and Ideology
o Volunteerism and desire to support a cause of the project
o Reciprocity, exchange and mutual help</td></tr>
<tr><td>This time the minor reward is of type enjoyment-based and thus, for what concerns this part it is valid what we already said in the previous scenarios. The main reward instead is opportunistic. This is the only scenario that exploits an extrinsic/opportunistic reward system. The users earn money by winning bet, i.e. by performing correct predictions. Thereby the opportunistic reward is a direct monetary compensation. Moreover, they get a monetary bonus every time they reach ten posted questions that provide new information for the sentiment analysis tool. This should stimulate both the answering to the questions (second task) and the creation of new questions (first task)</td></tr>
<tr><td>Incentive</td><td>→ <u>Connected to enjoyment-based reward</u>
o Sharing of the result because the user can access to all the predictions coming from the community.
→ <u>Connected opportunistic reward</u></td></tr>
</table>

	○ Money
	The fourth scenario has the money as incentive. Indeed, it exploits a direct monetary compensation. For what concerns the minor reward, the incentives are the same seen for the first and second scenarios
Remuneration (Qualitative)	Low, Medium, High
	The monetary remuneration provided by this platform varies according to the amount of the money used by the users to bet on predictions. Thus, it encompasses the whole range of Low, Medium and High. Moreover, we supposed that the compensation for posting new questions would be low. Anyway, in the survey we inserted a specific question at this regard
Task Type	Simple
	The fourth scenario again has the same value of the first one for what regard the Task Type dimension, thus it is valid all what we discussed before
Data Quality Mechanism	Group Evaluation [Averaging] Reward Accuracy
	The fourth scenario again has the Group Evaluation [Averaging] as value for the Data Quality Mechanism dimension. Thus, what we discussed before for the first scenario is valid also here. Moreover, this scenario has a Reward Accuracy system as data quality mechanism. Indeed, the system of the fourth scenario provides the monetary bonus only to people in the *crowd* able to post at least 10 questions providing new information for the sentiment analysis tool. In addiction, the user can earn money by betting, only if making correct predictions. This should ensure that the users would try to post each time question regarding new brands and sub-brands and to answer at their best to the questions.

4.4 The experiment and the prescriptive framework

In this section we will discuss to which extent the analyses performed in the context of the *prescriptive framework*, helped us in shaping the experiment. At this regard, we can state that the *prescriptive framework* told us how to choose the four scenarios to test.

We already know from the beginning that the experimental *crowdsourcing* web application that we have to build was a knowledge community, i.e. a "Knowledge Sharing", "Knowledge Acquisition" or "Collective Knowledge" *crowdsourcing* application. Indeed, this emerges from the paragraph 4.2 in which we described the Categorization dimension for the four scenarios.

In the analysis of the relationship Categorization – Incentive in the *prescriptive framework*, we understood that knowledge communities almost in any case, exploit "Sharing of the goals" and "Sharing of the results" incentives. Moreover, we saw that the greatest part of *crowdsourcing* communities belong to this category. Thus, we developed the base web architecture in order to harness these two incentives. As counter-argument we also tested the case of the third scenario where the "Sharing of the results" is not exploited, thus removing this incentive from the base architecture.

In the analysis of the relationship Main Reward – Task Type, we saw that generally *crowdsourcing* systems with simple tasks exploit enjoyment-based main rewards, while opportunistic rewards are usually offered in the case of complex tasks. Thus we focus more on the former and three out four scenarios focus on this type of reward mechanism. Moreover, usually enjoyment-based rewards are coupled with prestige-oriented rewards (see Minor Reward – Main Reward relationship) and we tested this design in the second scenario.

The analysis of the ternary relationships Incentive – Main Reward – Minor Reward, Incentive – Task Type and Incentive – Required Knowledge, showed us which other incentives we could exploit in our test cases. In particular, we saw that the incentives Competition, User Ranking and Voting systems were harnessed by *crowdsourcing* application with both enjoyment-based and opportunistic reward systems. Thus, we designed a limited version of these incentives in our scenarios: the second scenario exploits both a user ranking system coupled with user statistics and as consequence of this, also competition. The third scenario instead introduces competition by gaming.

The rationale behind the gaming system of the third scenario comes from several factors. First, we saw that our empirical dataset had few examples of this type of *crowdsourcing* task and thus we wanted to study more in deep the topic. Second, from several analyses in the *prescriptive framework* it turned out that these platforms usually present an original behavior w.r.t. the other approaches, and therefore they may offer some kind of advantages. Moreover, we already had in our mind the idea of building a game as final implementation for a *crowdsourcing* methodology for the *sentiment analysis* (see Conclusions).

Finally, some considerations regarding the fourth scenario: from the *prescriptive framework* analyses, in particular from the Main Reward – Minor Reward relationship, clearly emerges the importance of the extrinsic/opportunistic reward mechanism and of the monetary incentive. Previously, we formulated the hypothesis that this is probably the best reward for stimulating user participation. Thus, we built the fourth scenario to explore this type of systems and test our hypothesis.

As concluding remark, the review performed in the context of the *prescriptive framework* gave us some data regarding the priority of the reward systems and how to state if a reward would have been the minor or the main for our scenarios. In particular, we saw that the prestige-oriented reward mechanism never appears as main reward in real *crowdsourcing* implementation. Thus, we described our test cases accordantly. This is true also for the opportunistic reward mechanism that we saw being always the main reward when exploited (see Main Reward – Minor Reward relationship). At this regard, we will see that our experiment provides counter-arguments to this conclusion and we will widely discuss this finding.

4.5 Survey

Table 26 presents the list of questions composing the survey that we proposed to the test users. The survey consists of 23 questions and tries to cover all the relevant aspects of the *crowdsourcing* phenomenon and of our research. In particular, each question refers to a specific topic. Assessing the user satisfaction while using a *crowdsourcing* system, has been the main driver behind the survey, but the questions decline this goal in the specific sub-components of a *crowdsourcing* platform, according to the *descriptive framework.*

The first group of questions ("User Information") helps us to identify, classify and analyze our base of test users. They are general questions concerning the test users' age, sex and education level. Moreover, they also assess the amount of time spent online and the general attitude of the users toward the online web communities of knowledge sharing (with the question "I am likely to join online web communities and to share my knowledge").

Other questions refer to a specific dimension (or set of dimensions) discussed in the *descriptive framework* (for instance, Size dimension, Reward dimension, etc.). These questions may or not involve the test scenarios. The last group of questions ("Scenarios") refers to the scenarios themselves: they are used to perform general comparisons and analyses between the scenarios. The "Subject" column of Table 17 specifies to which group a question belongs or to which dimension(s) refers.

As we already discussed while describing the experiment, we divide it in two parts. In the first part the users answer to a set of questions before trying the *crowdsourcing* scenarios. Then, they have to play with the scenarios and finally they answer to the remaining questions. At this regard, the questions numbered in bold in Table 17 belong to the second part of the experiment.

Table 26 also shows the values that the answers can assume under the column "Value". The test user can provide just one answer for each question. The answers can be a number (for instance when specifying the age), or a value from a multiple choice. In the latter case we often used a Likert-type scale. A Likert-type scale, is a psychometric scale commonly used in questionnaires, and is the most widely used scale in survey research, such that the term is often used interchangeably with rating scale even though the two are not synonymous. When responding to a Likert questionnaire item, respondents specify their level of agreement or disagreement on a symmetric agree-disagree scale for a series of statements. Thus, the scale captures the intensity of their feelings according to four five possible values: strongly disagree, disagree, neither agree nor disagree, agree and strongly agree. The questions without a Likert-type answer use instead a text list of possible choices.

Finally, several questions ask the respondent to order the four scenarios or other items according to one of their properties or features.

The survey was designed using the tools offered by Google Docs and accessible by any web browser. The user, while answering to the questionnaire, was overseen by a researcher. Thus, the tester had the chance of asking any kind of explanations regarding the questions. Of course the researcher didn't interfere in any way with the testers providing help just when necessary and he didn't check the answers provided by the users at the moment of filling the survey. The questionnaire was completely anonymous. In this way, we ensured a higher quality of the experiment's outcome. All the questions in the survey were mandatory.

Lastly, a final notice concerning the question about the education level: the education level considered is the academic title that the respondent held or was pursuing at the time of the survey.

While presenting the results of our experiment, we will discuss each question analyzing the meaning and explaining why it was chosen as part of the survey. In particular, we will provide the disaggregate results for each question before analyzing the aggregate data.

Table 26: List of questions in the survey

#	*Question*	*Value*	*Subject*
1	Please enter your age	*Numeric [18 - 99]*	User Information
2	Please enter your sex	*[Male; Female]*	User Information
3	Please enter your education level	*[High school diploma; Bachelor Degree; Master Degree; PhD; Other]*	User Information
4	Please enter the number of hours spent online in a day	*Numeric [1 - 24]*	User Information
5	I am likely to join online web communities and to share my knowledge	*Likert-type scale*	User Information
6	I think the size of the community matters when	*Likert-type scale*	Size dimension

	choosing if joining or not an online community		
7	I would share my knowledge and spend my time contributing to online communities for free if I share the goals	*Likert-type scale*	Reward and Incentive dimensions
8	I would share my knowledge and spend my time contributing to online communities for free if I can access to the results and they are useful to me	*Likert-type scale*	Reward and Incentive dimensions
9	I would share my knowledge and spend my time contributing to online communities for free if I get fun	*Likert-type scale*	Reward and Incentive dimensions
10	I am likely to contribute much more to an online community if I can receive any kind of monetary compensation even if I share the goals and/or the results	*Likert-type scale*	Reward and Incentive dimensions
11	I am likely to contribute to online communities only if performing simple (tag images, assign values, etc.) tasks	*Likert-type scale*	Task Type dimension
12	I am likely to contribute to online communities also by performing complex (write long essays, translation of text, etc.) tasks if I have the skills	*Likert-type scale*	Task Type dimension
13	I would not join an online communities and/or perform tasks if they are not linked to my interests even if I get a monetary compensation	*Likert-type scale*	Reward and Incentive dimensions
14	It wasn't a problem for me to provides some additional information while placing new questions	*Likert-type scale*	Scenarios
15	Please order the four scenarios according to how likely would you join each of them	*[Scenario 1; Scenario 2; Scenario 3; Scenario 4]*	Scenarios
16	Please order the four scenarios according to their enjoyableness	*[Scenario 1; Scenario 2; Scenario 3; Scenario 4]*	Scenarios
17	Please order the four scenarios according to the number of questions that you are likely to answer or post in each of them	*[Scenario 1; Scenario 2; Scenario 3; Scenario 4]*	Reward and Incentive dimensions
18	Which range of compensation would you judge fair for providing us useful information by placing new questions?	*[0.01-0.10 euro for question; 0.10-1 euro for question; I don't care / everything would be fine]*	Remuneration dimension
19	Instead of money would you accepted some other form of compensation (for instance few free analysis	*[Yes, if I judge the monetary compensation*	Remuneration, Reward and

	by the sentiment engine for your purposes) for providing us useful information by placing new questions?	*low; Yes, even if I judge the monetary compensation adequate; No, in any case]*	Incentive dimensions
20	Please order the following incentives according to how much they encourage you to correctly answer to the questions	*[User ranking in the second scenario; Gaming in the third scenario; Monetary betting in the fourth scenario]*	Data Quality Mechanism dimension
21	I would try to correctly answer the questions at my best even in the first scenario	*Likert-type scale*	Data Quality Mechanism dimension
22	Would you consciously provide false/incorrect information in the fourth scenario just to benefit of the compensation?	*[Yes; Maybe; No]*	Data Quality Mechanism dimension
23	Would you consciously provide false/incorrect information in the fourth scenario just to benefit of the compensation knowing that the website uses automatic and/or manual mechanisms to avoid cheating?	*[Yes; Maybe; No]*	Data Quality Mechanism dimension

4.6 Test group description

In this paragraph we will describe the test group of users that we used for collecting the data for our experiment.

The test group was composed of 51 people and each of them answered to all the questions composing the survey that we presented in the previous section. As we said, the first five questions, were mean for discovering important information concerning the testers. In particular, we wanted to ensure that our test group was heterogeneous enough according to age, sex, education level, etc. As condition for becoming a tester we requested the users to be *active surfers*, i.e. to usually use the Web at home and not just at work. Moreover, we didn't request the users to have a specific level of technology expertise nor to have previous experiences with online web communities.

In the following part, we will review each of the first five questions and provide the results of the survey.

Age distribution

The majority of the testers belong to the range of age between 20 and 30 years and 30 and 40 years. The reason is that these users are the biggest Internet users and as consequence they have greater chances of joining online web communities. Anyway, we tried to have an as much as possible heterogeneous test group, thus it's possible to see that we interviewed also users with several other different ages, ranging from 19 to 66 years (see Appendix B for the complete list of testers' ages).
Figure 31 presents a chart depicting the age distribution.

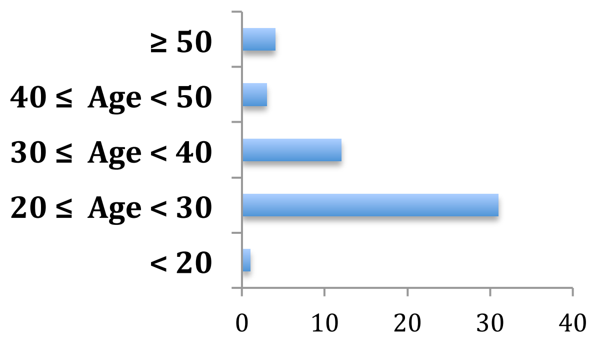

Figure 31: Age distribution chart among the test users

Table 27 presents the age distribution with percentages and the number of users belonging to each range. It also shows the average tester's age and the variance of the sample

Table 27: Age distribution among the test users

***Due to rounding, the percentages may not add up to 100%**

< 20	**1**	**2%**
20 ≤ Age < 30	**31**	**61%**
30 ≤ Age < 40	**12**	**24%**
40 ≤ Age < 50	**3**	**6%**
≥ 50	**4**	**8%**
Average	**30.45**	**-**
Variance	**116.25**	
Total	**51**	**100%**

Sex distribution

The majority of the testers are males. Anyway, the difference between the number of females and males in the test group is little. Thus, the test group is heterogeneous according to the sex distribution.

Figure 32 presents a chart depicting the sex distribution.

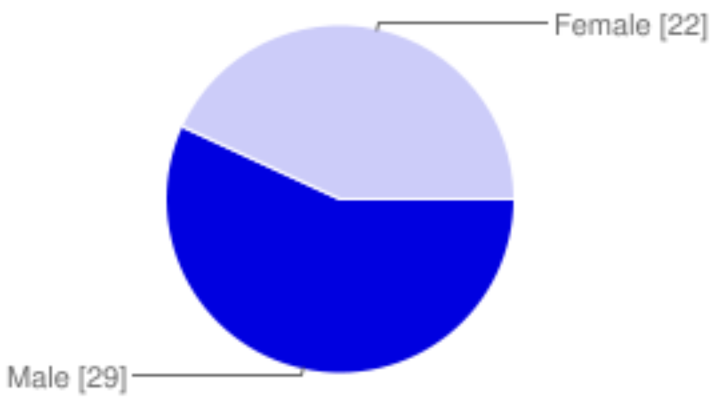

Figure 32: Sex distribution chart among the test users

Table 28 shows the sex distribution with percentages and the number of users belonging to each sex.

Table 28: Sex distribution among the test users

***Due to rounding, the percentages may not add up to 100%**

Male	**29**	**57%**
Female	**22**	**43%**
Total	**51**	**100%**

Education level distribution

As we already said, the education level is the academic title that the tester held or was pursuing at the time of the experiment.

It's possible to see from the next chart that the majority of the testers hold a Master Degree or at least a Bachelor Degree. Thus, we were not able to have an equal distribution of testers among the various categories of education level. Anyway, few users in the test groups hold or are pursuing a PhD while few less have just a High School Diploma. People belonging to the category Other either have a lower degree than the High School Diploma or have some kind of vocational education.

Figure 33 presents a chart depicting the education level distribution.

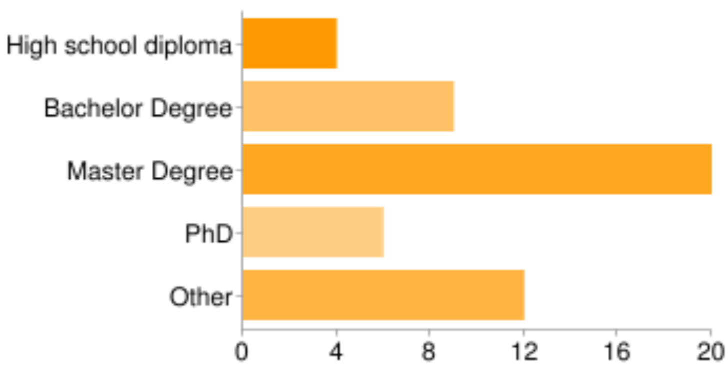

Figure 33: Education level distribution chart among the test users

Table 29 presents the education level distribution with percentages and the number of users holding each academic title, including Other.

Table 29: Education level distribution among the test users

***Due to rounding, the percentages may not add up to 100%**

High School Diploma	**4**	**8%**
Bachelor Degree	**9**	**18%**
Phd	**20**	**39%**
Other	**6**	**6%**
Total	**12**	**100%**

Hours spent online a day

We wanted to check that the testers were active Internet users. Thus we asked to provide the average number of hours spent online each day. While interviewing the users we found out that is a common practice to surf the Net during office time for personal purposes. Thus, we took into account this factor and the number of hours spent online is the overall average, including the time spent online for personal purposes at work.

The majority of testers spend about two or more hours a day surfing the Web for purposes unconnected to their jobs and/or education. Few users spend just one hour a day. Moreover, it's possible to see that many testers spend even 4 or more hours a day online. Thus, we can see that the test group was biased to very active Internet users. This is not a problem for our research because the probability of engaging with online web communities grows proportionally with the time spent online. Thus, active and very active Internet users are the main target users of our research.

Figure 34 shows a chart depicting the hours spent online a day by the test users.

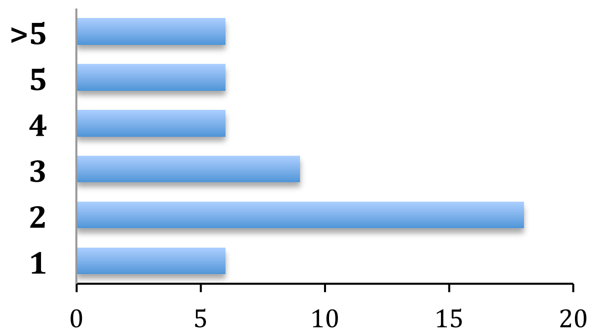

Figure 34: Hours spent online a day by the test users chart

Table 30 shows the hours spent online a day by testers with percentages and the number of users belonging to each time range.

Table 30: Hours spent online a day by the test users

***Due to rounding, the percentages may not add up to 100%**

1 hour a day	**6**	**11%**
2 hours a day	**18**	**35%**
3 hours a day	**9**	**17%**
4 hours a day	**6**	**11%**
5 hours a day	**6**	**11%**
More than 5 hours a day	**6**	**11%**
Total	**51**	**100%**

General attitude toward online web communities and knowledge sharing

We assessed the general attitude of the users in the test group, toward online web communities of knowledge sharing (see the *descriptive framework* for Categorization) through the question: “I am likely to join online web communities and to share my knowledge”. This question had a Likert-type scale as possible values for the answer. We assessed this general sentiment because we wanted to be sure that there was no strong bias among the people in the test group, against online communities. Indeed, a strong bias would have undermined the result of the research. This bias could come from a strong privacy sentiment that prevents from joining any type of community or from an overall low feeling toward the sharing of information on Internet, independently from the design, the purposes and the reward systems proposed by the *crowdsourcing* platform.

It's possible to see from the next chart that the testers are usually well oriented towards online web communities. The majority of them agrees or strongly agrees with the statement in the question. No tester has a strong bad feeling against online communities ("Strongly Disagree"), while some of them usually don't easily join web communities ("Disagree"). A good percentage of the users in the test group neither agree nor disagree with the statement in the question. They don't have any negative sentiments against web community but usually they think carefully before joining one of them.

Figure 35 shows a chart depicting the answer to the question according to a Likert-type scale.

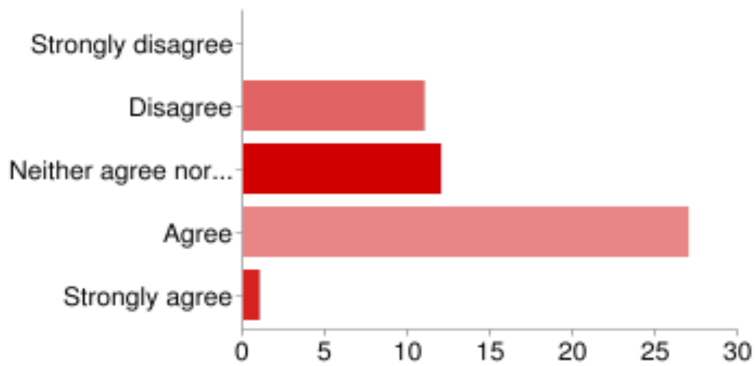

Figure 35: Answers distribution chart for the question "I am likely to join online web communities and to share my knowledge"

Table 31 shows the answer to the question with percentages and the number of users choosing each option.

Table 31: Answers distribution for the question
"I am likely to join online web communities and to share my knowledge"
***Due to rounding, the percentages may not add up to 100%**

Strongly Agree	**1**	**2%**
Agree	**27**	**53%**
Neither Agree or Disagree	**12**	**24%**
Disagree	**11**	**22%**
Strongly Disagree	**0**	**0%**
Total	**51**	**100%**

Heterogeneity and representativeness of the test group

As we said, we tried to select an as much as possible heterogeneous test group in order to not introduce biases in our analysis. Moreover, we wanted our test group to be representative of the typical users that surf the Web and actively use Internet in Italy. At this regard we used the study of Livraghi (2011), for comparing the characteristics of our test group, with the ones of the Italian Internet users. For what concerns the age and sex distributions, our test group approximately follows the same patterns that we have for the overall Italian case. However, the level of education of our testers, is higher then the one of the typical Italian Internet users. Indeed, the average level of education of the Italian Internet users is the High School Diploma (Livraghi, 2011), while in the case of our test group is a university degree (bachelor or master). Anyway, we think that this factor doesn't introduce a significant bias to our analysis, being all the other characteristics approximately distributed as in the general Italian case.

4.7 Results

We will now present the result of the survey. For each question we provide detailed statistics concerning the answers given by the testers. Moreover, we analyze the meaning of the results and how they reflex on our prospective. After listing all the results for each question, we will comment in the next paragraph, the general results of the survey, pointing out our overall conclusions gained from the experiment and how this experience can help us to shape a novel *crowdsourcing* methodology for the sentiment analysis tools.

Question 6. I think the size of the community matters when choosing if joining or not an online community

While analyzing the Size Dimensions (Qualitative and Quantitative) in the context of the *descriptive framework* we pointed out that the size of the user-base (or the *crowd*) is a decisive factor for any *crowdsourcing* project. The reason is that the size of the *crowd* influences both the quality of the *crowdsourcing's* output (see *descriptive framework* and *Wisdom of the Crowds,* Chapter 2-3) and the general attitude of the individuals toward the community. A vital and vast community has more chances of engaging a higher number of new joiners and, moreover, people tend to be more active in big communities (see *descriptive framework*, Chapter 3). With this question we tried to bring experimental data at the support of our statements. Thus, we asked the testers to express their opinion concerning the size of a generic *crowdsourcing* community. We can see that the majority of the testers agree or strongly agree with the statement of question 6 while no tester disagrees or strongly disagrees. We can conclude that as we stated in the *descriptive framework*, the community size is a fundamental aspect both for the quality of the crowd-sourced data and for the users in the *crowds* themselves. Thereby, the designers of *crowdsourcing* platform should try to maximize it.

Figure 36 shows a chart depicting the answer to the question according to a Likert-type scale.

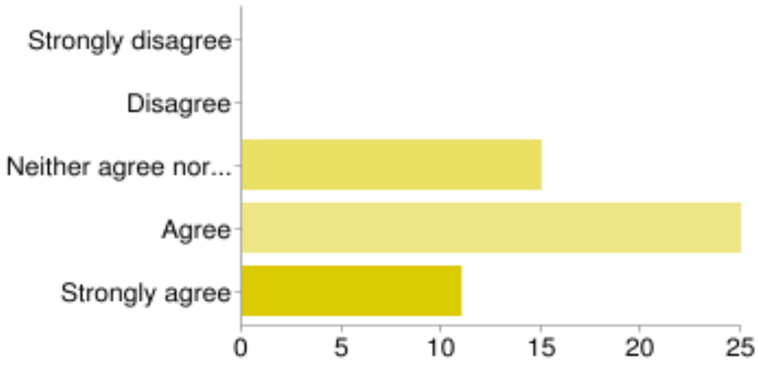

Figure 36: Answers distribution chart for question 6

Table 32 shows the answer to the question with percentages and the number of users choosing each option.

Table 32: Answers distribution for question 6
***Due to rounding, the percentages may not add up to 100%**

Strongly Agree	**11**	**22%**
Agree	**25**	**49%**
Neither Agree or Disagree	**15**	**29%**
Disagree	**0**	**0%**
Strongly Disagree	**0**	**0%**
Total	**51**	**100%**

Question 7. I would share my knowledge and spend my time contributing to online communities for free if I share the goals

In the *descriptive framework* we concluded that one of the component of the enjoyment-based reward mechanism is that the users often participate in *crowdsourcing* communities because they share the goals that power the community itself. Moreover, we pointed out that sharing of the goals, together with the sharing of the results, is a fundamental incentive for the enjoyment-based reward mechanism. This came out also from our considerations concerning the Incentive dimension in the *prescriptive framework* (see the *prescriptive framework*). We pointed out the example of Wikipedia where users basically contribute because they believe in the idea of free knowledge. Moreover, we showed the outcome of a survey from which comes out that the sharing of the goals ("Values" in the survey) is the second most important reason for becoming a Wikipedians (see *descriptive framework*). Thus, we argued that the sharing of the goals is often a decisive factor for encouraging the users' engagement. In our experiment, we asked to the testers, if they agree or not with the statement in question 7 that exactly focuses on assessing the importance of "sharing of the goals". In particular, the question tries to clear the ground from other reward mechanisms. Indeed, it contains the formula "for free" meaning that no other form of compensation would be given besides the "sharing of the goals". It results that all the 51 testers agree or strongly agree with the statement; thereby they provided strong evidence that sharing of the goals is a decisive factor for stimulating user participation in an online community of *crowdsourcing.* We can conclude that the designers should take care of clearly pointing out the goals behind a *crowdsourcing* project even in the case of a not-for-free initiative backed by a company for business reasons. Of course, *crowdsourcing* projects with good social goals such as Wikipedia or Kickstarter, benefit from this. However, the sharing of the goals isn't limited just to benefic social initiative. Consider for instance the case of Waze, it strongly tries to make its users aware of the benefits that they will receive joining its community and being active contributors. Thus, it moves the users toward the sharing of its goals in order to make them active contributors.

Figure 37 shows a chart depicting the answer to the question according to a Likert-type scale.

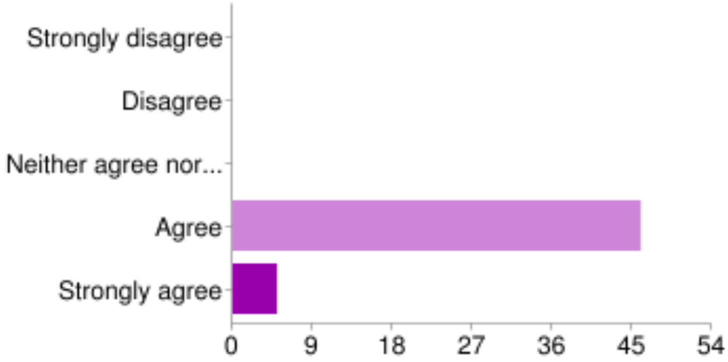

Figure 37: Answers distribution chart for question 7

Table 33 shows the answer to the question with percentages and the number of users choosing each option.

Table 33: Answers distribution for question 7

***Due to rounding, the percentages may not add up to 100%**

Strongly Agree	**5**	**10%**
Agree	**46**	**90%**
Neither Agree or Disagree	**0**	**0%**
Disagree	**0**	**0%**
Strongly Disagree	**0**	**0%**
Total	**51**	**100%**

Question 8. I would share my knowledge and spend my time contributing to online communities for free if I can access to the results and they are useful to me

Following what we did in question 7, here we analyze another component of the enjoyment-based reward system, in this case the Sharing of the result (see *descriptive framework* for a definition). Indeed, many online *crowdsourcing* communities share the results produced by their platforms. We had evidence of this from our analysis of the empirical dataset in the context of the *prescriptive framework*. In particular, sharing of the results was one of the most exploited incentives from communities with a main reward of type enjoyment-based. Moreover, it was one of the most exploited incentives overall. In our analysis of the Categorization – Incentive relationship in the *prescriptive framework* we also saw that the communities of knowledge ("Collective Knowledge", "Knowledge Sharing" and "Knowledge Acquisition") usually widely use this incentive for stimulating user participation. This result is particularly interesting to us, because as we said in the paragraph 4.2 our base web application, and all its declinations in the four scenarios, falls in the three categories that compose the knowledge communities taxonomy: "Collective Knowledge", "Knowledge sharing" and "Knowledge Acquisition". Thus, it turns out that we are clearly interested at assessing the validity of the incentive "Sharing of the results". Moreover, a *crowdsourcing* methodology for the sentiment engine would surely fall at least, in the category "Knowledge Acquisition". Question 8 tries to bring some statistical data at this regard. We asked to the users to which grade they agree with the statement of the question. As for question 7 we tried to clear the ground from other reward mechanisms with the formula "for free" meaning that no other form of compensation would be given. The result shows that the majority of testers agree or strongly agree with the statement, thus they are deeply more stimulated at contributing to a *crowdsourcing* project if they can access to the outcome of the *crowdsourcing*.

Figure 38 shows a chart depicting the answer to the question according to a Likert-type scale.

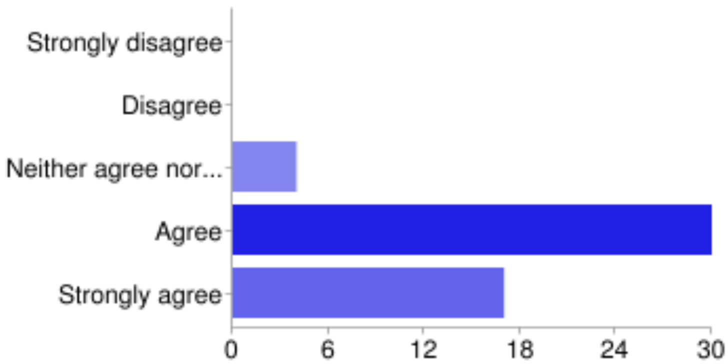

Figure 38: Answers distribution chart for question 8

Table 34 shows the answer to the question with percentages and the number of users choosing each option.

Table 34: Answers distribution for question 8
***Due to rounding, the percentages may not add up to 100%**

Strongly Agree	**17**	**33%**
Agree	**30**	**59%**
Neither Agree or Disagree	**4**	**8%**
Disagree	**0**	**0%**
Strongly Disagree	**0**	**0%**
Total	**51**	**100%**

<u>*Question 9*</u>. I would share my knowledge and spend my time contributing to online communities for free if I get fun

With this question we tried to assess how important is the fun part for a user who takes part to a *crowdsourcing* community. As we said in the *descriptive framework* talking of the type of task, several *crowdsourcing* communities try to stimulate user engagement by proposing tasks of game type. In every case these *crowdsourcing* projects are knowledge communities, i.e. they are “Collective Knowledge”, “Knowledge Sharing” or “Knowledge Acquisition” *crowdsourcing* platforms according to the Categorization dimension of our *descriptive framework*. This emerges from the analysis of the empirical dataset (see the relationships involving the Task Type dimension, at this regard). We saw that fun is a fundamental component of the enjoyment-based reward mechanism and we had evidence of this also analyzing the Wikipedia’s community: it emerged from the survey performed by Nov (2007) that the main reason why contributing to Wikipedia is because writing or editing articles is funny. Moreover, our third scenario is of type game and thus we as well tried to widely analyze to which extend the fun side can help at stimulating user contribution. With this question we try to provide an overall general view of the topic directly asking if the users would be interested at contributing to a *crowdsourcing* community just to spend their spare time in a “funny” activity without receiving any other form of compensation. As for question 7 and 8 we introduced the formula “for free” in order to clear the ground from other reward mechanisms. It turns out that a slightly majority of the testers (53%) disagrees or is neutral with the statement. Thus, we can argue that the overall sentiment toward taking part to *crowdsourcing* projects just for fun is low. Anyway, a good share of the user agrees or strongly agrees with the statement. They still represent a good pool of individuals to tap for *crowdsourcing* purposes. Moreover, the majority of users that doesn’t agree with the statement are neutral at this regard and don’t express an overall negative attitude to the fun side of online communities. Thus, we can conclude that a good *crowdsourcing* methodology should try to exploit other rewards and incentives coupled with the fun, in order to tap several different groups of people. Analyzing other questions concerning the third scenario we will come back on this topic.

Figure 39 shows a chart depicting the answer to the question according to a Likert-type scale.

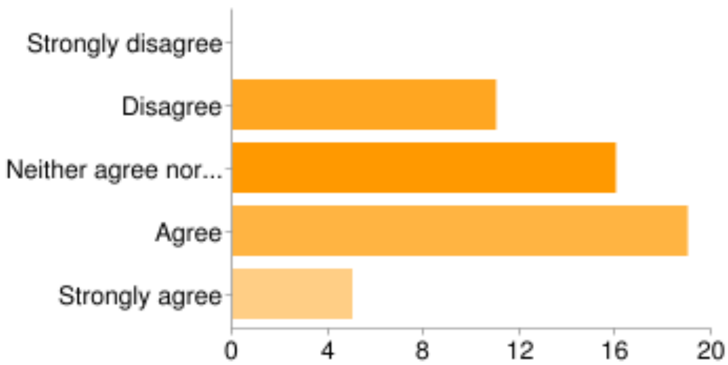

Figure 39: Answers distribution chart for question 9

Table 35 shows the answer to the question with percentages and the number of users choosing each option.

Table 35: Answers distribution for question 9

***Due to rounding, the percentages may not add up to 100%**

Strongly Agree	**5**	**10%**
Agree	**19**	**37%**
Neither Agree or Disagree	**16**	**31%**
Disagree	**11**	**22%**
Strongly Disagree	**0**	**0%**
Total	**51**	**100%**

Question 10. I am likely to contribute much more to an online community if I can receive any kind of monetary compensation even if I share the goals and/or the results

Question 10 tries to assess to which level the extrinsic/opportunistic reward mechanism and incentives stimulate user engagement in online *crowdsourcing* communities. In particular, it focuses on the direct monetary compensation as reward system and the money as incentive (see Rewards and Incentive dimensions in the *descriptive framework*). This is a first general question and we will study further the topic analyzing the results concerning the fourth scenario that, as we said, involves the betting of real money and monetary compensation. We wanted to focus on the contributions derived just from the extrinsic incentive and thus we formulated the statement of question 10, asking to the testers if they would contribute (much) more to an online community just because of the monetary incentive even if they already share its goals and they can access to the *crowdsourcing* results. These two incentives were already covered by the previous questions.
It's possible to see in Figure 40, that there is no strong bias toward a specific answer. The only clear result is that no tester strongly disagree with the statement thereby no one completely exclude the chance that a monetary compensation would increase his activity level in a *crowdsourcing* project. Actually, the relative majority of the testers agrees or strongly agrees with the statement. This brings support to the hypothesis

that we formulated in the *descriptive framework* (and that we also remarked in the *prescriptive framework*), i.e. the users are deeply stimulated by an opportunistic reward mechanism, in particular by a direct monetary compensation for their contributions. Moreover, we can argue again, and now with experimental statistical data supporting us, that the opportunistic reward mechanism is the main reward in all the systems that exploit it. We already reached this conclusion previously, while analyzing the relationships concerning the Main and Minor reward, in the *prescriptive framework*. Thus, we can already conclude that the direct monetary compensation seems the best way for stimulating user participation to *crowdsourcing* projects. However, we will come back with more details discussing the fourth scenario. Still there is a big share of testers that disagrees with the statement. Excluding any kind of moral prejudice toward the money by the testers, we can conclude that the money often doesn't suffice if considered alone, at stimulating part of the individuals that may compose a *crowd.* We already saw it commenting the result of the Nov's survey in the context of the *descriptive framework*. At this regard, it is valid again what we conclude for question 9, that a good combination of incentives and rewards helps at tapping the biggest share of the people.

Figure 40 shows a chart depicting the answer to the question according to a Likert-type scale.

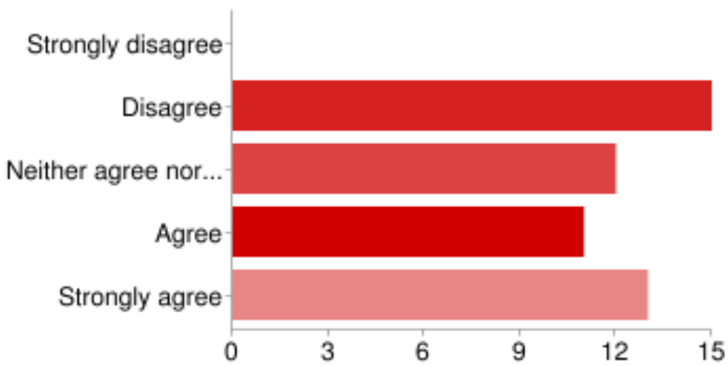

Figure 40: Answers distribution chart for question 10

Table 36 shows the answer to the question with percentages and the number of users choosing each option.

Table 36: Answers distribution for question 10
***Due to rounding, the percentages may not add up to 100%**

Strongly Agree	**13**	**25%**
Agree	**11**	**22%**
Neither Agree or Disagree	**12**	**24%**
Disagree	**15**	**29%**
Strongly Disagree	**0**	**0%**
Total	**51**	**100%**

<u>Question 11 and Question 12</u>.

I am likely to contribute to online communities only if performing simple (tag images, assign values, etc.) tasks

I am likely to contribute to online communities also by performing complex (write long essays, translation of text, etc.) tasks if I have the skills

We will analyze together question 11 and question 12 because they both focus on the Task Type dimension of our *descriptive framework*. From this point of view, they are complementary questions and must be studied together. We asked to the testers, which is the tasks' maximum complexity that they would accept in order to take part to a *crowdsourcing* community. At this regard, we provided in question 11 some examples of simple tasks (tag images, assign values, etc.) and assessed how much the testers are likely to join a community that requires the users to accomplish them. Both question 11 and question 12 are interesting to us for several reasons: first, we are interested to check which level of efforts can we ask to the users of a future *sentiment analysis crowdsourcing* community; second we wanted to link our theoretical analysis of the Task Type dimension with experimental data. Thus, we inserted also these two questions in the survey.

It's possible to see in Figure 41, that the relative majority of the testers (47%) agrees or strongly agree with the statement, thus they are interested only in communities with simple tasks. Anyway, 29% of the testers are neutral and 24% disagree. Thus, we need to move to the next questions to analyze better these results. In question 12 we preceded in the same way of question 11, proposing a list of possible complex tasks (write long essays, translation of text, etc.) and assessing how many users would accomplish them. While reading the answers of questions 11 and 12 we were expecting that the users would have provided complementary data. Namely, the number of users that agree or strongly agree with the question 11's statement would have disagreed or strongly disagreed with question 12's statement. However, no tester strongly disagrees with question 12 while only 27% disagree. This is far way from the 47% of users that expressed strong bias toward simple tasks. We can argue that we got these results because the testers tend to not exclude *a priori* any type of involvement, thus they don't express a strongly negative feeling in question 12. Moreover, question 11 and 12 concentrate just on the complexity of the tasks and not on the nature of them: some users may perform also really complex tasks if they strongly likely them while they will also perform simple boring tasks if stimulated by other form of incentives unlinked to the nature of the job, such as the direct monetary compensation. This emerges also from the strong component of neutral answers to both the questions. We can conclude that the individuals in the *crowds* don't exclude *a priori* any complexity of tasks but what really matters is the incentives and rewards system set up by the *crowdsourcing* platforms

Figure 41 shows a chart depicting the answer to the question 11 according to a Likert-type scale.

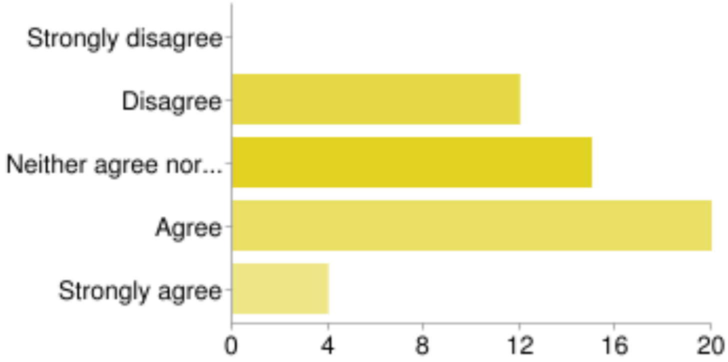

Figure 41: Answers distribution chart for question 11

Table 37 shows the answer to the question 11 with percentages and the number of users choosing each option.

Table 37: Answers distribution for question 11
***Due to rounding, the percentages may not add up to 100%**

Strongly Agree	**4**	**8%**
Agree	**20**	**39%**
Neither Agree or Disagree	**15**	**29%**
Disagree	**12**	**24%**
Strongly Disagree	**0**	**0%**
Total	**51**	**100%**

Figure 42 shows a chart depicting the answer to the question 12 according to a Likert-type scale.

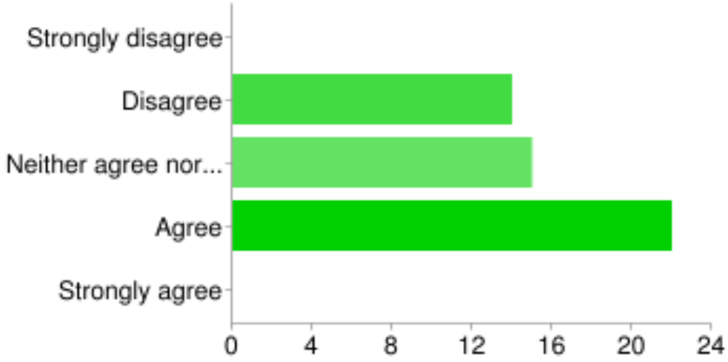

Figure 42: Answers distribution chart for question 12

Table 38 shows the answer to the question 12 with percentages and the number of users choosing each option.

Table 38: Answers distribution for question 12

***Due to rounding, the percentages may not add up to 100%**

Strongly Agree	**0**	**0%**
Agree	**22**	**43%**
Neither Agree or Disagree	**15**	**29%**
Disagree	**14**	**27%**
Strongly Disagree	**0**	**0%**
Total	**51**	**100%**

Question 13. I would not join an online communities and/or perform tasks if they are not linked to my interests even if I get a monetary compensation

Question 13 is another question concerning the opportunistic/extrinsic reward mechanism and incentives. As we already stated, we are deeply interested to this topic because from our analyses, until now appears strong evidence that the direct monetary compensation is the best mechanism to stimulate user involvement in *crowdsourcing* projects. However, we saw that often this mechanism needs to be coupled with other forms of reward and/or incentive to work well. The monetary

compensation alone doesn't suffice to move the majority of the people to join the *crowds*. We wanted to bring experimental data at support of this conclusion in order to finally reach a groundbreaking conclusion at this regard. Moreover, the monetary mechanism requires a good level of financial investments to be exploited by an online community backed by an organization. Thus, we would like to bring more data supporting this hypothesis in order to clear the ground from any doubts. Finally, our first research problem, namely developing a *crowdsourcing* methodology for a typical sentiment engine, moves us to consider other options before looking to the monetary incentive.
Question 13 tries to assess if the users are generally oriented at performing tasks even if they are not interested on them but they can receive an amount of money for their job. In other words, we checked if the money incentive alone is a good option for a *crowdsourcing* community of any kind. We introduced the formula "if they are not linked to my interests" to clear the ground from all the forms of enjoyment-based reward. We didn't take into consideration the prestige-oriented reward because as we saw in the *prescriptive framework* it is never used as main reward mechanism. Looking at Figure 43, it's possible to see that indeed, a very strong share of the testers would not be effectively encouraged at taking part to a *crowdsourcing* platform by just the money. 92% of the total set of testers agrees or strongly agrees with the statement. No tester is neutral or disagrees with the question. Only a small percentage (8%) of testers strongly disagree with the statement. As usually we can exclude a moral bias from the testers, indeed the test was completely anonymous thus the users were not concerned at showing an attitude different from the reality. We can conclude that our hypothesis on the direct monetary compensation is valid. However, we will come back on the opportunistic reward system again, talking of the fourth scenario. This conclusion will be part of the general compound of results that we will put together in the second part of this section.

Figure 43 shows a chart depicting the answer to the question according to a Likert-type scale.

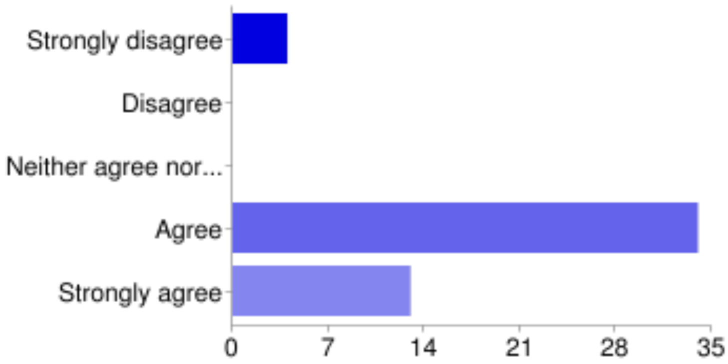

Figure 43: Answers distribution chart for question 13

Table 39 shows the answer to the question 12 with percentages and the number of users choosing each option.

Table 39: Answers distribution for question 13

***Due to rounding, the percentages may not add up to 100%**

Strongly Agree	**13**	**25%**
Agree	**34**	**67%**
Neither Agree or Disagree	**0**	**0%**
Disagree	**0**	**0%**
Strongly Disagree	**4**	**8%**
Total	**51**	**100%**

Question 14. It wasn't a problem for me to provides some additional information while placing new questions

Question 14 enters in the specific problem of assessing which would be an effective *crowdsourcing* methodology for populating the domain knowledge of a sentiment analysis tool. In particular, this question directly assesses if the task that we designed in our base web application, and thus in all its four declinations into scenarios, would be easily accomplished by the users of a future hypothetical *crowdsourcing* community. In other words, we wanted to know to which extent the users are likely to provide useful information for us while taking part to our experimental community. As we stated several times throughout this thesis, one of our main goals, besides modeling the novel domain of *crowdsourcing,* has been understanding how we can effectively develop *crowdsourcing* applications that encourage user participation in an online community. This is an important step linked to our wish of shaping a *crowdsourcing* methodology for a sentiment analysis tool. A great part of our work has been, as consequence, the analysis of the underlying factors that bring people into the *crowds.* The following questions regarding the four scenarios, try to put a light on this topic. Question 14 instead collects experimental data concerning the general architecture of the base web application. It provides data to check that the design choices that we made while developing the base architecture don't create a negative bias. As we discussed in the paragraph 4.2, all the four scenarios collects the information for the sentiment engine in the same way thus, besides their different declinations, they all share a common methodology for accomplishing the most important task of populating the domain knowledge. Moreover, this task comes alongside the other task of making predictions. The later derives from our choice of a *Prediction Market* web applications but it's not directly linked to the *sentiment analysis*, although it is fundamental to us for understanding the *crowdsourcing.* Question 14 checks the testers' sentiment towards the technical methodology for collecting data over the football and fashion trend domains. The problem is if the testers have been annoyed by having to provide new information every time they posted questions with unknown brands or sub-brands. With this question in the survey, we tested if the choice of asking new information in the process of posting new questions in the web application was easily accepted by the users thus not influencing the answers to the remaining part of the survey. Moreover, we wanted to check our hypothesis that we can build a *crowdsourcing* community that provides data to a sentiment analysis tool, as *side-effect* to its processes. An affirmative answer is particularly interesting because it confirms the correctness of our methodology.

Looking at Figure 44, it clearly emerges that all the testers don't express any negative attitude toward the way of collecting the information implemented in our web application. Indeed, 100% of the test group agrees or strongly agrees with the statement in question 14.

We can conclude that our methodology of producing semantic data from *crowdsourcing* was not invasive and well accepted. We will propose again this conclusion in the final compound at the end of this section.

Figure 44 shows a chart depicting the answer to the question according to a Likert-type scale.

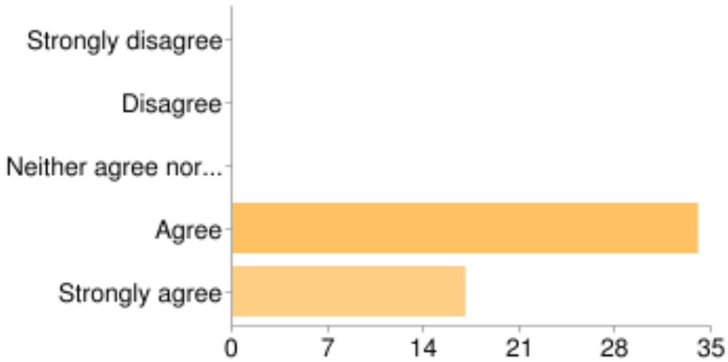

Figure 44: Answers distribution chart for question 14

Table 40 shows the answer to the question 12 with percentages and the number of users choosing each option.

Table 40: Answers distribution for question 14
***Due to rounding, the percentages may not add up to 100%**

Strongly Agree	**17**	**33%**
Agree	**34**	**67%**
Neither Agree or Disagree	**0**	**0%**
Disagree	**0**	**0%**
Strongly Disagree	**0**	**0%**
Total	**51**	**100%**

Question 15. Please order the four scenarios according to how likely would you join each of them

Question 15, 16 and 17 try to assess the effectiveness of the four scenarios and of the respective design choices, in stimulating the testers' involvement. Question 15 starts by checking which scenario is more likely to move the people to join it.

Looking at Figure 45, we can see that the third scenario, the game scenario (see paragraph 4.2), is the most effective while the first scenarios is the less one. Indeed, although the first scenario doesn't appear with the highest frequency in the last position (the fourth scenario does), it is still placed in the majority of the cases in the third and last position. Thus, a first conclusion is that the gaming is a strong incentive for the users when deciding to join or not a *crowdsourcing* community. We already partly discussed it while describing the result for question 9 concerning the fun aspect. We saw that fun is a good incentive and we can argue that gaming is the most effective way of implementing it. It also emerges that the "Sharing of the goals" and the "Sharing of the result" are not effective enough as incentives for our web application. Indeed, they are the main incentives exploited by the first scenario, which we must stress it again, is just the base web architecture without any modifications.

Coming to the second and fourth scenario we see the former is usually placed in good position, just after the third one, while the former performs worse and 26 (51%) of the testers put it in the last position. We can argue that the competition and user ranking system introduced with the second scenario works well in stimulating user engagement. Going to the fourth scenario, we can argue that there was a strong negative bias from the testers towards the betting of real money and that this bias was not counterbalanced by the monetary bonus that the fourth scenario gave as incentive to post new questions on the community. This bias could undermine our conclusions, indeed it seems that the monetary incentive doesn't work well as incentive, contradicting all the evidences we got until now against this fact. Thus, we can already conclude that our design choice of a betting system as monetary incentive for the fourth scenario has undermined, in part, the validity of our conclusions. Of course we couldn't imagine this result while designing the survey and this problem is part of any experimental effort involving real data, which often presents bias and hidden factors, unknown at the moment of the design. Moreover, this result gave us new information for the sake of developing a *crowdsourcing* methodology for a sentiment analysis engine. In fact, we can state that a *Prediction Market crowdsourcing* web application oriented to the betting isn't an effective way to stimulate user involvement. Anyway, the set of all questions concerning the scenario provides a better characterization and help us to understand the real situation. When we will provide our general conclusions taking into account the whole survey and not just a question at the time we will see that indeed the results confirm our original hypothesis. Anyway, we can argue that a system implementing only a financial bonus for posting new useful question would be better than a betting application. We will include these considerations in the final compound.

Figure 45 shows a chart depicting the answer to the question according to the positions of the four scenarios.

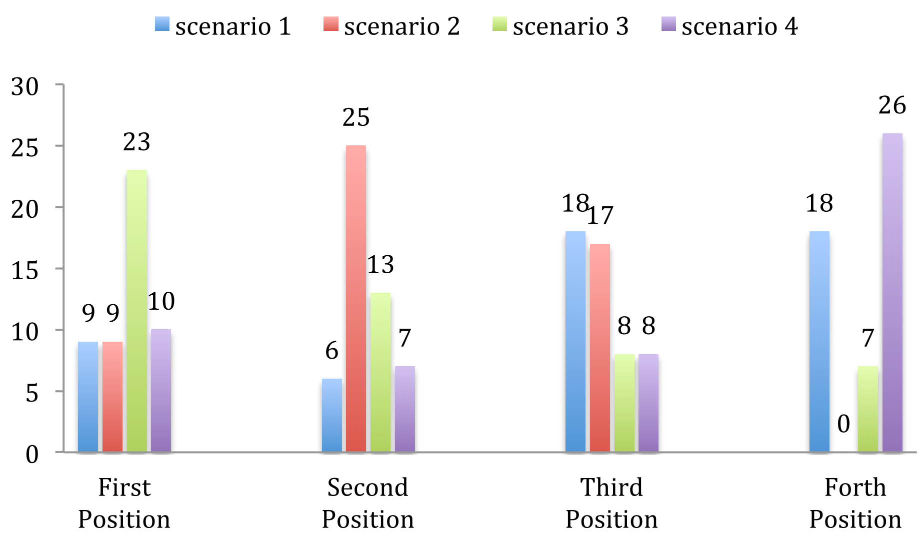

Figure 45: Ranking distribution of the four scenarios in the answers to question 15

<u>*Question 16*</u>. Please order the four scenarios according to their enjoyableness

Question 16 assess the overall sentiment of the testers toward the four scenarios after that played with them for a certain, variable according to the wish of the tester, amount of time. As we would aspect, the results are similar to the one seen for the previous question and we could repeat here almost all of the considerations we made for question 15. The main difference is that the scenario number four performs better, placing in the third position with the highest frequency. This seems to support our idea that the outcome of question 15 was biased by a general negative sentiment toward online betting. Indeed, even if the tester looks less likely join the fourth scenario, he expresses a better appreciation for it than for the first scenario, implicitly confirming our hypothesis. For what concerns the scenario number three and two, the results are almost identical in their meanings to the ones of question 15: the third scenario confirms to be the most appreciated followed by the scenario number two and, finally, we can see that without exploiting any kind of incentives besides the “Sharing of the goals” and “Sharing of the result”, the first scenario fails in gaining the appreciation of the test group.

Figure 46 shows a chart depicting the answer to the question according to the positions of the four scenarios.

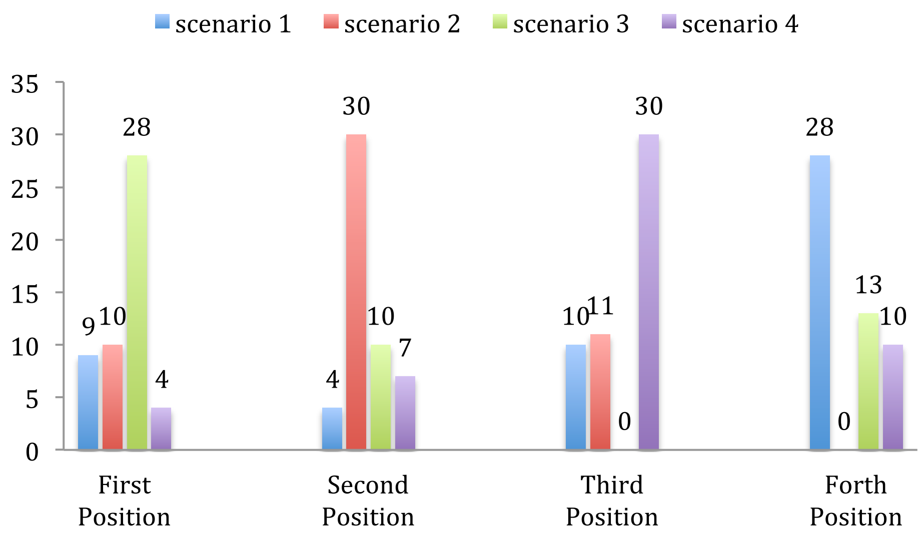

Figure 46: Ranking distribution of the four scenarios in the answers to question 16

<u>Question 17</u>. Please order the four scenarios according to the number of questions that you are likely to answer or post in each of them

Question 17 looks at how the four scenarios encourage users' contribution. Indeed, we want to check which scenario and the respective design choices, brings more people in the *crowd* at posting and answering more questions. In other words, which scenario would have the most vital community.

Looking at Figure 47, it is possible to see that scenario number two perform better overall, placing in the first position with the highest frequency, thus overcoming even the third scenario, that until now has presented the best performance. Anyway, the third scenario seems to work well as well. We can argue that the user ranking and user statistics implemented in the second scenario are strong incentive for contributions. Moreover, the competition as incentive, which is exploited both by the second and third scenario, performs well overall. The first scenario, confirming what we have seen until now, places in the last position in the majority of the cases. Finally the fourth scenario again comes often as third preference. If this result didn't surprise us in question 16, which concerns the enjoyableness of the scenarios and not the contribution level, here instead tells us that the monetary bonus isn't a strong enough incentive both for answering to the questions and for posting new questions. Again, we think that we can trace back these findings to the bias against betting that seems to transpire from all the outcomes of question 15, 16 and 17.

Figure 47 shows a chart depicting the answer to the question according to the positions of the four scenarios.

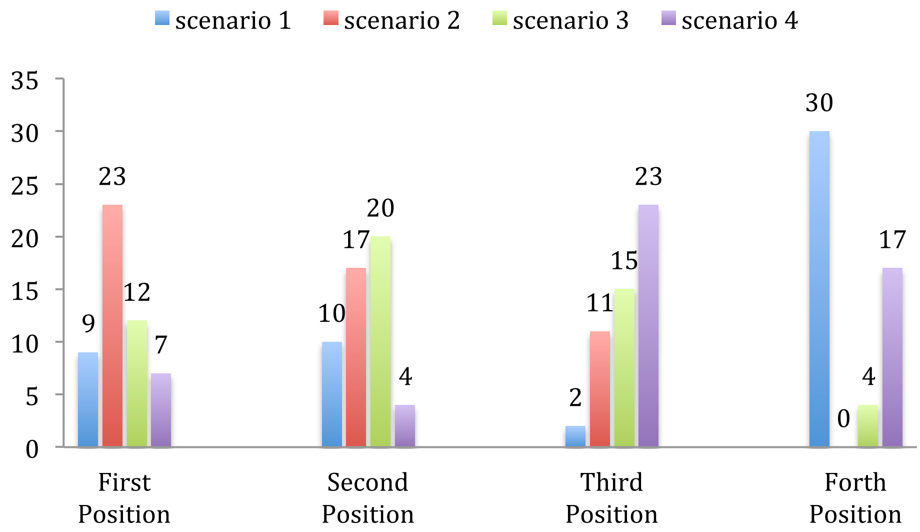

Figure 47: Ranking distribution of the four scenarios in the answers to question 17

<u>*Question 18*</u>. Which range of compensation would you judge fair for providing us useful information by placing new questions?

Question 18 refers to the Remuneration dimension of our *descriptive framework*. It tries to assess which level of remuneration would be judge fair by the testers, for providing us information for the domain knowledge by placing new question. It is directly linked to the fourth scenario that is the only scenario with an opportunistic reward system. Moreover, this scenario, as we already discussed, stimulate the users to post new questions by giving them a monetary bonus for reaching a threshold of ten useful new questions posted. In the prospective of building a future working implementation of a *crowdsourcing* methodology for a sentiment analysis engine, we already asked to the testers which level of remuneration they would like for this bonus. Moreover, this gave us more information concerning the *crowdsourcing* phenomenon itself. Indeed, the data coming from this question can be used as aid by the developers of new *crowdsourcing* applications that exploit a structure similar to our test cases, according to the *descriptive framework*. Of course, this remark is valid also for the majority of the data produced by this research, and, in fact, it has always been one of our driving goals.

Looking to the Figure 48, we can see the majority of the testers (73%) didn't express any preferences concerning the level of remuneration they would judge fair for their job, accepting instead any amount of money. This is the most important result for what concerns the designing of new *crowdsourcing* applications: most of the time the monetary compensation works well at encouraging user involvement, independently from its level. This is particularly true for *crowdsourcing* communities that propose small and simple tasks and we already had a preview of this while discussing of Amazon Mechanical Turk (see the *descriptive framework*). Moreover, in the *prescriptive framework* we avoided to analyze the Remuneration dimension, but anyway we still can bring some new findings also in this topic thanks to our experiment and survey. Surely, these findings are not always generalizable. At this regard, consider for instance the case of InnoCentive (we already discussed it several times in this thesis): the tasks that it proposes are really complex tasks thus the monetary incentive has to be proportionally increased and it's amount is a determining factor for stimulating the user involvement.

Figure 48 shows a chart depicting the answer to the question.

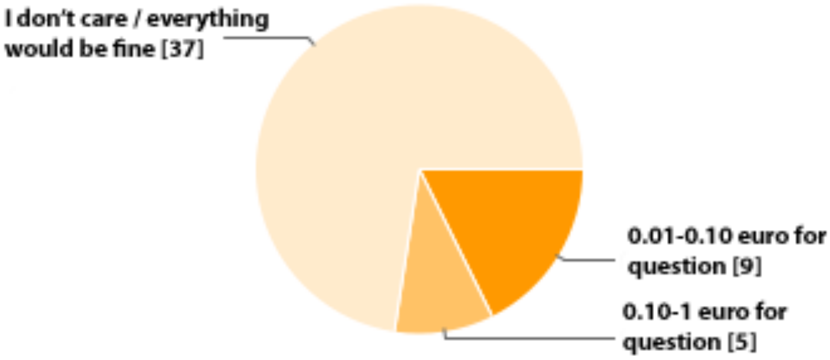

Figure 48: Answers distribution chart for question 18

Table 41 shows the answer to the question with percentages and the number of users choosing each option.

Table 41: Answers distribution for question 18
***Due to rounding, the percentages may not add up to 100%**

0.01-0.10 euro for question	**9**	**18%**
0.10-1 euro for question	**5**	**10%**
I don't care / everything would be fine	**37**	**73%**
Total	**51**	**100%**

Question 19. Instead of money would you accepted some other form of compensation (for instance few free analysis by the sentiment engine for your purposes) for providing us useful information by placing new questions?

We saw several times how the monetary incentive can be a strong aid in stimulating user involvement in *crowdsourcing* projects. However, we were interested in finding alternatives to the monetary incentive for the sake of developing a *crowdsourcing* methodology. In particular, we wanted to discover if there could a market for the sentiment analyses produced by a specific sentiment analysis tool and if these analyses could be used as a form of “payment” for the contributions given by the *crowd.* Moreover, we considered that involving the user-base in the usage of the sentiment analysis tool, would increase the effectiveness of the “Sharing the goals” and “Sharing the results” incentives. Thus, we investigated this topic by asking the testers if they would accept some other form of compensation instead of money (and in particular free sentiment analyses performed by a sentiment analysis engine), for providing to the *crowdsourcing* system useful information by placing new questions. Unfortunately, it is possible to see from Figure 49 that this is not the case. The greatest majority of the test users refuse this possibility. Anyway, we can see that there is still a good share of the testers (33%) that would accept this alternative in any case. We can make two considerations looking to the data. First that it is possible to propose a combined payment composed both of a monetary incentive and of another form of compensation involving the analyses produced by a sentiment analysis tool. Moreover, that the strong negative answer may come from the fact the majority of the testers were not able to understand the topic of *sentiment analysis* because they didn’t know it before. Thus, further development may open new possibilities.

Figure 49 shows a chart depicting the answer to the question.

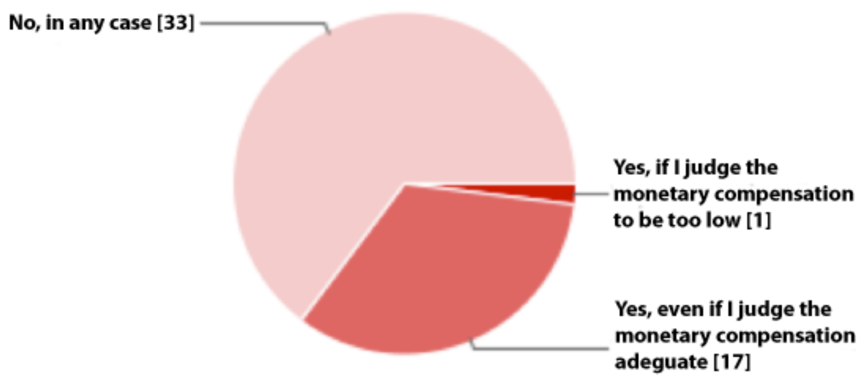

Figure 49: Answers distribution chart for question 19

Table 42 shows the answer to the question with percentages and the number of users choosing each option.

Table 42: Answers distribution for question 19
***Due to rounding, the percentages may not add up to 100%**

Yes, if I judge the monetary compensation low	**1**	**2%**
Yes, even if I judge the monetary compensation adequate	**17**	**33%**
No, in any case	**33**	**65%**
Total	**51**	**100%**

Question 20. Please order the following incentives according to how much they encourage you to correctly answer to the questions

Question 20, 21, 22 and 23 refers to the Data Quality Mechanism dimension. At this point, we can see how we tried to shape the survey in order to cover all the main factors that are involved in *crowdsourcing.* The Data Quality Mechanism is the last dimension proposed in our *descriptive framework* and it will be analyzed it.
As we stated already, the *crowdsourcing* web application in all its four declinations doesn't provide any kind of data quality mechanism for the sake of ensuring a good quality of the information reaching the sentiment analysis tool. However, it exploits some data quality mechanisms for the other minor task of making predications. Thus, we can study the results coming from this approach in order to either test the validity of our *descriptive framework* and to gain new knowledge that would be of great interest in future researches involving the quality of the data for a specific sentiment analysis tool.
In paragraph 4.3 we described the Data Quality Mechanism dimension of the four scenarios and we saw that all of them exploit the Group Consensus [Average] mechanism in the same way. However, we also stated that the secondo and third scenarios harness also the Competition mechanism, while the fourth adds a Rewards

Accuracy mechanism. Moreover, the second and third scenarios implement the competition in a different way: scenario number two introduces the competition through user ranking and user statistics; scenario number three instead uses the gaming and the challenges between friends.

At this regard, it's interesting to us a comparative analysis of these three ways to the data quality, to discover which one is the most effective. Thus, in question 20, we asked to the users to rank the scenarios according to how much they encourage them to correctly answering the predictions.

Looking at Figure 50, we can see that the User Ranking in scenario two is regarded as the best mechanism by the testers. Gaming also performs well, being in second position with the highest frequency. Finally, remarkably for us, the monetary betting in the fourth scenario isn't regarded as being a good incentive to provide correct answers. Thus, in our experiment, the reward of the accuracy performs worse than the other two. This is a counter-argument to Shaw, Horton & Chen (2011), which proposed the reward of the best *crowdsourcing* productions as the foremost mechanisms for fostering data quality.

Figure 50 shows a chart depicting the answer to the question according to the positions of the four scenarios.

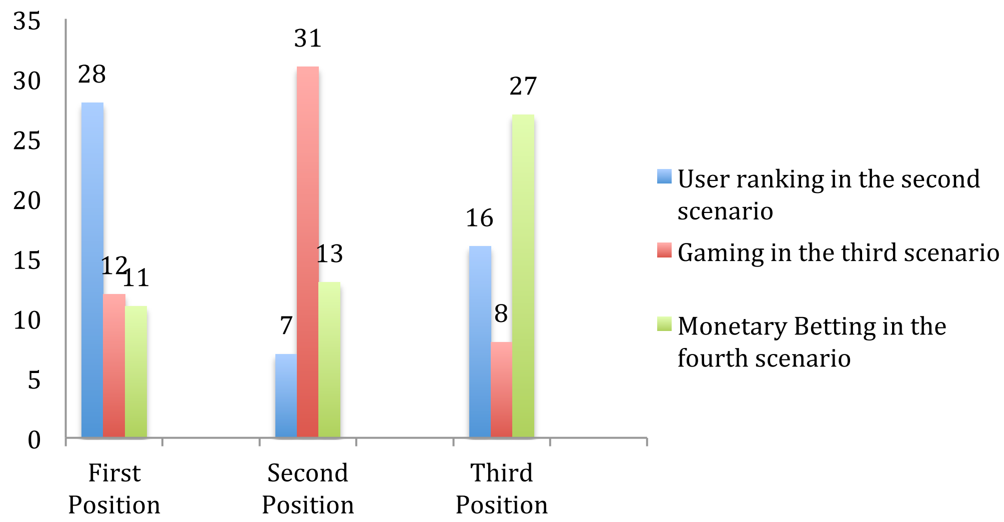

Figure 50: Ranking distribution of the four scenarios in the answers to question 17

<u>*Question 21, Question 22 and Question 23*</u>.

I would try to correctly answer the questions at my best even in the first scenario

Would you consciously provide false/incorrect information in the fourth scenario just to benefit of the compensation?

Would you consciously provide false/incorrect information in the fourth scenario just to benefit of the compensation knowing that the website uses automatic and/or manual mechanisms to avoid cheating?

We will conclude our presentation of the data coming from the survey, analyzing the question 21, 22 and 23. We analyze these questions together because they try to assess first if a data quality mechanism isn't unavoidable to ensure a good quality of the answers and second if using the a monetary bonus would not undermine the methodology we chose for *crowdsourcing* data for the domain knowledge, producing low quality outcomes.
In Figure 51, it is possible to see that almost 100% of the test group agrees with the statement of question 21. Thus, they would try to provide correct answer also without any additive data quality incentive such as Competition or Reward Accuracy. From this result we can state that Group Consensus [averaging] should already perform well in producing good answers.
Figure 52 and 53 show the answers to the last two questions. It results, that the people in the test group are not strongly oriented at cheating just for earning monetary bonus. Moreover, question 23 tries to check if a data quality mechanism of type Surveillance (see *descriptive framework*) would foster a better data quality overall. Indeed, it does, rising the number of people that would surely not cheat from the 65% to the 90% of the test base.
It is possible to argue to which extent the people answered sincerely to these last three questions. The questionnaire was completely anonymous thus we have no evidence of lying or giving false answers and we think the data are reliable. However, for what concerns the data quality a bigger and more heterogeneous test

group may have produced slightly different results.

Figure 51 shows a chart depicting the answer to the question 21 according to a Likert-type scale.

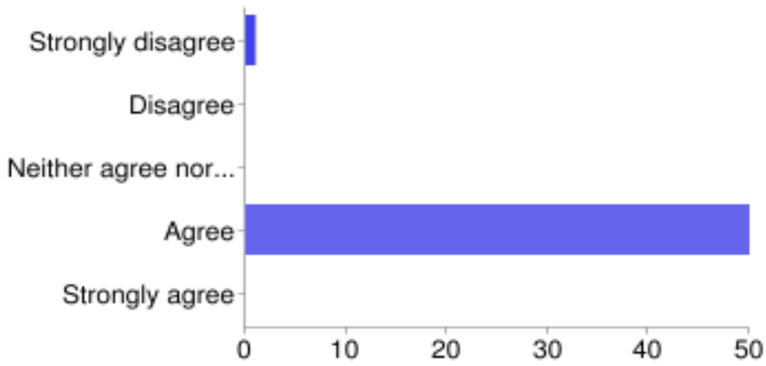

Figure 51: Answers distribution chart for question 21

Table 43 shows the answer to the question 21 with percentages and the number of users choosing each option.

Table 43: Answers distribution for question 21

***Due to rounding, the percentages may not add up to 100%**

Strongly Agree	**0**	**0%**
Agree	**50**	**98%**
Neither Agree or Disagree	**0**	**0%**
Disagree	**0**	**0%**
Strongly Disagree	**1**	**2%**
Total	**51**	**100%**

Figure 52 shows a chart depicting the answer to the question 22.

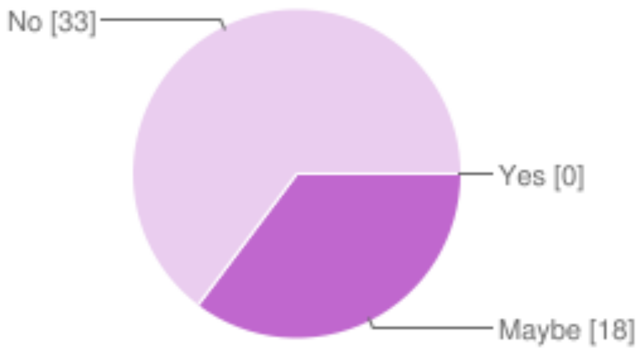

Figure 52: Answers distribution chart for question 22

Table 44 shows the answer to the question 22 with percentages and the number of users choosing each option.

Table 44: Answers distribution for question 22

***Due to rounding, the percentages may not add up to 100%**

Yes	**0**	**0%**
Maybe	**18**	**35%**
No	**33**	**65%**
Total	**51**	**100%**

Figure 53 shows a chart depicting the answer to the question 23.

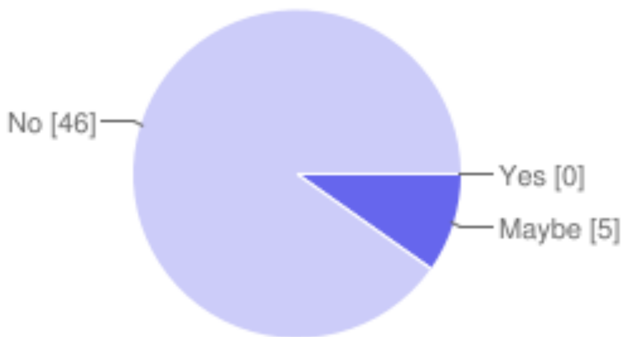

Figure 53: Answers distribution chart for question 23

Table 45 shows the answer to the question 23 with percentages and the number of users choosing each option.

Table 45: Answers distribution for question 23
***Due to rounding, the percentages may not add up to 100%**

Yes	**0**	**0%**
Maybe	**5**	**10%**
No	**46**	**90%**
Total	**51**	**100%**

4.8 Analysis of the results

We will now provide our general conclusions concerning the experiment that we carried out and that can be drawn from the answers to the survey. These conclusions are aimed at proposing a final *crowdsourcing* model for an application capable of enhancing a generic sentiment analysis tool with a methodology to crowd-source the data in order to populate its domain knowledge. This model maximizes the user satisfaction variable as represented by the opinions of the test group. Thus, with this last paragraph we conclude our research work, answering to the first research question that we posed at the beginning of our effort.
While developing the survey we tried to cover all the aspects that we regarded as meaningful and relevant for our purposes, i.e. to further assess the validity of the modeling tools (the *descriptive* and *prescriptive frameworks*), in addition to all the analyses focused on the empirical dataset that we made previously, and to finally developed a groundbreaking *crowdsourcing* methodology for a sentiment analysis tool with a focus on the user satisfaction. We still have to conclude with a complete and coherent description of our proposal for a community achieving the later task. We will try to offer a schematic presentation of this community, of course again exploiting our *descriptive framework*. The overall analysis that we will draw here takes into account our main goal of understanding how to foster user participation to

online *crowdsourcing* communities and at the same time tries to put together all the pieces that we presented throughout the thesis.

The community that we want to build for the sentiment engine is a knowledge community, i.e. a community with the following values for the Categorization dimension of the *descriptive framework*:

- “Knowledge Sharing”
- “Collective Knowledge”
- “Knowledge Acquisition”

We already discussed this in the paragraph 4.2 when we presented the common dimensions of the four scenarios. Thus, we will read the answers provided by the testers at the light of this consideration.
From question 7 (“I would share my knowledge and spend my time contributing to online communities for free if I share the goals”) we can derive the importance of “Sharing of the goals” incentive. Clearly stating the goals of a *crowdsourcing* application greatly helps in attracting the users into the *crowd* and in fostering their participation and satisfaction. Thus, this could be an affordable option for a future implementation of a *sentiment analysis crowdsourcing* methodology. From question 6 (“I think the size of the community matters when choosing if joining or not an online community”) clearly emerges that a *crowdsourcing* community should tries to opt for all the incentives that can increase the size of its user community. This finding offers another good support to our decision of focusing on the study of the stimuli of user involvement. The answers to question 8 (“I would share my knowledge and spend my time contributing to online communities for free if I can access to the results and they are useful to me”) tell us that “Sharing of the results” should be coupled with “Sharing of the goals”. Indeed, we saw in the *prescriptive framework* context that this practice is exploited by the majority of knowledge communities. A *crowdsourcing* community for a sentiment analysis tool should not make an exception to this. The problem is which kind of results the sentiment

analysis tool's *crowdsourcing* community should share. For what we saw in the analysis of question 19 ("Instead of money would you accepted some other form of compensation (for instance few free analysis by the sentiment engine for your purposes) for providing us useful information by placing new questions?") the testers are not interested in the *sentiment analysis.* It results that the proposal for a community should consider other form of knowledge sharing. We think that the idea developed in our base web application is going in the correct direction at this regard. Our base application proposes a system where the *crowd* accomplishes several tasks: the people make predictions answering to the questions or provide useful information about new brands and sub-brands while posting new questions. Thus, the *crowdsourcing* of semantic data for the sentiment analysis tool comes as *side-effect* of another *crowdsourcing* task, namely the placing of new questions. In this way we provide to the community a database of crowd-sourced data (the predictions), which can be seen as a value for the users in the community but it is not vital for sentiment engine's purposes. The database of predictions is the result that is shared, and is the implementation of the "Sharing of the results" incentive. Moreover, this is a way of dribbling the problem emerged from the answers to question 19, i.e. the users seem not willing to engage directly with *sentiment analysis*. A *crowdsourcing* methodology for a sentiment analysis tool cannot avoid from these considerations and thus we think it should always take the form of a web application aimed at some other purposes, producing as *side-effect,* the data for the tool. This conclusion is also backed by the results of question 14 ("It wasn't a problem for me to provides some additional information while placing new questions"), in which we saw that the testers were not annoyed by the necessity to provide additional information while posting new questions. The only shortcoming is that this approach requires developing not intrusive mechanisms to collect semantic data. We made a first step into this direction proposing the solution represented by *Prediction Markets* web application model but of course many other ways can be studied and tried.

This finding offers another opportunity. We saw from the answers to question 9 ("I would share my knowledge and spend my time contributing to online communities for free if I get fun"), 15 ("Please order the four scenarios according to how likely would you join each of them"), 16 ("Please order the four scenarios according to their enjoyableness") and 17 ("Please order the four scenarios according to the number of questions that you are likely to answer or post in each of them"), that the fun side of participating to a *crowdsourcing* community should be properly encouraged. We saw how the gaming task in the third scenario gained good consensus among the testers. Thus, the design of our *crowdsourcing* community should follow this track: proposing tasks that can be regarded as funny to be accomplished. This cannot be easily achieved by just asking to the *crowd* to provide their knowledge concerning brands or sub-brands of interest for the sentiment analysis tool, but should be managed by designing collateral tasks. Moreover, we saw from the answers to questions 20 ("Please order the following incentives according to how much they encourage you to correctly answer to the questions") and 17, that competition holds several roles in our model. Competition can be effectively exploited to increase the quality of the crowd-sourced data and it is also an incentive to the enjoyment-based reward mechanism, in particular to its component of fun.

However, we saw from question 9, that the fun of performing a task alone, often doesn't suffice at stimulating the user to accomplish it. The drawback is even bigger if our purpose is encouraging the users at performing many times the same task. And indeed, the tool needs a lot of information to keep updated its knowledge domain. Thus, we explored the opportunity of opportunistic rewards coupled with monetary incentive. At this regard, we got contradictory results. From one hand, the users seems to effectively support our thesis, extracted from the experience of the empirical dataset and the *prescriptive framework*, that the direct monetary compensation is a strong incentive to boost user contribution and satisfaction. In particular, this turns out from the answers to question 10 ("I am likely to contribute much more to an online community if I can receive any kind of monetary compensation even if I share the goals and/or the results"). At the other hand, when the questions directly involve our implementation of an opportunistic scheme (fourth scenario), the testers seem to dislike the approach (question 15,16,17 and 20). We argued that the betting model that we designed was not well regarded by the majority of the testers and this partly undermined the results of our survey concerning the direct monetary compensation and monetary incentive. However, we were still able to get worthy data at this regard. In particular, we saw that the individuals in the *crowds* are not deeply interested to the level of the money that they can get from by accomplishing the jobs. We linked this result to the consideration we made in the *descriptive framework* and we concluded that for simple tasks the level of compensation is not a determining factor. Thus, we were even able to deepen our analyses of the Remuneration dimension.

The answers to question 11(“I am likely to contribute to online communities only if performing simple (tag images, assign values, etc.) tasks”) and 12 (“I am likely to contribute to online communities also by performing complex (write long essays, translation of text, etc.) tasks if I have the skills”) tell us that simple tasks have much greater chances to attract contributors. This can be coupled with the results regarding the third scenario, i.e. the gaming one, and with the answers to question 15, 16, 17 and 20 that directly compare the four scenarios. It emerges that scenario two and scenario three overall perform better then the others in the majority of the situations, both regarding the user participation stimulation and the data quality of the *crowdsourcing* outcomes. We can conclude that simple tasks coupled with user ranking (exploiting competition) and gaming (another way of exploiting competition and fun) are the most effective solutions. At the end, the former solution is overall the best one and practically, the easiest to be implemented.

Finally, questions 20, 21 (“I would try to correctly answer the questions at my best even in the first scenario“), 22 (“Would you consciously provide false/incorrect information in the fourth scenario just to benefit of the compensation?”) and 23 (“Would you consciously provide false/incorrect information in the fourth scenario just to benefit of the compensation knowing that the website uses automatic and/or manual mechanisms to avoid cheating?”), gave us some material to reflex on the data quality mechanism. We can state that the users are not willing to cheat while participating to a *crowdsourcing* community, even if a direct monetary compensation is offered. Of course, we can argue that to an increasing level of compensation corresponds a greater level of cheating. But as far as the amount of money involved is low, this doesn't encourage bad behaviors. Following the answers coming from these questions, it seems that just aggregating the crowd-sourced data could be a good mechanism of *crowdsourcing*. Competition itself, that as we saw can be easily introduced in several ways, is greatly helping in increasing the quality of the crowd-sourced data. Moreover, the second scenario looks better at ensuring data quality than the third one. Again it looks that our *crowdsourcing* methodology should not rely on money to foster better performance, participation and overall satisfaction, of the contributors. Indeed, the testers were not stimulated at providing better answers by this incentive. Finally surveillance seems to work well to prevent the cheating.

Finally, we can conclude that a mixed combinations of all the factors that we discussed until now should be the main route to follow while developing a *crowdsourcing* methodology for any sentiment analysis tool that has the architecture design we assumed in Chapter 2 and should achieve the goal of attracting the highest amount of users and boosting their satisfaction. Thus, our final model puts all these considerations together in a schematic way built according to our *descriptive framework*. In particular, for what concerns the reward mechanism we propose a mix of opportunistic and enjoyment-based rewards. In our case, the traditional distinction between main and minor reward falls and we put them at the same level, keeping the prestige-oriented reward as minor reward. Thus, our situation is quite complex, presenting three reward systems.

The next table provides a list of all the dimensions with the values that we chose for them. Some of them descend directly from the description of the base web application discussed in paragraph 4.2.

Table 46: Our model of a *crowdsourcing* methodology for a sentiment analysis tool according to the *descriptive framework*

Dimension	Value(s)
Categorization	Collective Knowledge, Knowledge Sharing, Knowledge Acquisition
Crowdsourcing Type	Integrative
Required Knowledge	Low
Community Size	Big (as big as possible)
User Type	Any
Task Type	Simple
Rewards (Main and Minor)	Main Reward → <u>*Enjoyment-based*</u> • Fun • Curiosity and desire to test if it work • Desire to do something different from your work • Desire to express yourself • Values and Ideology o Volunteerism and desire to support a cause of the project o Reciprocity, exchange and mutual help → Opportunistic • Direct monetary compensation for providing useful new information for the domain knowledge • (Alternative/In addiction) Indirect compensation by offering the services powered by the sentiment analysis engine Minor Reward → Prestige-oriented mechanism • Increasing online reputation and recognition

Incentive	→ *<u>Connected to enjoyment-based reward</u>* • Sharing of the result because the user can access to all the predictions coming from the community • Competition → *<u>Connected to prestige-oriented reward</u>* • User ranking and statistics of the best users • Position inside the community → *<u>Connected opportunistic reward</u>* • Money (direct)
Remuneration	Low
Data Quality Mechanism	Group Evaluation [Averaging] Competition Reward Accuracy

4.9 Testing of the methodology

In the previous paragraph we used the data coming from the experiment to design our final proposal of a *crowdsourcing* methodology for a sentiment analysis tool. In particular, we focused on the specific subset of sentiment analysis tools of *Web reputation* analysis, but our results are easily transferable to the general case. This methodology should maximize our quality metric, i.e. the user satisfaction, because it has been built according to the data coming from the survey where we assessed this specific variable using a test group of users. Indeed, the final model is composed of sub-components that derive from the testing of several different solutions, the ones implemented in the four scenarios, with respect to the user satisfaction and participation expressed by the test group. Each of the sub-components (the rewards mechanism, the type of task, the incentives, etc.) of the final architecture has been selected because it was the most effective according to our analyses. Thus, the overall final methodology is the one that maximizes our quality metric and results to be the most effective.

Chapter 5

Conclusions and future developments

We started our research effort facing the problem yielded by the domain knowledge component of a typical sentiment analysis engine. Indeed, we saw that there was the need of an automated methodology to populate the knowledge database used by these tools, for performing semantic analyses. At this scope, we looked at a novel research domain, namely the *crowdsourcing.* We thought that it would have been possible to develop an automatic *crowdsourcing* methodology for populating this knowledge domain and we developed a research methodology for assessing the best *crowdsourcing* model for this goal, according to the user satisfaction variable.

However, we saw from the beginning of our research that the *crowdsourcing* phenomenon was still not well covered from a scientific and academic point of view and that we need to perform a deep study of it in order to design an effective way of exploiting the *crowdsourcing* for our purposes.

In the first part of this research we fulfilled this lack providing a complete set of modeling tools for *crowdsourcing* applications. We named them the *descriptive* and *prescriptive framework.* The *descriptive framework* provides ten analysis dimensions in order to describe and model a *crowdsourcing* platform. These dimensions emerge from both a strong theoretical effort that we made reviewing the academic literature in several different contexts, and from experimental and empirical effort. We tested the validity of the *descriptive framework* in two ways: by modeling the *crowdsourcing* applications in the empirical dataset, and by applying it to develop an experimental *crowdsourcing* methodology for a generic sentiment analysis engine. We can conclude the validity and coherence of this framework. Moreover, we hope that its scope will not be limited to this research but that instead it will offer a novel instrument in the field of *crowdsourcing*. This has always been one of our primary goals and it is the main reason why we put such a strong effort and care in its design and developing.

The *prescriptive framework* should offer a further view on the *crowdsourcing* phenomenon. It has been developed both as part of the testing of the *descriptive framework* and for aiding us in shaping the experimental phase by transmitting the expertise accumulated by a wide sample of real *crowdsourcing* applications currently active on the market. The results of the survey that we conducted as part of the experimental effort, confirmed the majority of the findings that emerge from the analyses of the relationships among dimensions contained in the *prescriptive framework*. We are satisfied by this outcome and we hope that it might also offer a precious set of guidelines and best practices for the developers of *crowdsourcing* applications.

In the Chapter 4, we outlined a *crowdsourcing* methodology for a sentiment analysis tool, emerging from the experiment that we made with a test group composed of 51 people. The presented *crowdsourcing* methodology maximizes the quality indicator that we chose for our research: the user satisfaction. Indeed, this *crowdsourcing* methodology exploits the most effective ways to stimulate the user participation to the *crowdsourcing* platform and at the same time provides a great amount of crowd-sourced data with a good quality.

Eventually we can state that one of our primary goals has been to study the links between the technologies aspects and the human factors. In particular the question has been which are the technologies that more encourage the user participation and satisfaction in *crowdsourcing* projects and how to best design these technologies. The *crowdsourcing* methodology that we outlined here answers to this question.

This research has been a first step towards the final solution of the problem. Indeed, our *crowdsourcing* methodology gives many ideas and directions to effectively reach a real implementation of a working *crowdsourcing* system tapping the *crowds* in order to maintain and enhance the domain knowledge of any sentiment analysis tool that follow the architecture we presented in Chapter 2. The problem of automatically modeling a knowledge domain is still an open challenge. The research methodology that we followed has showed to be a valid for the scope of proposing a solution to this problem.

We hope that our effort can put a new light on this topic and that the *crowdsourcing* would be regarded in future as a possible solution.

With this research, we didn't exhaust the whole topic of *crowdsourcing* applied to *sentiment analysis* because our resources compared to the magnitude of the job prevented us from moving further, for the moment. Anyway, there are still many open questions and problems that arise and many of them already came out while working on this research. We can list some of these open challenges from which we would like to have an answer sooner or later:

1) Increase the number and the type of information that can be collected for the sentiment analysis tool by the *crowdsourcing* methodology. Indeed here we provided a methodology just for the concepts of brand and sub-brand

2) Develop and apply quantitative measures to assess the quality of the crowd-sourced information transferred to the sentiment analysis tool

3) Following the direction shaped by the previous point, develop new mechanism to foster the quality of crowd-sourced data. This should involve also a system to detect and resolve the ambiguities that the *crowdsourcing* of information naturally brings

4) Develop a way to get information from the *crowds* to better model the relationships existing among the semantic concepts. This should involve also

the study of novel automatic techniques to build hierarchies of concepts, composing the domain knowledge.

and many more.

The most pressing challenge for us is probably developing a real implementation of our *crowdsourcing* methodology and many ideas have already been studied at this regard. In particular, it would be interesting to develop a Facebook application following the final recommendations that we outlined at the end of our experiment. Facebook is a social network with many millions of users and it would be the ideal environment to test on the field our conclusions. Moreover, it would give us a large enough community to develop several metrics to evaluate in a precise and quantitative fashion, the findings that we already exposed in this research. For sure, we think that this would be a stimulating and challenging next research step.

APPENDIX A

- Website URL, Location, Description, Categorization

#	NAME	WEBSITE URL	LOCATION	DESCRIPTION	*CATEGORIZATION*
1	**Pledgemusic**	www.pledgemusic.com	United Kingdom	Music Band Funding	Crowdfunding
2	**global for me**	www.globalfm.com	United Kingdom	Community-Funded Reporting, Journalism	Crowdfunding
3	**IMDB**	www.imdb.com	U.S.A.	Product Recommendation, Movies, Shows	Collective Knowledge, Knowledge Sharing
4	**Delicious**	www.delicious.com	U.S.A.	Social Bookmarking	Collective Knowledge, Knowledge Sharing
5	**Amazon Mechanical Turk**	www.mturk.com	U.S.A.	Task Allocating, Job Marketplace	Cloud Labor, Problem Solving
6	**Wikipedia**	www.wikipedia.org	U.S.A.	Collective Encyclopedia	Collective Knowledge, Knowledge Sharing
7	**Foursquare**	www.foursquare.com	U.S.A.	Location-Based Social Network, Geoinformation	Collective Knowledge, Knowledge Sharing
8	**Article One Partners**	www.articleonepartners.com	U.S.A.	Prior Art Search Community	Collective Knowledge, Knowledge Acquisition
9	**crowdSPRING**	www.crowdspring.com	U.S.A.	Creativity, Design, Logo Design, Graphic Design, Brand Design	Cloud Labor, Collective Creativity
10	**clickworker**	www.clickworker.com	Germany	Task Allocating, Job Marketplace	Cloud Labor, Problem Solving
11	**Waze**	www.waze.com	U.S.A.	Community-Driven GPS Application, Geoinformation	Collective Knowledge, Knowledge Sharing
12	**Urtak**	www.urtak.com	U.S.A.	Collaborative Public Opinion, Q&A	Collective Knowledge, Knowledge Acquisition
13	**CloudCrowd**	www.cloudcrowd.com	U.S.A.	Task Allocating, Job Marketplace	Cloud Labor, Problem Solving
14	**CrowdFlower**	www.crowdflower.com	U.S.A.	Task Allocating, Job Marketplace, Content Generation	Cloud Labor, Problem Solving
15	**Listener Driven Radio**	www.ldrradio.com	U.S.A.	Community-Driven Radio Station	Collective Creativity, Knowledge Acquisition
16	**Kickstarter**	www.kickstarter.com	U.S.A.	Creativity Projects Funding	Crowdfunding
17	**Amazon.com**	www.amazon.com	U.S.A.	Product Recommendation, Books, Goods, Marketplace	Collective Knowledge, Knowledge Sharing
18	**Funding4Learning**	www.funding4learning.com	Spain	Education Funding	Crowdfunding

#	NAME	WEBSITE URL	LOCATION	DESCRIPTION	*CATEGORIZATION*
19	**RocketHub**	www.rockethub.com	U.S.A.	Creativity Projects Funding	Crowdfunding
20	**Grow Venture Community**	www.growvc.com	Hong Kong	Technology Start-Ups Funding	Crowdfunding
21	**Sponsume**	www.sponsume.com	United Kingdom	Creativity Projects Funding	Crowdfunding
22	**Stardust@home**	stardustathome.ssl.berkeley.edu	U.S.A.	Search Images, Science, Astronomy	Cloud Labor, Problem Solving
23	**Zooniverse**	www.zooniverse.org	United Kingdom	Search Images, Science, Astronomy	Cloud Labor, Problem Solving
24	**Distributed Proofreaders**	www.pgdp.net	N.A.	Proofreading E-Texts, OCR, Error Search	Cloud Labor, Problem Solving
25	**wikimapia**	www.wikimapia.org	Russia	Map, Geoinformation	Collective Knowledge, Knowledge Sharing
26	**google maps**	www.google.com/maps	U.S.A.	Map, Geoinformation	Collective Knowledge, Knowledge Sharing
27	**Facebook Translation**	www.facebook.com/translations	U.S.A.	Text Translation	Cloud Labor, Problem Solving
28	**spot.us**	www.spot.us	U.S.A.	Community-Funded Reporting, Journalism	Crowdfunding
29	**Ebay**	www.ebay.com	U.S.A.	User Recommendation, Goods, Online Auction, Marketplace	Collective Knowledge, Knowledge Sharing
30	**buyacredit**	www.buyacredit.com	United Kingdom	Community-Funded Movies	Crowdfunding
31	**Zazzle T-shirt**	www.zazzle.co.uk	U.S.A.	Creativity, Product Design, Clothes	Collective Creativity, Knowledge Acquisition
32	**Sellaband**	www.sellaband.com	Germany	Music Band Funding	Crowdfunding
33	**threadless**	www.threadless.com	U.S.A.	Creativity, Product Design, Clothes	Collective Creativity, Knowledge Acquisition
34	**OpenStreetMap**	www.openstreetmap.org	United Kingdom	Map, Geoinformation	Collective Knowledge, Knowledge Sharing
35	**Innocentive**	www.innocentive.com	U.S.A.	High Level Problem Solving, Job Marketplace	Open Innovation, Cloud Labor, Problem Solving
36	**LinkedIn Answers**	www.linkedin.com/answers	U.S.A.	Q&A	Collective Knowledge, Knowledge Sharing
37	**Answerbag**	www.answerbag.com	U.S.A.	Q&A	Collective Knowledge, Knowledge Sharing
38	**Free Knowledge Exchange**	3form.org	N.A.	Collective Intelligence Algorithm, Q&A	Collective Knowledge, Knowledge Sharing

#	NAME	WEBSITE URL	LOCATION	DESCRIPTION	*CATEGORIZATION*
39	**Google Answers**	www.google.com/answers	U.S.A.	Q&A	Collective Knowledge, Knowledge Sharing
40	**WikiAnswers**	www.wiki.answers.com	Israel	Q&A	Collective Knowledge, Knowledge Sharing
41	**FunAdvice.com**	www.funadvice.com	U.S.A.	Q&A	Collective Knowledge, Knowledge Sharing
42	**Yahoo! Answers**	www.yahoo.com/answers	U.S.A.	Q&A	Collective Knowledge, Knowledge Sharing
43	**Digg**	www.digg.com	U.S.A.	Community-Driven News, Website Sharing	Collective Knowledge, Knowledge Sharing
44	**Reddit**	www.reddit.com	U.S.A.	Community-Driven News, Website Sharing	Collective Knowledge, Knowledge Sharing
45	**Phylo**	phylo.cs.mcgill.ca	U.S.A.	ESP Online Game, Genetics	Collective Knowledge, Knowledge Acquisition
46	**Foldit**	fold.it	U.S.A.	ESP Online Game, Genetics	Collective Knowledge, Knowledge Acquisition
47	**gwap**	www.gwap.com	U.S.A.	ESP Online Game, Science, Image Recognition	Collective Knowledge, Knowledge Acquisition
48	**BlueServo**	www.blueservo.net	U.S.A.	Image Recognition	Collective Knowledge, Knowledge Acquisition
49	**cerberusgame**	www.cerberusgame.com	The Netherlands	ESP Online Game, Image Recognition, Science, Astronomy	Collective Knowledge, Knowledge Acquisition
50	**Freelancer**	www.freelancer.com	Australia	High Level Problem Solving, Task Allocating, Job Marketplace	Cloud Labor, Problem Solving
51	**Get a Slogan**	www.getaslogan.com	N.A.	Slogan Creation	Collective Creativity, Cloud Labor, Knowledge Acquisition
52	**Prova**	www.prova.com	N.A.	Creativity, Web Design, Logo Design, Graphic Design	Collective Creativity, Cloud Labor
53	**SocialAttire**	www.socialattire.com	U.S.A.	Creativity, Product Design, Clothes	Collective Creativity, Knowledge Acquisition
54	**FundIt**	www.fundit.ie	Ireland	Creativity Projects Funding	Crowdfunding
55	**Philoptima**	www.philoptima.org	U.S.A.	High Level Problem Solving, Job Marketplace	Open Innovation, Cloud Labor, Problem Solving
56	**TopCoder**	www.topcoder.com	U.S.A.	Freelancer Coders, Programming, Job Marketplace	Cloud Labor, Problem Solving
57	**uTest**	www.utest.com	U.S.A.	Software Testing	Cloud Labor, Problem Solving
58	**Fixya**	www.fixya.com	U.S.A.	Community-Driven Costumer Care	Cloud Labor, Collective Knowledge, Knowledge Acquisition, Knowledge Sharing

#	NAME	WEBSITE URL	LOCATION	DESCRIPTION	*CATEGORIZATION*
59	**bootB**	www.bootb.com	Italy	Creativity, Design, Logo Design, Graphic Design, Brand Design	Collective Creativity, Cloud Labor
60	**Zooppa**	www.zooppa.com	Italy	Creativity, Design, Logo Design, Graphic Design, Brand Design	Collective Creativity, Cloud Labor
61	**iStockPhoto**	www.istockphoto.com	Canada	Creativity, Stock, Photography, Illustrations, Video	Collective Creativity, Cloud Labor

- Crowdsourcing Type, Task Type, Required Knowledge, Community Size (Quantitative), Community Size (Qualitative), User Type

#	NAME	*CROWDSOURCING TYPE*	*TASK TYPE*	*REQUIRED KNOWLEDGE*	*COMMUNITY SIZE (QUANTITATIVE)*	*COMMUNITY SIZE (QUALITATIVE)*	*USER TYPE*
1	**Pledgemusic**	Integration	Simple	Low	N.A.	N.A.	Amateur
2	**global for me**	Integration	Simple	Low	N.A.	N.A.	Amateur
3	**IMDB**	Integration	Simple	Low	N.A.	Big (*) Estimated	Amateur
4	**Delicious**	Integration	Simple	Low	6M *source: http://en.wikipedia.org/wiki/Delicious_(website)	Big	Amateur
5	**Amazon Mechanical Turk**	Selection	Simple	Low, Medium, High	100K *source: http://www.propublica.org/article/propublicas-guide-to-mechanical-turk	Small	Amateur, Professional
6	**Wikipedia**	Integration	Simple, Complex	Medium, High	15M *source: http://en.wikipedia.org/wiki/Wikipedia:Wikipedians	Big	Amateur, Professional
7	**Foursquare**	Integration	Simple	Low	7M *source: http://techcrunch.com/2011/02/21/foursquare-closing-in-on-7-million-users/	Big	Amateur
8	**Article One Partners**	Selection	Complex	Medium, High	11K *source: http://www.articleonepartners.com/patent-research-community/?u=10920	Small	Amateur, Professional
9	**crowdSPRING**	Selection	Simple, Complex	Medium, High	100K *source: http://www.crowdspring.com/how-it-works/	Small	Amateur, Professional

#	NAME	CROWDSOURCING TYPE	TASK TYPE	REQUIRED KNOWLEDGE	COMMUNITY SIZE (QUANTITATIVE)	COMMUNITY SIZE (QUALITATIVE)	USER TYPE
10	**clickworker**	Selection	Simple, Complex	Medium, High	100K *source: http://www.clickworker.com/en/das-clickworker-prinzip/	Small	Amateur, Professional
11	**Waze**	Integration	Simple	Low	4M *source: http://www.waze.com/about/the_buzz/more_press/	Big	Amateur
12	**Urtak**	Integration	Simple	Medium, High	7K *source: https://urtak.com/	Small	Amateur
13	**CloudCrowd**	Selection	Simple, Complex	Medium, High	80K *source: http://www.serv.io/company/about	Small	Amateur, Professional
14	**CrowdFlower**	Selection	Simple, Complex	Medium, High	1M *source: http://venturebeat.com/2011/05/17/crowdflower-100-million-tasks/	Medium	Amateur, Professional
15	**Listener Driven Radio**	Integration	Simple	Low	N.A.	Small (*) Estimated	Amateur
16	**Kickstarter**	Integration	Simple	Low	700K *source: http://www.quora.com/How-many-users-does-Kickstarter-have	Medium	Amateur, Professional
17	**Amazon.com**	Integration	Simple	Low	80M *source: http://hpc.isti.cnr.it/~claudio/corsi/ecomm2010/Amazon.pdf	Big	Amateur
18	**Funding4Learning**	Integration	Simple	Low	N.A.	N.A.	Amateur
19	**RocketHub**	Integration	Simple	Low	N.A.	N.A.	Amateur
20	**Grow Venture Community**	Integration	Simple	Low	N.A.	N.A.	Amateur
21	**Sponsume**	Integration	Simple	Low	N.A.	N.A.	Amateur
22	**Stardust@home**	Selection	Simple	Low, Medium	25K *source: http://stardustathome.ssl.berkeley.edu/p1_rankings.php	Small	Amateur
23	**Zooniverse**	Selection	Simple	Low, Medium	350K *source: http://blogs.zooniverse.org/blog/2010/12/05/350000-zooniverse-users/	Medium	Amateur
24	**Distributed Proofreaders**	Selection	Complex	Low, Medium	3K *source: http://www.pgdp.net/c/stats/stats_central.php	Small	Amateur
25	**wikimapia**	Integration	Simple	Low	1M *source: http://blog.wikimapia.org/node/33	Medium	Amateur
26	**google maps**	Integration	Simple	Low	N.A.	Big (*) Estimated	Amateur
27	**Facebook Translation**	Integration	Simple, Complex	Low, Medium	N.A.	N.A.	Amateur

#	NAME	CROWDSOURCING TYPE	TASK TYPE	REQUIRED KNOWLEDGE	COMMUNITY SIZE (QUANTITATIVE)	COMMUNITY SIZE (QUALITATIVE)	USER TYPE
28	**spot.us**	Integration	Simple	Low	11K *source: http://www.spot.us/	Small	Amateur
29	**Ebay**	Integration	Simple	Low	95M *source: http://en.wikipedia.org/wiki/EBay	Big	Amateur
30	**buyacredit**	Integration	Simple	Low	N.A.	N.A.	Amateur
31	**Zazzle T-shirt**	Selection	Complex	Medium	N.A.	N.A.	Amateur, Professional
32	**Sellaband**	Integration	Simple	Low	N.A.	N.A.	Amateur
33	**threadless**	Selection	Complex	Medium	900K *source: http://news.cnet.com/8301-17939_109-10211721-2.html	Medium	Amateur, Professional
34	**OpenStreetMap**	Integration	Simple	Low	400K *source: http://en.wikipedia.org/wiki/OpenStreetMap	Medium	Amateur
35	**Innocentive**	Selection	Complex	Medium, High	250K *source: http://www.innocentive.com/about-innocentive/facts-stats	Medium	Amateur, Professional
36	**LinkedIn Answers**	Integration	Complex	Medium, High	1M *source: http://www.crunchbase.com/company/linkedin	Big	Amateur, Professional
37	**Answerbag**	Integration	Complex	Medium, High	2M *source: http://www.crunchbase.com/company/answerbag	Big	Amateur, Professional
38	**Free Knowledge Exchange**	Integration	Complex	Medium, High	N.A.	N.A.	Amateur, Professional
39	**Google Answers**	Integration	Complex	Medium, High	500 *source: http://en.wikipedia.org/wiki/Google_Answers	Small	Professional
40	**WikiAnswers**	Integration	Complex	Medium, High	5M *source: http://www.nostupidanswers.com/about-2/	Big	Amateur, Professional
41	**FunAdvice.com**	Integration	Complex	Medium, High	2M *source: http://en.wikipedia.org/wiki/FunAdvice.com	Big	Amateur, Professional
55	**Philoptima**	Selection	Complex	High	N.A.	N.A.	Amateur, Professional
56	**TopCoder**	Selection	Complex	Medium, High	300K *source: http://www.topcoder.com/	Medium	Professional
57	**uTest**	Integration	Complex	Medium	30K *source: http://www.crunchbase.com/company/utest	Small	Amateur, Professional
58	**Fixya**	Integration	Complex	Medium, High	10M *source: http://www.crunchbase.com/company/fixya	Big	Amateur, Professional
59	**bootB**	Selection	Complex	Medium, High	N.A.	N.A.	Amateur, Professional
60	**Zooppa**	Selection	Complex	Medium, High	N.A.	N.A.	Amateur, Professional

#	NAME	CROWDSOURCING TYPE	TASK TYPE	REQUIRED KNOWLEDGE	COMMUNITY SIZE (QUANTITATIVE)	COMMUNITY SIZE (QUALITATIVE)	USER TYPE
61	**iStockPhoto**	Selection	Complex	Medium, High	6M *source: http://press.istockphoto.com/pr/isp/background.aspx	Big	Amateur, Professional

- Main Reward, Minor Reward, Remuneration (Quantitative), Remuneration (Qualitative), Incentive, Data Quality Mechanism

#	NAME	MAIN REWARD	MINOR REWARD	REMUNERATION (QUANTITATIVE $)	REMUNERATION (QUALITATIVE)	INCENTIVE	DATA QUALITY MECHANISM
1	**Pledgemusic**	Enjoyment-based motivation	N.A.	N.A.	N.A.	Sharing of the result, Sharing of the goal	N.A.
2	**global for me**	Enjoyment-based motivation	N.A.	N.A.	N.A.	Sharing of the result, Sharing of the goal	N.A.
3	**IMDB**	Enjoyment-based motivation	Prestige-oriented	N.A.	N.A.	Sharing of the result, Sharing of the goal, User ranking and voting systems	Group Evaluation [Voting, Averaging], Reward Accuracy, Competition
4	**Delicious**	Enjoyment-based motivation	N.A.	N.A.	N.A.	Sharing of the result, Sharing of the goal	Group Evaluation [Voting, Averaging]
5	**Amazon Mechanical Turk**	Opportunistic	Prestige-oriented	0.10 - 50	Low	Money, Competition, User ranking and voting systems, Position inside community and user power scaling	Group Evaluation [Voting, Averaging], Reward Accuracy, Competition, Surveillance
6	**Wikipedia**	Enjoyment-based motivation	Prestige-oriented	N.A.	N.A.	Sharing of the result, Sharing of the goal, User ranking and voting systems, Position inside community and user power scaling	Group Evaluation [Voting, Consensus] Reward Accuracy, Competition, Surveillance
7	**Foursquare**	Enjoyment-based motivation	Opportunistic	N.A.	N.A.	Sharing of the result, Sharing of the goal	Group Evaluation [Voting, Averaging]
8	**Article One Partners**	Opportunistic	N.A.	5000 - 50000	Medium	Money	Reward Accuracy, Competition, Surveillance
9	**crowdSPRING**	Opportunistic	Enjoyment-based motivation	100 - 2000	Medium	Money, User ranking and voting systems	Reward Accuracy, Competition, Surveillance
10	**clickworker**	Opportunistic	Prestige-oriented	0.10 - 10	Low	Money, User ranking and voting systems	Reward Accuracy, Competition, Surveillance
11	**Waze**	Enjoyment-based motivation	N.A.	N.A.	N.A.	Sharing of the result, Sharing of the goal, User ranking and voting systems, Position inside community and user power scaling	Group Evaluation [Voting, Averaging]
12	**Urtak**	Enjoyment-based motivation	N.A.	N.A.	N.A.	Sharing of the result, Sharing of the goal	Group Evaluation [Voting, Averaging]

#	NAME	*MAIN REWARD*	*MINOR REWARD*	*REMUNERATION (QUANTITATIVE $)*	*REMUNERATION (QUALITATIVE)*	*INCENTIVE*	*DATA QUALITY MECHANISM*
13	**CloudCrowd**	Opportunistic	N.A.	0.10 - 15	Low	Money, User ranking and voting systems	Reward Accuracy, Competition, Surveillance
14	**CrowdFlower**	Opportunistic	N.A.	N.A.	N.A.	Money, User ranking and voting systems	Reward Accuracy, Competition, Surveillance
15	**Listener Driven Radio**	Enjoyment-based motivation	N.A.	N.A.	N.A.	Sharing of the result, Sharing of the goal	Group Evaluation [Voting]
16	**Kickstarter**	Enjoyment-based motivation	N.A.	N.A.	N.A.	Sharing of the result, Sharing of the goal	N.A.
17	**Amazon.com**	Enjoyment-based motivation	N.A.	N.A.	N.A.	Sharing of the result, Sharing of the goal	Group Evaluation [Voting, Averaging], Reward Accuracy, Competition
18	**Funding4Learning**	Enjoyment-based motivation	N.A.	N.A.	N.A.	Sharing of the result, Sharing of the goal	N.A.
19	**RocketHub**	Enjoyment-based motivation	N.A.	N.A.	N.A.	Sharing of the result, Sharing of the goal	N.A.
20	**Grow Venture Community**	Enjoyment-based motivation	N.A.	N.A.	N.A.	Sharing of the result, Sharing of the goal	N.A.
21	**Sponsume**	Enjoyment-based motivation	N.A.	N.A.	N.A.	Sharing of the result, Sharing of the goal	N.A.
22	**Stardust@home**	Enjoyment-based motivation	N.A.	N.A.	N.A.	Sharing of the result, Sharing of the goal	Group Evaluation [Averaging]
23	**Zooniverse**	Enjoyment-based motivation	N.A.	N.A.	N.A.	Sharing of the result, Sharing of the goal	Group Evaluation [Averaging]
24	**Distributed Proofreaders**	Enjoyment-based motivation	N.A.	N.A.	N.A.	Money	Group Evaluation [Averaging], Surveillance
25	**wikimapia**	Enjoyment-based motivation	N.A.	N.A.	N.A.	Sharing of the result, Sharing of the goal	Group Evaluation [Averaging]
26	**google maps**	Enjoyment-based motivation	N.A.	N.A.	N.A.	Sharing of the result, Sharing of the goal	Group Evaluation [Averaging]
27	**Facebook Translation**	Enjoyment-based motivation	N.A.	N.A.	N.A.	Sharing of the result, Sharing of the goal	Group Evaluation [Averaging], Surveillance
28	**spot.us**	Enjoyment-based motivation	N.A.	N.A.	N.A.	Sharing of the result, Sharing of the goal	N.A.
29	**Ebay**	Opportunistic	Prestige-oriented	N.A.	N.A.	Sharing of the result, Sharing of the goal, User ranking and voting systems, Position inside community and user power scaling	Group Evaluation [Voting, Averaging], Reward Accuracy, Competition, Surveillance
30	**buyacredit**	Enjoyment-based motivation	N.A.	N.A.	N.A.	Sharing of the result, Sharing of the goal	N.A.
31	**Zazzle T-shirt**	Enjoyment-based motivation	Prestige-oriented	N.A.	N.A.	Sharing of the result, Competition, User ranking and voting systems, Position inside community and user power scaling	Group Evaluation [Voting], Reward Accuracy, Competition, Surveillance

#	NAME	MAIN REWARD	MINOR REWARD	REMUNERATION (QUANTITATIVE $)	REMUNERATION (QUALITATIVE)	INCENTIVE	DATA QUALITY MECHANISM
32	**Sellaband**	Enjoyment-based motivation	N.A.	N.A.	N.A.	Sharing of the result, Sharing of the goal	N.A.
33	**threadless**	Enjoyment-based motivation	Prestige-oriented	N.A.	N.A.	Sharing of the result, Competition, User ranking and voting systems, Position inside community and user power scaling	Group Evaluation [Voting], Reward Accuracy, Competition, Surveillance
34	**OpenStreetMap**	Enjoyment-based motivation	N.A.	N.A.	N.A.	Sharing of the result, Sharing of the goal	Group Evaluation [Averaging]
35	**Innocentive**	Opportunistic	Prestige-oriented	10000 - 10000000	High	Money, Sharing the result, Sharing the goal	Reward Accuracy, Competition
36	**LinkedIn Answers**	Enjoyment-based motivation	Prestige-oriented	N.A.	N.A.	Sharing of the result, Sharing of the goal	Group Evaluation [Voting], Reward Accuracy, Competition
37	**Answerbag**	Enjoyment-based motivation	Prestige-oriented	N.A.	N.A.	Sharing of the result, Sharing of the goal, User ranking and voting systems	Group Evaluation [Voting], Reward Accuracy, Competition
38	**Free Knowledge Exchange**	Enjoyment-based motivation	N.A.	N.A.	N.A.	Sharing of the result, Sharing of the goal	Group Evaluation [Voting]
39	**Google Answers**	Enjoyment-based motivation	Prestige-oriented	0 - 1000	Medium	Money, Sharing of the result, Sharing of the goal, User ranking and voting systems	Group Evaluation [Voting], Reward Accuracy, Competition
40	**WikiAnswers**	Enjoyment-based motivation	Prestige-oriented	N.A.	N.A.	Sharing of the result, Sharing of the goal, User ranking and voting systems	Group Evaluation [Voting], Reward Accuracy, Competition
41	**FunAdvice.com**	Enjoyment-based motivation	Prestige-oriented	N.A.	N.A.	Sharing of the result, Sharing of the goal, User ranking and voting systems	Group Evaluation [Voting], Reward Accuracy, Competition
42	**Yahoo! Answers**	Enjoyment-based motivation	Prestige-oriented	N.A.	N.A.	Sharing of the result, Sharing of the goal, User ranking and voting systems	Group Evaluation [Voting], Reward Accuracy, Competition
43	**Digg**	Enjoyment-based motivation	Prestige-oriented	N.A.	N.A.	Sharing of the result, Sharing of the goal	Group Evaluation [Voting]
56	**TopCoder**	Opportunistic	Prestige-oriented	100 - 10000	Medium, High	Money, Competition, User ranking and voting systems	Reward Accuracy, Competition
57	**uTest**	Opportunistic	Enjoyment-based motivation	100 - 1000	Medium	Money, User ranking and voting systems	Reward Accuracy, Competition
58	**Fixya**	Enjoyment-based motivation	Prestige-oriented	N.A.	N.A.	Money, Sharing of the result, Sharing of the goal, User ranking and voting systems	Group Evaluation [Voting], Reward Accuracy, Competition
59	**bootB**	Opportunistic	Prestige-oriented	1000 - 10000	Medium, High	Money, Competition, User ranking and voting systems	Reward Accuracy, Competition
60	**Zooppa**	Opportunistic	Prestige-oriented	10000 - 50000	Medium, High	Money, Competition, User ranking and voting systems	Reward Accuracy, Competition
61	**iStockPhoto**	Opportunistic	Enjoyment-based motivation	1 - 100	Low	Money, Competition, User ranking and voting systems	Reward Accuracy, Competition

The data about the Remuneration of the elements in the dataset were retrieved by joining the community and observing the trends

APPENDIX B

- First 7 questions

#User	Please enter your age	Please enter your sex	Please enter your education level	Please enter the number of hours spent online in a day	I am likely to join online web communities and to share my knowledge	I think the size of the community matters when choosing if joining or not an online community	I would share my knowledge and spend my time contributing to online communities for free if I share the goals
1	28	Female	Master Degree	4	Agree	Agree	Agree
2	25	Male	Bachelor Degree	6	Agree	Strongly agree	Agree
3	25	Male	Master Degree	4	Agree	Strongly agree	Strongly agree
4	26	Male	Master Degree	3	Agree	Agree	Agree
5	33	Male	PhD	10	Neither agree nor disagree	Agree	Agree
6	33	Male	PhD	10	Neither agree nor disagree	Agree	Agree
7	33	Male	PhD	10	Neither agree nor disagree	Agree	Agree

#User	Please enter your age	Please enter your sex	Please enter your education level	Please enter the number of hours spent online in a day	I am likely to join online web communities and to share my knowledge	I think the size of the community matters when choosing if joining or not an online community	I would share my knowledge and spend my time contributing to online communities for free if I share the goals
8	33	Male	PhD	10	Neither agree nor disagree	Agree	Agree
9	20	Male	Bachelor Degree	4	Strongly agree	Strongly agree	Agree
10	38	Female	Master Degree	2	Neither agree nor disagree	Agree	Agree
11	42	Male	High School Diploma	1	Neither agree nor disagree	Agree	Agree
12	39	Female	High School Diploma	1	Neither agree nor disagree	Agree	Agree
13	25	Male	Master Degree	3	Neither agree nor disagree	Agree	Agree

#User	Please enter your age	Please enter your sex	Please enter your education level	Please enter the number of hours spent online in a day	I am likely to join online web communities and to share my knowledge	I think the size of the community matters when choosing if joining or not an online community	I would share my knowledge and spend my time contributing to online communities for free if I share the goals
14	25	Male	Bachelor Degree	3	Neither agree nor disagree	Agree	Agree
15	25	Female	Master Degree	2	Neither agree nor disagree	Agree	Agree
16	25	Female	Bachelor Degree	2	Neither agree nor disagree	Agree	Agree
17	27	Female	High School Diploma	2	Neither agree nor disagree	Agree	Agree
18	60	Male	Master Degree	2	Disagree	Neither agree nor disagree	Agree
19	25	Female	High School Diploma	2	Disagree	Neither agree nor disagree	Agree
20	25	Male	High School Diploma	2	Disagree	Neither agree nor disagree	Agree

#User	Please enter your age	Please enter your sex	Please enter your education level	Please enter the number of hours spent online in a day	I am likely to join online web communities and to share my knowledge	I think the size of the community matters when choosing if joining or not an online community	I would share my knowledge and spend my time contributing to online communities for free if I share the goals
21	25	Male	Master Degree	5	Disagree	Neither agree nor disagree	Agree
22	26	Female	Master Degree	2	Disagree	Neither agree nor disagree	Agree
23	32	Female	High School Diploma	1	Disagree	Neither agree nor disagree	Agree
24	66	Male	High School Diploma	2	Disagree	Neither agree nor disagree	Agree
25	65	Male	High School Diploma	1	Disagree	Neither agree nor disagree	Agree
26	43	Female	High School Diploma	1	Disagree	Neither agree nor disagree	Agree
27	47	Male	High School Diploma	2	Disagree	Neither agree nor disagree	Agree
28	55	Female	High School Diploma	1	Disagree	Neither agree nor disagree	Agree

#User	Please enter your age	Please enter your sex	Please enter your education level	Please enter the number of hours spent online in a day	I am likely to join online web communities and to share my knowledge	I think the size of the community matters when choosing if joining or not an online community	I would share my knowledge and spend my time contributing to online communities for free if I share the goals
29	25	Male	Master Degree	4	Agree	Agree	Agree
30	24	Female	Master Degree	2	Agree	Agree	Agree
31	24	Female	Master Degree	4	Agree	Agree	Agree
32	24	Female	Master Degree	2	Agree	Agree	Agree
33	23	Male	Master Degree	4	Agree	Agree	Agree
34	23	Male	Master Degree	3	Agree	Agree	Agree
35	24	Male	Master Degree	3	Agree	Agree	Agree
36	24	Male	Master Degree	3	Agree	Agree	Agree

#User	Please enter your age	Please enter your sex	Please enter your education level	Please enter the number of hours spent online in a day	I am likely to join online web communities and to share my knowledge	I think the size of the community matters when choosing if joining or not an online community	I would share my knowledge and spend my time contributing to online communities for free if I share the goals
37	24	Female	Master Degree	2	Agree	Agree	Agree
38	25	Male	PhD	5	Agree	Agree	Agree
39	27	Female	Master Degree	2	Agree	Agree	Agree
40	26	Male	Bachelor Degree	5	Agree	Strongly agree	Agree
41	21	Male	Bachelor Degree	5	Agree	Strongly agree	Agree
42	22	Male	Bachelor Degree	6	Agree	Strongly agree	Agree

#User	Please enter your age	Please enter your sex	Please enter your education level	Please enter the number of hours spent online in a day	I am likely to join online web communities and to share my knowledge	I think the size of the community matters when choosing if joining or not an online community	I would share my knowledge and spend my time contributing to online communities for free if I share the goals
43	25	Male	High School Diploma	3	Agree	Strongly agree	Agree
44	25	Female	Bachelor Degree	3	Agree	Strongly agree	Agree
45	26	Female	PhD	5	Agree	Strongly agree	Agree
46	31	Male	Bachelor Degree	5	Agree	Strongly agree	Agree
47	25	Male	Master Degree	3	Agree	Strongly agree	Agree
48	30	Female	High school diploma	2	Agree	Neither agree nor disagree	Strongly agree
49	30	Female	High school diploma	2	Agree	Neither agree nor disagree	Strongly agree

#User	Please enter your age	Please enter your sex	Please enter your education level	Please enter the number of hours spent online in a day	I am likely to join online web communities and to share my knowledge	I think the size of the community matters when choosing if joining or not an online community	I would share my knowledge and spend my time contributing to online communities for free if I share the goals
50	30	Female	High school diploma	2	Agree	Neither agree nor disagree	Strongly agree
51	30	Female	High school diploma	2	Agree	Neither agree nor disagree	Strongly agree

- Next 7 questions

#User	I would share my knowledge and spend my time contributing to online communities for free if I can access to the results and they are useful for me	I would share my knowledge and spend my time contributing to online communities for free if I get fun	I am likely to contribute much more to an online community if I can receive any kind of monetary compensation even if I share the goals and/or the results	I am likely to contribute to online communities only if performing simple (tag images, assign values, etc.) tasks	I am likely to contribute to online communities also performing complex (write long essays, translation of text, etc.) tasks if I have the skills	I would not join an online communities and/or perform tasks if they are not linked to my interests even if I get a monetary compensation	It wasn't a problem for me to provides some additional information while placing new questions
1	Strongly agree	Neither agree nor disagree	Neither agree nor disagree	Disagree	Agree	Agree	Agree
2	Agree	Agree	Strongly agree	Agree	Disagree	Agree	Agree
3	Strongly agree	Strongly agree	Agree	Agree	Disagree	Agree	Strongly agree
4	Agree	Agree	Agree	Agree	Agree	Strongly agree	Strongly agree

#User	I would share my knowledge and spend my time contributing to online communities for free if I can access to the results and they are useful for me	I would share my knowledge and spend my time contributing to online communities for free if I get fun	I am likely to contribute much more to an online community if I can receive any kind of monetary compensation even if I share the goals and/or the results	I am likely to contribute to online communities only if performing simple (tag images, assign values, etc.) tasks	I am likely to contribute to online communities also performing complex (write long essays, translation of text, etc.) tasks if I have the skills	I would not join an online communities and/or perform tasks if they are not linked to my interests even if I get a monetary compensation	It wasn't a problem for me to provides some additional information while placing new questions
5	Strongly agree	Neither agree nor disagree	Strongly agree	Strongly agree	Disagree	Strongly disagree	Strongly agree
6	Strongly agree	Neither agree nor disagree	Strongly agree	Strongly agree	Disagree	Strongly disagree	Strongly agree
7	Strongly agree	Neither agree nor disagree	Strongly agree	Strongly agree	Disagree	Strongly disagree	Strongly agree
8	Strongly agree	Neither agree nor disagree	Strongly agree	Strongly agree	Disagree	Strongly disagree	Strongly agree
9	Agree	Agree	Agree	Agree	Agree	Strongly agree	Agree

#User	I would share my knowledge and spend my time contributing to online communities for free if I can access to the results and they are useful for me	I would share my knowledge and spend my time contributing to online communities for free if I get fun	I am likely to contribute much more to an online community if I can receive any kind of monetary compensation even if I share the goals and/or the results	I am likely to contribute to online communities only if performing simple (tag images, assign values, etc.) tasks	I am likely to contribute to online communities also performing complex (write long essays, translation of text, etc.) tasks if I have the skills	I would not join an online communities and/or perform tasks if they are not linked to my interests even if I get a monetary compensation	It wasn't a problem for me to provides some additional information while placing new questions
10	Agree	Agree	Agree	Agree	Agree	Agree	Agree
11	Agree	Agree	Agree	Agree	Agree	Agree	Agree
12	Agree	Agree	Agree	Agree	Agree	Agree	Agree
13	Agree	Agree	Agree	Agree	Agree	Agree	Agree
14	Agree	Agree	Agree	Agree	Agree	Agree	Agree
15	Agree	Agree	Agree	Agree	Agree	Agree	Agree

#User	I would share my knowledge and spend my time contributing to online communities for free if I can access to the results and they are useful for me	I would share my knowledge and spend my time contributing to online communities for free if I get fun	I am likely to contribute much more to an online community if I can receive any kind of monetary compensation even if I share the goals and/or the results	I am likely to contribute to online communities only if performing simple (tag images, assign values, etc.) tasks	I am likely to contribute to online communities also performing complex (write long essays, translation of text, etc.) tasks if I have the skills	I would not join an online communities and/or perform tasks if they are not linked to my interests even if I get a monetary compensation	It wasn't a problem for me to provides some additional information while placing new questions
16	Agree	Agree	Agree	Agree	Agree	Agree	Agree
17	Agree	Agree	Agree	Agree	Agree	Agree	Agree
18	Agree	Disagree	Disagree	Neither agree nor disagree	Neither agree nor disagree	Strongly agree	Strongly agree
19	Agree	Disagree	Disagree	Neither agree nor disagree	Neither agree nor disagree	Strongly agree	Strongly agree
20	Agree	Disagree	Disagree	Neither agree nor disagree	Neither agree nor disagree	Strongly agree	Strongly agree
21	Agree	Disagree	Disagree	Neither agree nor disagree	Neither agree nor disagree	Strongly agree	Strongly agree
22	Agree	Disagree	Disagree	Neither agree nor disagree	Neither agree nor disagree	Strongly agree	Strongly agree

#User	I would share my knowledge and spend my time contributing to online communities for free if I can access to the results and they are useful for me	I would share my knowledge and spend my time contributing to online communities for free if I get fun	I am likely to contribute much more to an online community if I can receive any kind of monetary compensation even if I share the goals and/or the results	I am likely to contribute to online communities only if performing simple (tag images, assign values, etc.) tasks	I am likely to contribute to online communities also performing complex (write long essays, translation of text, etc.) tasks if I have the skills	I would not join an online communities and/or perform tasks if they are not linked to my interests even if I get a monetary compensation	It wasn't a problem for me to provides some additional information while placing new questions
23	Agree	Disagree	Disagree	Neither agree nor disagree	Neither agree nor disagree	Strongly agree	Strongly agree
24	Agree	Disagree	Disagree	Neither agree nor disagree	Neither agree nor disagree	Strongly agree	Strongly agree
25	Agree	Disagree	Disagree	Neither agree nor disagree	Neither agree nor disagree	Strongly agree	Strongly agree
26	Agree	Disagree	Disagree	Neither agree nor disagree	Neither agree nor disagree	Strongly agree	Strongly agree
27	Agree	Disagree	Disagree	Neither agree nor disagree	Neither agree nor disagree	Strongly agree	Strongly agree
28	Agree	Disagree	Disagree	Neither agree nor disagree	Neither agree nor disagree	Strongly agree	Strongly agree
29	Strongly agree	Neither agree nor disagree	Neither agree nor disagree	Disagree	Agree	Agree	Agree

#User	I would share my knowledge and spend my time contributing to online communities for free if I can access to the results and they are useful for me	I would share my knowledge and spend my time contributing to online communities for free if I get fun	I am likely to contribute much more to an online community if I can receive any kind of monetary compensation even if I share the goals and/or the results	I am likely to contribute to online communities only if performing simple (tag images, assign values, etc.) tasks	I am likely to contribute to online communities also performing complex (write long essays, translation of text, etc.) tasks if I have the skills	I would not join an online communities and/or perform tasks if they are not linked to my interests even if I get a monetary compensation	It wasn't a problem for me to provides some additional information while placing new questions
30	Strongly agree	Neither agree nor disagree	Neither agree nor disagree	Disagree	Agree	Agree	Agree
31	Strongly agree	Neither agree nor disagree	Neither agree nor disagree	Disagree	Agree	Agree	Agree
32	Strongly agree	Neither agree nor disagree	Neither agree nor disagree	Disagree	Agree	Agree	Agree
33	Strongly agree	Neither agree nor disagree	Neither agree nor disagree	Disagree	Agree	Agree	Agree
34	Strongly agree	Neither agree nor disagree	Neither agree nor disagree	Disagree	Agree	Agree	Agree
35	Strongly agree	Neither agree nor disagree	Neither agree nor disagree	Disagree	Agree	Agree	Agree
36	Strongly agree	Neither agree nor disagree	Neither agree nor disagree	Disagree	Agree	Agree	Agree

#User	I would share my knowledge and spend my time contributing to online communities for free if I can access to the results and they are useful for me	I would share my knowledge and spend my time contributing to online communities for free if I get fun	I am likely to contribute much more to an online community if I can receive any kind of monetary compensation even if I share the goals and/or the results	I am likely to contribute to online communities only if performing simple (tag images, assign values, etc.) tasks	I am likely to contribute to online communities also performing complex (write long essays, translation of text, etc.) tasks if I have the skills	I would not join an online communities and/or perform tasks if they are not linked to my interests even if I get a monetary compensation	It wasn't a problem for me to provides some additional information while placing new questions
37	Strongly agree	Neither agree nor disagree	Neither agree nor disagree	Disagree	Agree	Agree	Agree
38	Strongly agree	Neither agree nor disagree	Neither agree nor disagree	Disagree	Agree	Agree	Agree
39	Strongly agree	Neither agree nor disagree	Neither agree nor disagree	Disagree	Agree	Agree	Agree
40	Agree	Agree	Strongly agree	Agree	Disagree	Agree	Agree
41	Agree	Agree	Strongly agree	Agree	Disagree	Agree	Agree

#User	I would share my knowledge and spend my time contributing to online communities for free if I can access to the results and they are useful for me	I would share my knowledge and spend my time contributing to online communities for free if I get fun	I am likely to contribute much more to an online community if I can receive any kind of monetary compensation even if I share the goals and/or the results	I am likely to contribute to online communities only if performing simple (tag images, assign values, etc.) tasks	I am likely to contribute to online communities also performing complex (write long essays, translation of text, etc.) tasks if I have the skills	I would not join an online communities and/or perform tasks if they are not linked to my interests even if I get a monetary compensation	It wasn't a problem for me to provides some additional information while placing new questions
42	Agree	Agree	Strongly agree	Agree	Disagree	Agree	Agree
43	Agree	Agree	Strongly agree	Agree	Disagree	Agree	Agree
44	Agree	Agree	Strongly agree	Agree	Disagree	Agree	Agree
45	Agree	Agree	Strongly agree	Agree	Disagree	Agree	Agree
46	Agree	Agree	Strongly agree	Agree	Disagree	Agree	Agree
47	Agree	Agree	Strongly agree	Agree	Disagree	Agree	Agree

#User	I would share my knowledge and spend my time contributing to online communities for free if I can access to the results and they are useful for me	I would share my knowledge and spend my time contributing to online communities for free if I get fun	I am likely to contribute much more to an online community if I can receive any kind of monetary compensation even if I share the goals and/or the results	I am likely to contribute to online communities only if performing simple (tag images, assign values, etc.) tasks	I am likely to contribute to online communities also performing complex (write long essays, translation of text, etc.) tasks if I have the skills	I would not join an online communities and/or perform tasks if they are not linked to my interests even if I get a monetary compensation	It wasn't a problem for me to provides some additional information while placing new questions
48	Neither agree nor disagree	Strongly agree	Disagree	Neither agree nor disagree	Neither agree nor disagree	Agree	Agree
49	Neither agree nor disagree	Strongly agree	Disagree	Neither agree nor disagree	Neither agree nor disagree	Agree	Agree
50	Neither agree nor disagree	Strongly agree	Disagree	Neither agree nor disagree	Neither agree nor disagree	Agree	Agree
51	Neither agree nor disagree	Strongly agree	Disagree	Neither agree nor disagree	Neither agree nor disagree	Agree	Agree

- Next 4 questions

#User	Please order the four scenarios according to how likely would you join each of them	Please order the four scenarios according to their enjoyableness	Please order the four scenarios according to the number of questions that you are likely to answer or post in each of them	Which range of compensation would you judge fair for providing us useful information by placing new questions?
1	3,2,1,4	3,2,4,1	2,3,4,1	I don't care / everything would be fine
2	1,2,3,4	1,2,4,3	1,2,3,4	0.01-0.10 euro for question

#User	Please order the four scenarios according to how likely would you join each of them	Please order the four scenarios according to their enjoyableness	Please order the four scenarios according to the number of questions that you are likely to answer or post in each of them	Which range of compensation would you judge fair for providing us useful information by placing new questions?
3	1,4,2,3	1,4,2,3	1,2,4,3	I don’t care / everything would be fine
4	3,4,2,1	3,4,,2,1	4,3,2,1	0.10-1 euro for question
5	4,3,2,1	3,4,,2,1	4,3,2,1	I don’t care / everything would be fine
6	4,3,2,1	3,4,,2,1	4,3,2,1	I don’t care / everything would be fine
7	4,3,2,1	3,4,,2,1	4,3,2,1	I don’t care / everything would be fine
8	4,3,2,1	3,4,,2,1	4,3,2,1	I don’t care / everything would be fine
9	3,4,2,1	3,4,,2,1	4,3,2,1	I don’t care / everything would be fine
10	3,2,4,1	3,2,4,1	3,2,4,1	I don’t care / everything would be fine

#User	Please order the four scenarios according to how likely would you join each of them	Please order the four scenarios according to their enjoyableness	Please order the four scenarios according to the number of questions that you are likely to answer or post in each of them	Which range of compensation would you judge fair for providing us useful information by placing new questions?
11	3,2,4,1	3,2,4,1	3,2,4,1	I don’t care / everything would be fine
12	3,2,4,1	3,2,4,1	3,2,4,1	I don’t care / everything would be fine
13	3,2,4,1	3,2,4,1	3,2,4,1	I don’t care / everything would be fine
14	3,2,4,1	3,2,4,1	3,2,4,1	I don’t care / everything would be fine
15	3,2,4,1	3,2,4,1	3,2,4,1	I don’t care / everything would be fine
16	3,2,4,1	3,2,4,1	3,2,4,1	I don’t care / everything would be fine

#User	Please order the four scenarios according to how likely would you join each of them	Please order the four scenarios according to their enjoyableness	Please order the four scenarios according to the number of questions that you are likely to answer or post in each of them	Which range of compensation would you judge fair for providing us useful information by placing new questions?
17	3,2,4,1	3,2,4,1	3,2,4,1	I don’t care / everything would be fine
18	2,3,1,4	2,3,1,4	2,1,3,4	I don’t care / everything would be fine
19	2,3,1,4	2,3,1,4	2,1,3,4	I don’t care / everything would be fine
20	2,3,1,4	2,3,1,4	2,1,3,4	I don’t care / everything would be fine
21	2,3,1,4	2,3,1,4	2,1,3,4	I don’t care / everything would be fine
22	2,3,1,4	2,3,1,4	2,1,3,4	I don’t care / everything would be fine
23	4,1,2,3	2,3,1,4	2,1,3,4	I don’t care / everything would be fine
24	2,3,1,4	2,3,1,4	2,1,3,4	I don’t care / everything would be fine

#User	Please order the four scenarios according to how likely would you join each of them	Please order the four scenarios according to their enjoyableness	Please order the four scenarios according to the number of questions that you are likely to answer or post in each of them	Which range of compensation would you judge fair for providing us useful information by placing new questions?
25	4,1,2,3	2,3,1,4	2,1,3,4	I don’t care / everything would be fine
26	2,3,1,4	2,3,1,4	2,1,3,4	I don’t care / everything would be fine
27	2,3,1,4	2,3,1,4	2,1,3,4	I don’t care / everything would be fine
28	2,3,1,4	4,1,2,3	2,1,3,4	I don’t care / everything would be fine
29	3,2,1,4	3,2,4,1	2,3,4,1	I don’t care / everything would be fine
30	3,2,1,4	3,2,4,1	2,3,4,1	I don’t care / everything would be fine
31	3,2,1,4	3,2,4,1	2,3,4,1	I don’t care / everything would be fine
32	3,2,1,4	3,2,4,1	2,3,4,1	I don’t care / everything would be fine

#User	Please order the four scenarios according to how likely would you join each of them	Please order the four scenarios according to their enjoyableness	Please order the four scenarios according to the number of questions that you are likely to answer or post in each of them	Which range of compensation would you judge fair for providing us useful information by placing new questions?
33	4,1,2,3	4,1,2,3	2,3,4,1	I don’t care / everything would be fine
34	3,2,1,4	3,2,4,1	2,3,4,1	I don’t care / everything would be fine
35	3,2,1,4	3,2,4,1	2,3,4,1	I don’t care / everything would be fine
36	4,1,2,3	4,1,2,3	2,3,4,1	I don’t care / everything would be fine
37	3,2,1,4	3,2,4,1	2,3,4,1	I don’t care / everything would be fine
38	3,2,1,4	3,2,4,1	2,3,4,1	I don’t care / everything would be fine

#User	Please order the four scenarios according to how likely would you join each of them	Please order the four scenarios according to their enjoyableness	Please order the four scenarios according to the number of questions that you are likely to answer or post in each of them	Which range of compensation would you judge fair for providing us useful information by placing new questions?
39	4,1,2,3	3,2,4,1	2,3,4,1	I don't care / everything would be fine
40	1,2,3,4	1,2,4,3	1,2,3,4	0.01-0.10 euro for question
41	1,2,3,4	1,2,4,3	1,2,3,4	0.01-0.10 euro for question
42	1,2,3,4	1,2,4,3	1,2,3,4	0.01-0.10 euro for question
43	1,2,3,4	4,1,2,3	1,2,3,4	0.01-0.10 euro for question
44	1,2,3,4	1,2,4,3	1,2,3,4	0.01-0.10 euro for question

#User	Please order the four scenarios according to how likely would you join each of them	Please order the four scenarios according to their enjoyableness	Please order the four scenarios according to the number of questions that you are likely to answer or post in each of them	Which range of compensation would you judge fair for providing us useful information by placing new questions?
45	1,2,3,4	1,2,4,3	1,2,3,4	0.01-0.10 euro for question
46	1,2,3,4	1,2,4,3	1,2,3,4	0.01-0.10 euro for question
47	4,1,2,3	1,2,4,3	1,2,3,4	0.01-0.10 euro for question
48	3,4,2,1	3,2,4,1	3,4,2,1	0.10-1 euro for question
49	3,4,2,1	3,2,4,1	3,4,2,1	0.10-1 euro for question
50	3,4,2,1	3,2,4,1	3,4,2,1	0.10-1 euro for question
51	3,4,2,1	3,2,4,1	3,4,2,1	0.10-1 euro for question

- Last 5 questions

#User	Instead of money would you accepted some other form of compensation (for instance a number of free analysis by the sentiment engine for your purposes) for providing us useful information by placing new questions?	Please order the following incentives according to how much they encourage you to correctly answer to the questions [User ranking in the second scenario, Gaming in the third scenario, Monetary Betting in the fourth scenario]	I would try to correctly answer the questions at my best even in the first scenario	Would you consciously provide false/incorrect information in the forth scenario just to benefit of the compensation?	Would you consciously provide false/incorrect information in the forth scenario just to benefit of the compensation knowing that the website uses automatic and/or manual mechanisms to avoid cheating?
1	No, in any case	1,2,3	Agree	No	No
2	Yes, even if I judge the monetary compensation adequate	3,2,1	Agree	Maybe	No
3	Yes, if I judge the monetary compensation low	2,3,1	Agree	No	No
4	No, in any case	3,2,1	Agree	Maybe	Maybe
5	No, in any case	1,3,2	Agree	Maybe	No
6	No, in any case	3,2,1	Agree	Maybe	No
7	No, in any case	3,2,1	Agree	Maybe	No

#User	Instead of money would you accepted some other form of compensation (for instance a number of free analysis by the sentiment engine for your purposes) for providing us useful information by placing new questions?	Please order the following incentives according to how much they encourage you to correctly answer to the questions [User ranking in the second scenario, Gaming in the third scenario, Monetary Betting in the fourth scenario]	I would try to correctly answer the questions at my best even in the first scenario	Would you consciously provide false/incorrect information in the forth scenario just to benefit of the compensation?	Would you consciously provide false/incorrect information in the forth scenario just to benefit of the compensation knowing that the website uses automatic and/or manual mechanisms to avoid cheating?
8	No, in any case	3,2,1	Agree	Maybe	No
9	No, in any case	2,3,1	Strongly disagree	No	No
10	Yes, even if I judge the monetary compensation adequate	2,1,3	Agree	No	No
11	Yes, even if I judge the monetary compensation adequate	2,1,3	Agree	No	No
12	Yes, even if I judge the monetary compensation adequate	2,1,3	Agree	No	No
13	Yes, even if I judge the monetary compensation adequate	2,1,3	Agree	No	No

#User	Instead of money would you accepted some other form of compensation (for instance a number of free analysis by the sentiment engine for your purposes) for providing us useful information by placing new questions?	Please order the following incentives according to how much they encourage you to correctly answer to the questions [User ranking in the second scenario, Gaming in the third scenario, Monetary Betting in the fourth scenario]	I would try to correctly answer the questions at my best even in the first scenario	Would you consciously provide false/incorrect information in the forth scenario just to benefit of the compensation?	Would you consciously provide false/incorrect information in the forth scenario just to benefit of the compensation knowing that the website uses automatic and/or manual mechanisms to avoid cheating?
14	Yes, even if I judge the monetary compensation adequate	2,1,3	Agree	No	No
15	Yes, even if I judge the monetary compensation adequate	1,3,2	Agree	No	No
16	Yes, even if I judge the monetary compensation adequate	2,1,3	Agree	No	No
17	Yes, even if I judge the monetary compensation adequate	2,1,3	Agree	No	No
18	No, in any case	1,2,3	Agree	No	No
19	No, in any case	1,2,3	Agree	No	No

#User	Instead of money would you accepted some other form of compensation (for instance a number of free analysis by the sentiment engine for your purposes) for providing us useful information by placing new questions?	Please order the following incentives according to how much they encourage you to correctly answer to the questions [User ranking in the second scenario, Gaming in the third scenario, Monetary Betting in the fourth scenario]	I would try to correctly answer the questions at my best even in the first scenario	Would you consciously provide false/incorrect information in the forth scenario just to benefit of the compensation?	Would you consciously provide false/incorrect information in the forth scenario just to benefit of the compensation knowing that the website uses automatic and/or manual mechanisms to avoid cheating?
20	No, in any case	1,3,2	Agree	No	No
21	No, in any case	1,2,3	Agree	No	No
22	No, in any case	1,2,3	Agree	No	No
23	No, in any case	1,2,3	Agree	No	No
24	No, in any case	1,2,3	Agree	No	No
25	No, in any case	1,2,3	Agree	No	No
26	No, in any case	1,2,3	Agree	No	No
27	No, in any case	1,2,3	Agree	No	No

#User	Instead of money would you accepted some other form of compensation (for instance a number of free analysis by the sentiment engine for your purposes) for providing us useful information by placing new questions?	Please order the following incentives according to how much they encourage you to correctly answer to the questions [User ranking in the second scenario, Gaming in the third scenario, Monetary Betting in the fourth scenario]	I would try to correctly answer the questions at my best even in the first scenario	Would you consciously provide false/incorrect information in the forth scenario just to benefit of the compensation?	Would you consciously provide false/incorrect information in the forth scenario just to benefit of the compensation knowing that the website uses automatic and/or manual mechanisms to avoid cheating?
28	No, in any case	1,2,3	Agree	No	No
29	No, in any case	1,2,3	Agree	No	No
30	No, in any case	1,2,3	Agree	No	No
31	No, in any case	1,3,2	Agree	No	No
32	No, in any case	1,2,3	Agree	No	No
33	No, in any case	1,2,3	Agree	No	No
34	No, in any case	1,3,2	Agree	No	No

#User	Instead of money would you accepted some other form of compensation (for instance a number of free analysis by the sentiment engine for your purposes) for providing us useful information by placing new questions?	Please order the following incentives according to how much they encourage you to correctly answer to the questions [User ranking in the second scenario, Gaming in the third scenario, Monetary Betting in the fourth scenario]	I would try to correctly answer the questions at my best even in the first scenario	Would you consciously provide false/incorrect information in the forth scenario just to benefit of the compensation?	Would you consciously provide false/incorrect information in the forth scenario just to benefit of the compensation knowing that the website uses automatic and/or manual mechanisms to avoid cheating?
35	No, in any case	1,2,3	Agree	No	No
36	No, in any case	1,2,3	Agree	No	No
37	No, in any case	1,2,3	Agree	No	No
38	No, in any case	1,2,3	Agree	No	No
39	No, in any case	1,2,3	Agree	No	No
40	Yes, even if I judge the monetary compensation adequate	1,3,2	Agree	Maybe	No
41	Yes, even if I judge the monetary compensation adequate	3,2,1	Agree	Maybe	No

#User	Instead of money would you accepted some other form of compensation (for instance a number of free analysis by the sentiment engine for your purposes) for providing us useful information by placing new questions?	Please order the following incentives according to how much they encourage you to correctly answer to the questions [User ranking in the second scenario, Gaming in the third scenario, Monetary Betting in the fourth scenario]	I would try to correctly answer the questions at my best even in the first scenario	Would you consciously provide false/incorrect information in the forth scenario just to benefit of the compensation?	Would you consciously provide false/incorrect information in the forth scenario just to benefit of the compensation knowing that the website uses automatic and/or manual mechanisms to avoid cheating?
42	Yes, even if I judge the monetary compensation adequate	3,2,1	Agree	Maybe	No
43	Yes, even if I judge the monetary compensation adequate	3,2,1	Agree	Maybe	No
44	Yes, even if I judge the monetary compensation adequate	1,3,2	Agree	Maybe	No
45	Yes, even if I judge the monetary compensation adequate	3,2,1	Agree	Maybe	No
46	Yes, even if I judge the monetary compensation adequate	3,2,1	Agree	Maybe	No
47	Yes, even if I judge the monetary compensation adequate	3,2,1	Agree	Maybe	No

#User	Instead of money would you accepted some other form of compensation (for instance a number of free analysis by the sentiment engine for your purposes) for providing us useful information by placing new questions?	Please order the following incentives according to how much they encourage you to correctly answer to the questions [User ranking in the second scenario, Gaming in the third scenario, Monetary Betting in the fourth scenario]	I would try to correctly answer the questions at my best even in the first scenario	Would you consciously provide false/incorrect information in the forth scenario just to benefit of the compensation?	Would you consciously provide false/incorrect information in the forth scenario just to benefit of the compensation knowing that the website uses automatic and/or manual mechanisms to avoid cheating?
48	No, in any case	2,3,1	Agree	Maybe	Maybe
49	No, in any case	2,3,1	Agree	Maybe	Maybe
50	No, in any case	2,3,1	Agree	Maybe	Maybe
51	No, in any case	1,3,2	Agree	Maybe	Maybe

Bibliography

[1] aBitAbout (2011), 'List of crowdsourcing projects', retrieved June 2011 from URL http://abitabout.com/List+of+crowdsourcing+projects

[2] von Ahn, L. & Dabbish, L. (2004), 'Labeling images with a computer game', *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 319-326

[3] von Ahn, L.; Kedia, M. & Blum, M. (2006), 'Verbosity: a game for collecting common-sense facts', *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 75-78

[4] Albors, J.; Ramos, J. & Hervas, J. (2008), 'New learning network paradigms: Communities of objectives, crowdsourcing, wikis and open source', *International Journal of Information Management* **28**(3), pp. 194-202

[5] Alonso, O.; Rose, D. E. & Stewart, B. (2008), 'Crowdsourcing for relevance evaluation', *SIGIR Forum* **42**, pp. 9-15

[6] Andreevskaia, A. & Bergler, S. (2008), 'When specialists and generalists work together: overcoming domain dependence in sentiment tagging', *Proceedings of ACL08 HLT* (6), pp. 290-298

[7] Anholt, S. (2006), *Competitive Identity: the new brand management for nations, cities and regions*, Palgrave Macmillan, United Kingdom

[8] Aue, A. & Gamon, M. (2005), 'Customizing sentiment classifiers to new domains: A case study', *Proceedings of Recent Advances in Natural Language Processing (RANLP)* **49**(2)

[9] Balahur, A. & Montoyo, A. (2009), 'A semantic relatedness approach to classifying opinion from Web reviews', *Procesamiento del lenguaje natural* ***42***, pp. 47-54

[10] Bayn, D. (2008), 'On the Wisdom of Crowds as a general critique and as it applies to social software', available at URL http://knol.google.com/k/daniel-bayn/on-the-wisdom-of-crowds/wxxygttievcn/7

[11] Bonaccorsi, A. & Rossi, C. (2004), 'Altruistic individuals, selfish firms? The structure of motivation in Open Source software', *First Monday* **9**(1)

[12] Brabham, D. C. (2008a), 'Crowdsourcing as a Model for Problem Solving: An Introduction and Cases', *Convergence* **14**, pp. 75-90

[13] Brabham, D. C. (2008b), 'Moving the crowd at iStockphoto: The composition of the crowd and motivations for participation in a crowdsourcing application', *First Monday* **13**

[14] Brabham, D. C. (2009), 'Moving the Crowd at Threadless: Motivations for Participation in a Crowdsourcing Application', *AEJMC conference August 2009* **13**(8), pp. 1-16

[15] Brabham, D. C. (2010), 'Crowdsourcing as a model for problem solving: Leveraging the collective intelligence of online communities for public good',

Unpublished doctoral dissertation, University of Utah

[16] Bretzke, H. & Vassileva, J. (2003), 'Motivating Cooperation on Peer to Peer Networks', *in* Brusilovsky, P.; Corbett, A. & de Rosis, F. ed., *User Modeling*, Springer, Berlin, pp. 148-148

[17] Buecheler, T.; Sieg, J. H.; Füchslin, R. M. & Pfeifer, R. (2011), *Crowdsourcing, Open Innovation and Collective Intelligence in the Scientific Method: A Research Agenda and Operational Framework*, MIT Press, Cambridge, pp. 679-686

[18] Calder, B. J. & Staw, B. M. (1975), 'Self-perception of intrinsic and extrinsic motivation', *Journal of Personality and Social Psychology* **31**(4), pp. 599-605

[19] Chanal, V. & Caron-Fasan, M.-L. (2008), 'How to invent a new business model based on crowdsourcing : the Crowdspirit ® case', *Actes de la Conference Internationale de Management Strategique,* Nice, pp. 1-27

[20] Chandler, D. & Kapelner, A. (2010), 'Breaking monotony with meaning: Motivation in crowdsourcing markets', *University of Chicago Mimeo*

[21] Cheng, R. & Vassileva, J. (2005a), 'Adaptive Reward Mechanism for Sustainable Online Learning Community', *Proc. of the International Conference on Artificial Intelligence in Education*, pp. 152-159

[22] Cheng, R. & Vassileva, J. (2005b), 'User Motivation and Persuasion Strategy for Peer-to-Peer Communities', *Hawaii International Conference on System Sciences* **7**, pp. 193-201

[23] Chesbruugh, H. W. (2003), 'The era of open innovation', *MIT Sloan Management Review* **44**(3), pp. 34-41

[24] Coar, K. (2006), 'The Open Source Definition', *Open Source Initiative (OSI),* available at URL http://www.opensource.org/docs/definition.php

[25] Crowdsourcing.org (2011), 'Directory of Sites', retrieved June 2011 from URL http://www.crowdsourcing.org/directory

[26] Enkel, E. ; Gassmann, O. & Chesbrough, H. (2009), 'Open R&D and open innovation: exploring the phenomenon', *R&D Management* **39**(4), pp. 311-316

[27] Erickson, T. (2011), 'Some Thoughts on a Framework for Crowdsourcing', *CHI 2011 Workshop on Crowdsourcing and Human Computation*

[28] Frey, B. S. (1997), *Not Just for the Money: An Economic Theory of Personal Motivation*, Edward Elgar Publishing

[29] Frei, B. (2009), *Paid Crowdsourcing: current state & progress toward mainstream business use*, produced by Smartsheet.com

[30] Ghosh, R. A. (1998), 'FM Interview with Linus Torvalds: What motivates free software developers?', *First Monday* **3**(3)

[31] von Hippel, E. & von Krogh, G. (2006), 'Free revealing and the private-collective model for innovation incentives', *R&D Management* **36**(3), pp. 295-306

[32] Holmes, N. (2006), 'Collecting the Wisdom of the Crowds', *Business Week Magazine* **39**(6)

[33] Howe, J. (2006), 'The Rise of Crowdsourcing', *Wired* **14**(6), retrieved June 2011 from URL http://www.wired.com/wired/archive/14.06/crowds.html

[34] Howe, J. (2008), *Crowdsourcing*, Crown Publishing Group, New York

[35] Hsueh, P. Y.; Melville, P. & Sindhwani, V. (2009), 'Data quality from crowdsourcing: a study of annotation selection criteria', *Proceedings of the NAACL HLT 2009 Workshop on Active Learning for Natural Language Processing', Association for Computational Linguistics*, pp. 27-35

[36] Ipeirotis, P. G. (2010a), 'Demographics of Mechanical Turk', *SSRN eLibrary*

[37] Ipeirotis, P. G. (2010b), 'Analyzing the Amazon Mechanical Turk marketplace', *XRDS* **17**(2), pp. 16-21

[38] Jeong, Y.; Kim, Y.; Kim, S. & Myaeng, H. (2009), 'Generating and mixing feature sets from language models for sentiment classification', *Natural Language Processing and Knowledge Engineering*, pp.1-8, 24-27

[39] Jones, Q. & Rafaeli, S. (1999), 'User population and user contributions to virtual publics: a systems model', *Proceedings of the international ACM SIGGROUP conference on Supporting group work*, New York, pp. 239-248

[40] Kittur, A.; Chi, E.; Pendleton, B.; Suh, B. & Mytkowicz, T. (2007), 'Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie', *World Wide Web* **1**(2)

[41] Kittur, A. & Kraut, R. E. (2008), 'Harnessing the wisdom of crowds in wikipedia: quality through coordination', *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, pp. 37-46

[42] Kolbitsch, J. & Maurer, H. (2006), 'Community Building around Encyclopaedic Knowledge', *Journal of Computing and Information Technology* **14**(3), pp. 175–190

[43] Lakhani, K. R. & von Hippel, E. (2003), 'How open source software works: “free” user-to-user assistance', *Research Policy* **32**(6), pp. 923-943

[44] Lakhani, K. & Wolf, R. G. (2003), 'Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects', *Social Science Research Network*, MIT Sloan, pp. 1-27

[45] Lakhani, K. R.; Lohse, P. A.; Panetta, J. A. & Karim, C. (2006), *The Value of Openness in Scientific Problem Solving*

[46] Lambiotte, R. & Ausloos, M. (2006), 'Collaborative Tagging as a Tripartite Network', *Computational Science*, Springer Berlin, pp. 1114-1117.

[47] Ling, K.; Beenen, G. & Kraut, R. (2005), 'Using Social Psychology to Motivate Contributions to Online Communities', *Journal of Computer-Mediated Communication* **10**(4)

[48] Livraghi G. (2011), 'Dati sull’Internet in Italia', retrieved August 2011 from URL http://web.mclink.it/MC8216/dati/dati3.htm

[49] Lorenz, J.; Rauhut, H.; Schweitzer, F. & Helbing, D. (2011), 'How social influence can undermine the wisdom of crowd effect', *Proceedings of the National Academy of Sciences* **108**(22), pp. 9020-9025

[50] Malone, T. W.; Laubacher, R. & Dellarocas, C. (2009), 'Harnessing Crowds: Mapping the Genome of Collective Intelligence', *Elements*, SSRN, pp. 1-20

[51] Mason, W. & Watts, D. J. (2010), 'Financial incentives and the "performance of crowds"', *SIGKDD Explorer Newsletter* **11**, pp. 100-108

[52] Matsumoto, S.; Takamura, H. & Okumura, M. (2005), 'Advances in

Knowledge Discovery and Data Mining', *Lecture Notes in Computer Science* **3518**(2005), pp. 301– 311

[53] Nov, O. (2007), 'What motivates Wikipedians?', *Commun. ACM* **50**, pp. 60-64

[54] Osterloh, M. & Frey, B. S. (2000), 'Motivation, Knowledge Transfer, and Organizational Forms', *Organization Science* **11**, pp. 538-550

[55] Quinn, A. J. & Bederson, B. B. (2009), 'A Taxonomy of Distributed Human Computation', *Technical Report HCIL*, University of Maryland

[56] Pang, B.; Lee, L. & Vaithyanathan, S. (2002), 'Thumbs up?: sentiment classification using machine learning techniques', *EMNLP '02: Proceedings of the ACL-02 conference on Empirical methods in natural language processing*, pp. 79–86

[57] Peters, I. (2009), *Folksonomies. Indexing and Retrieval in Web 2.0*, Walter de Gruyter & Co., Hawthorne, USA

[58] Prendergast, C. (1999), 'The Provision of Incentives in Firms', *Journal of Economic Literature* **37**(1), pp. 7-63

[59] Rockart, J. F. (1979), 'Chief executives define their own data needs.', *Harvard Business Review* **57**(2), pp. 81-93.

[60] Sharratt, M., & Usoro, A. (2003), 'Understanding Knowledge-Sharing in Online Communities of Practice', *Electronic Journal of Knowledge Management* **1**(2), pp. 187-197

[61] Shaw, A. D.; Horton, J. J. & Chen, D. L. (2011), 'Designing incentives for

inexpert human raters', *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, New York, pp. 275-284

[62] Schenk, E. & Guittard, C. (2011), 'Towards a characterization of crowdsourcing practices', *Journal of Innovation Economics*, **7**(1), pp. 93-107

[63] Sinha, R. & Swearingen, K. (2001), 'Comparing Recommendations Made by Online Systems and Friends', *Proceedings of the DELOS-NSF Workshop on Personalization and Recommender Systems in Digital Libraries*

[64] Spangler, S.; Proctor, L. & Chen, Y. (2008), 'Multi-taxonomy: Determining perceived brand characteristics from Web data', *Proceedings of IEEE/WIC/ACM Int. Conf. WI-IAT* **1**, pp. 258-259

[65] Steyvers, M.; Lee, M.; Miller, B. & Hemmer, P. (2009), 'The Wisdom of Crowds in the Recollection of Order Information', *Twenty-Third Annual Conference on Neural Information Processing Systems*

[66] Surowiecki, J. (2004), *The Wisdom of Crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations*, Doubleday, New York

[67] Vul, E. & Pashler, H. (2008), 'Measuring the Crowd Within: Probabilistic Representations Within Individuals', *Psychological Science* **19**(7), pp. 645-647

[68] Whitelaw, C.; Garg, N. & Argamon, S. (2005), 'Using appraisal groups for sentiment analysis', *CIKM '05: Proceedings of the 14th ACM international conference on Information and knowledge management*, pp. 625–631

[69] Whitla, P. (2009), 'Crowdsourcing and Its Application in Marketing

Activities', *Contemporary Management Research* **5**, 15—28

[70] Wikipedia (2011), 'List of crowdsourcing projects', retrieved June 2011 from URL http://en.wikipedia.org/wiki/List_of_crowdsourcing_projects