ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

A REFERENCE MODEL PROPOSAL FOR CROWDSOURCING AS A SERVICE

M.Sc. THESIS

Arbër MURTURI

Department of Computer Engineering

Computer Engineering Graduate Programme

AUGUST 2015

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

A REFERENCE MODEL PROPOSAL FOR CROWDSOURCING AS A SERVICE

M.Sc. THESIS

Arbër MURTURI
(504091530)

Department of Computer Engineering

Computer Engineering Graduate Programme

Thesis Advisor: Prof. Dr. Sema Oktuğ

AUGUST 2015

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

BİR KİTLE-KAYNAK SERVİSİ İÇİN REFERANS MODEL ÖNERİSİ

YÜKSEK LİSANS TEZİ

Arbër MURTURI
(504091530)

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Yüksek Lisans Programı

Tez Danışmanı: Prof. Dr. Sema Oktuğ

AUGUST 2015

Arbër Murturi, a M.Sc. student of ITU Department of Computer Engineering student ID 504091530, successfully defended the thesis entitled "A REFERENCE MODEL PROPOSAL FOR CROWDSOURCING AS A SERVICE ", which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Prof. Dr. Sema Oktuğ** İstanbul Technical University	.....
Jury Members :	**Doç. Dr. Feza BUZLUCA** İstanbul Technical University	.....
	Doç. Dr. Haluk BİNGÖL Boğaziçi University	.....

Date of Submission : 04.05.2015
Date of Defense : 19.08.2015

To my father,

FOREWORD

From the autum of 2014 to the spring of 2015, the process of writing this thesis has been a truly amazing and challenging journey, during which I have received tremendous support and guidance from many great individuals.

First and foremost, I would like to express my genuine gratitude towards my thesis supervisor Professor Sema Oktuğ. Professor Oktuğ, firstly taught computer networks and than introduced me to the field of social networks and especially the subject of crowdsourcing. Professor Oktuğ, guided me in choosing an interesting research topic concerning crowdsourcing, and always had the time to help me, comment on as well as improve my research. I have the utmost respect for you. You have taught me a lot through our numerous discussions, especially regarding conducting academic research. You were at the same time extremely challenging, supportive and inspirational, which enabled me to experience great learning curves in the process, I sincerely thank you for all that you have done for me.

I want to thank my family especially my father and mother for their support in every aspect of life. I am greatfull for every thing you have done for me. Finally, to my best friend Katya Kaya "Thank You" is just not enough for the support you have given to me.

August 2015

Arbër MURTURI
SAP Senior Technical Consultant

TABLE OF CONTENTS

Page

ABBREVIATIONS

ALS	**:** Amyotrophic Lateral Sclerosis
AMT	**:** Amazon Mechanical Turk
ATM	**:** Automatic Teller Machine
CG	**:** Control Group
CO_2	**:** Carbon Dioxide
GPS	**:** Global Positioning System
GWAP	**:** Games with a Purpose
HIT	**:** Human Intelligence Tasks
IaaS	**:** Infrastructure as a Service
ICT	**:** Information and Communication Technology
ID	**:** Identity
LTE	**:** Long-Term Evolution
MD	**:** Majority Decision
PaaS	**:** Platform as a Service
PC	**:** Personal Computer
R&D	**:** Research and Development
SaaS	**:** Software as a Service
SCOUT	**:** Super Contributor Outlier
SRM	**:** Single Round Match
WiFi	**:** Wireless Fidelity
XaaS	**:** Everything as a Service

LIST OF TABLES

Page

LIST OF FIGURES

Page

A REFERENCE MODEL PROPOSAL FOR CROWDSOURCING AS A SERVICE

SUMMARY

Crowdsourcing is a distributed problem-solving model in which a crowd of undefined size is engaged to solve a basic or a complicated problem through an open call. This thesis gives definitions, importance, advantages and disadvantages of crowdsourcing. It presents the traditional system overview of crowdsourcing. Moreover, popular crowdsourcing applications are described and analyzed.

Our aim is to present the important issues in crowdsourcing and how those are realized so far. Components and activities within crowdsourcing process are identified. We extract all known components and properties of crowdsourcing applications and design a reference model based on these features for cloud systems. The reference model is outlined in four phases. In each phase, we present interactions between entities and give a clear picture of crowdsourcing working logic.

Thesis also studies in detail all factors that have effect on the crowdsourcing process. It explores how companies, organizations or individuals leverage the latest internet technologies to build applications that attract the crowd. Participation of crowd as one of the most important factor in the success of crowdsourcing platforms is explored. Task design process, types and features of tasks are identified by studying applications developed so far. In addition, the thesis shows how tasks should be designed to accomplish communication among the crowd or with the requester efficiently. Tasks manager is also discussed, which is responsible to control and distribute tasks in the crowdsourcing platform.

The motivation behind crowdsourcing, both from the company's and the crowd's perspectives are investigated. It is easy to see why companies want to adopt crowdsourcing, however, it is hard to explain why so many people are willing to spend their time on activities that they will pay low (or even none).

As cloud-based services have become widely adopted, a cloudified reference model has been emergent for crowdsourcing platforms and applications. This thesis, for the first time, introduces a cloudified, four-phase reference model for crowdsourcing along with a generic workflow for crowdsourcing development utilizing the facilities offered by cloud service providers. Moreover, useful insights are presented for the evolution of today's online crowdsourcing applications and platforms towards the concept of crowdsourcing as a service.

The detailed reference model introduced in the thesis will be helpful to show directions to the crowdsourcing platform/application developers. This research aims to contribute for a better understanding of the crowdsourcing process.

BİR KİTLE-KAYNAK SERVİSİ İÇİN REFERANS MODEL ÖNERİSİ

ÖZET

Kitle-Kaynak, açık çağrı aracılığıyla ilgili platformlar ile (web ya da mobil uygulamalar kullanılarak) kitlelerden fikir alma veya belli sorunları (işleri) kitlelere yaptırma yöntemidir.

Kitle-Kaynak günümüzde giderek yaygınlaşan bir olgudur. Şirketler, araştırmacılar ve kurumlar faaliyetlerini bu alanda gerçekleştirmektedir. Bu faaliyetler (işler/görevler) bilimsel problemlerin çözümlenmesinden, tekrarlamalı ve sıkıcı görevlere kadar sıralanmaktadır.

Kitle-Kaynak, kitle tarafından fikir almak (yardım almak) için en etkili ve en yaygın kullanıma sahip yöntemdir. Teknolojinin günden güne gelişmesine rağmen, hala bilgisayarlaştırılmada zorluk yaşanılan farklı sorunlar (görüntü açıklama, görüntü sınıflandırılması, metin açıklama, görüntü tanıma, yazılım geliştirme, çevresel ve sağlık sorunları, vb.) yer almaktadır. Bu görevlerin bilgisayarlarla çözülmesi zor olduğundan dolayı insan desteği gereklidir. Kitle-Kaynak kavramı bu tür sorunların üstesinden gelmek için kullanılmaktadır. Geleneksel olarak, şirketler, bu gibi görevleri dış kaynaklı başka şirketlere veya profesyonel bireyleri kiralayarak yaptırmaktadır. Kitle Kaynak, dağıtılmış görev tamamlama modeli sayesinde, dış kaynaklı şirketlere iş yaptırma modelinin yerine geçmektedir. Kitle Kaynak modeli içerisinde görevlerin dağıtımının uzmanlar ya da seçilmiş adaylarla sınırlı olmadığının vurgulanması önemlidir. Kitle Kaynak uygulamalarında, kitlelerin verilen görev veya proje üzerinde eş zamanlı olarak çalışması mümkündür. Bu platform, mevcut görevlerin yer aldığı listeyi, işverenler (Crowdsourcer) tarafından verilen ödül ve süre ile ilişkilendirerek sunulmaktadır. Bu süre zarfında, çalışanlar en iyiyi sunmak için yarışmaktadırlar.

Çalışan, ödülü kazanmak için, görev listesi içerisinden görev seçerek tamamlamaktadır. Süre sonunda, gönderimlerden doğru olanlar seçilerek, işverenler tarafından ilgili çalışanlara ödülleri verilmektedir. Çalışan yaptığı işin, işveren tarafından kabul edilmesiyle ödüle ek olarak güvenilirlik kazanmaktadır. Bazen işveren, görevi belirtilen ihtiyaçlar doğrultusunda yerine getiren her bir çalışana para ödemek zorunda kalabilir.

Bazı durumlarda, çalışanlar verilen ödüllerle motive olmamaktadır fakat eğlenmek ya da yardım sever olarak çalışmaktadırlar. Sonuç olarak işveren, en iyi şekilde ihtiyaçlarının karşılandığı sonucu seçecektir. Kitle Kaynak, işveren (şirket, araştırmacılar, vb.) için azımsanmayacak ölçüde yarar sağlamaktadır. Başarısızlık riskinin göz ardı edilmesi ile, ürün ya da servisler için ödeme yapılarak beklentilerin karşılanması mümkündür.

Çalışmada bulut sistemlerde kullanılabilecek bir Kitle Kaynak sistem mimarisi önerilmiştir. Kitle Kaynak ile ilgili tanımlardan ve öneminden bahsedilerek avantaj ve dezajavantajları tartışılmıştır. Günümüzde Kitle Kaynak esasları ile çalışan

yüzlerce uygulama mevcuttur. Bu uygulamaların birçoğu özel ihtiyaçlara göre tasarlanmıştır. Uygulamaların bazıları sadece özel bir iş/görev için tasarlanmış olup bazıları ise birden fazla farklı türde görevlerin tanımlanabildigi sistem mimarisini desteklemektedir. Kitle Kaynak alanında bazı popüler uygulamaların çalışma mantığı tezde örnekler halinde sunulmuştur.

Araştırmacılar son birkaç yıldır Kitle Kaynak alanına dahil konular üzerine çalışmalarını sürdürerek çeşitli sayıda uygulamalar geliştirmişlerdir. Şimdiye kadar Kitle Kaynak bileşenleri hakkında ya da belirli sorunlar hakkında çalışma olmasına rağmen, Kitle Kaynak yöntemi için genel bir referans modeli üzerinde çok az çalışma yapılmıştır. Bu konudaki çalışmaların yetersiz olmasından dolayı, Kitle Kaynak alanında yapılan geliştirme çalışmalarında geçmişten süre gelen bazı sorunlar devam etmektedir.

Bu çalışmadaki amacımız Kitle Kaynak içerisindeki önemli konuları ve şimdiye kadar yapılan çalışmaların nasıl gerçekleştirildiğini ortaya koymaktır. Referans modeli oluşturmaktaki en büyük sorunlardan biri çok sayıda farklı alanlarda çalışabilen Kitle Kaynak uygulaması olmasıdır. Uygulamaların genel özelliklerinin, bileşenlerinin ve bileşenler arasındaki etkileşimin tanımlanması çalışmanın temel öğelerini oluşturmaktadır. Tez içerisinde bu özellikler baz alınarak Kitle Kaynak için yeni bir referans modeli tasarlanmıştır. İlgili bileşenler ve nitelikleri de ayrıntılı olarak tanımlanmıştır.

Çalışma içerisinde, Kitle Kaynak süreci dört aşamaya ayrılarak uygulamaların arkasındaki çalışma mantığı sunulmuştur.

İlk aşama, işveren (crowdsourcer) ve çalışanın (crowd) sisteme kayıt sürecini içermektedir. Kayıt sürecinin çeşitliliği ve kimlik bilgi doğrulama bölümü detaylı olarak anlatılmıştır. Bu aşamada işveren görevleri oluşturarak Görev Ambarına (Task Storage) koymaktadır. Görevler açık çağrı (Open Call) olarak çalışanlara duyurulmaktadır.

İkinci aşama, Görev Yöneticisini (Task Manager) içermektedir. Görev dağıtımı, listeleme ve atama gibi görev yöneticisinin işlevleri sunulmuştur.

Üçüncü aşama, görev dağıtımından sonra çalışanların belli bir süre içerisinde görevi tamamlayıp sonuçları Değerlendirme Birimine (Evaluation Engine) göndermesini içermektedir. Değerlendirme Birimi, çalışanlar tarafından gelen sonuçları çalışma içerisinde sunulmuş değerlendirme yöntemlerini kullanarak değerlendirmeyi gerçekleştirmektedir. Bu birimin sonucuna göre çalışanın yapmış olduğu çalışmanın kalite kotrolü yapılır.

Dördüncü aşama ise yapılan değerlendirmenin ardından, Sıralama ve Ödeme aşamasını içermektedir. Bu aşamada çalışanın başarı yüzdesi hesaplanmaktadır. Başarı yüzdesi çalışanın sıralamasını belirlemektedir. Sıralama, daha sonraki görevlere Görev Yöneticisi tarafından atanabilmesi için önemli bir veri olarak kullanılmaktadır. Görevin başarılı olarak tamamlandığına dair onay verildikten sonra ödeme yapılmaktadır.

Referans modelinin tasarlanması için Kitle Kaynak bileşenleri ve özellikleri hakkında detaylı bilgiye gerek vardır. Bu tezin amaçlarından biri de Kitle Kaynak süreci ve referans model tasarımı içerisindeki bileşenlerin ve aralarındaki etkileşimlerinin detaylı bir şekilde tanımlanmasıdır. Çalışmamızda, geliştirilmiş mevcut Kitle Kaynak uygulamalarındaki bileşenler ve özellikleri kullanılarak referans modeli tasarlanmıştır. Çalışmada Kitle Kaynağın tüm öğeleri detaylı bir

şekilde incelenmiştir. Genel olarak özel bir bileşenin performansını geliştirmek veya tamamen tek amaçlı yapılan geliştirmeler nedeniyle Kitle Kaynak uygulamaları için genel bir yapının (referans modelinin) oluşturulmasında zorluklar yaşanmıştır.

Bu tez çalışması, kitle kaynak hakkında yayınlanan kitap, makale ve blog yazılarını kapsayan mevcut literatür çalışmalarını içermektedir. Bu da bize referans modelinin, şimdiye kadar Kitle Kaynak alanında tamamlanmış tüm görevlere dayanarak tasarlanmasında yardımcı olmuştur.

Bulut-tabanlı servislerin yaygınlaşmasıyla birlikte bulutlaştırılmış Kitle-Kaynak referans modelinin tasarlanması bir ihtiyaç haline gelmiştir. Bu tezde ilk kez kapsamlı iş akışı ile bulutlaştırılmış dört aşamalı Kitle-Kaynak referans modeli sunulmuştur. Bununla birlikte bugün çevrimiçi çalışmakta olan Kitle-Kaynak uygulamaları ve platformları hizmet tabanlı Kitle-Kaynak kavramı kapsamında sunulmuştur.

Sonuç olarak çalışmada sunulan detaylandırılmış referans modeli, Kitle Kaynak uygulama/platform geliştiricilerine yönlendirme konusunda yardımcı olacağını umuyoruz. Geliştiriciler bu çalışmayı dikkate alarak uygulamalarını/platformlarını Kitle Kaynak özelliklerine ve niteliklerine bağlı olarak daha verimli çalışabilecek bir şekilde tasarlayabilirler.

1. INTRODUCTION

It has been several years that researchers are working on different fields of crowdsourcing. There has been various number of applications developed. This thesis introduces a new reference model for crowdsourcing, where the related components and attributes described in detail. Moreover, useful insights are presented for the evolution of today's online crowdsourcing applications and platforms towards the concept of crowdsourcing as a service.

Developers can define and design their applications/platforms based on the properties and attributes of the crowdsourcing model described in thesis. Here, the crowdsourcing process is studied and presented with generic workflow for crowdsourcing development utilizing the facilities offered by cloud service providers.

1.1 Thesis Structure

This thesis consists of five chapters, which are described in the following paragraphs for better illustration of the structure.

Chapter 1 gives an introduction and motivation of the research. Social aspects of crowdsourcing are discussed in this chapter.

Chapter 2 focuses on reviewing the literature of crowdsourcing. Definitions and traditional system overview of crowdsourcing is presented. Importance, advantages and disadvatages of crowdsourcing are covered also in this chapter.

Chapter 3 presents crowdsourcing applications implemented by researchers, organizations and companies. Web based and mobile application samples are described.

Chapter 4 presents a reference model proposal, which is framework of this research. In this chapter, we describe process of crowdsourcing sperately in phases. This chapter also outlines all entities of crowdsourcing model in detail. In addition, an applicability of the reference model in cloud-centric architecture is shown.

Chapter 5 presents the research summary, discusses future work as well as suggestions for further research.

1.2 Motivation of Research

Crowdsourcing is an increasingly popular phenomenon where companies, researchers, organizations are involved on accomplishing activities in this field. These activities can range from solving scientific problems to repetitive and boring tasks.

Although there are many works done in field of crowdsourcing studying separately components or specific issues until now, but there is few work done on designing a reference model for crowdsourcing process as service which can run on cloud architecture. Lack of work done on this direciton, have motivated us to study deeply literature and design a reference model for crowdsourcing.

Designing a reference model requires knowledge of components and properties of crowdsourcing. The main research objective of thesis is to identify the components and activities within crowdsourcing process and design a reference model. Extracting all known components and properties of such applications and design a reference model based on these features. This thesis analyzes in detail all entities of crowdsourcing.

The research consists of studying existing literature on crowdsourcing, including books, papers, articles and blogs. This helped us to design a proposal model based on all work done until now in field of crowdsourcing.

We hope that detailed reference model introduced in the thesis will be helpful to show directions to the crowdsourcing platform/application developers. This research aims to contribute for a better understanding of crowdsourcing process.

1.3 Social Aspects

The social structure denotes a set of relationships that occur among individuals involved in pursuing a goal (for instance, the boss-collaborator relationship, collaboration among community members, and so on). Social norms have a strong influence on the channels of communication, coordination mechanisms, beliefs and views, feelings, and motivations that affect these relationships.

In a crowdsourcing initiative, the actors include both the corwdsourcer (usually called the initiator) and the contributing crowd. Many such initiatives exploit a peer community in which the hierarchy is neutral; relationships among crowd users don't heavily depend on what role individuals have but rather on their reputation in the group. To ensure participation at a sustainable level and maintain performance (both in quantity and quality), the application designer must identify members' contribution style, past performance and practices, system of values, and the set of rewards that better suits their needs.

Crowdsourcing can also occur in a corporate environment. In such cases, the social structure is usually hierarchical, and workers interact at various levels. Typically, workers must deal with a supervisor, and their reputations depend on this person's feedback. The supervisor has a strategic role in the company because he or she can communicate the aims and expectations of those at a higher level to lower levels, support goal clarification, control and manage constraints, recognize positive or negative behaviors, and reward or punish people for their performance. In this context, the so-called principal-agent problem might occur. This happens when the supervisor (the principal) delegates a job to a worker (the crowd user) who performs it. Typically, this is the employer-employee relationship in which a principal hires an agent to pursue a specific interest. In a perfect situation, the agent acts exactly as the principal wants, but often the agent's interests do not converge with the principal's. In this case, the principal must provide incentives to align the agent's interests with the company's. Various mechanisms are available for this, such as periodical work assessments, payment based on piece rates, discretionary bonuses, promotions, profit sharing, and deferred compensation [1].

2. CROWDSOURCING

2.1 What is Crowdsourcing?

In Wired Magazine, the term “Crowdsourcing” first was coined in article in June by Jeff Howe (2006) [2]. The article describes an emerging trend where companies start engaging the public in helping perform activities such as content creation and problem solving. The term was intended to be wordplay on outsourcing and so it wasn't defined in the article. As people started referring to this term in a loosely defined way, Howe decided to offer a formal definition on his blog. In this thesis, the usage of this term will be consistent with Howe's definition: "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".

2.2 Definitions of Crowdsourcing

Crowdsourcing has been adapted to be used as an effective and powerful practice, however, it is difficult to be defined and categorized, and thus varying definitions of crowdsourcing exist. Ever since the introduction of crowdsourcing’s concept in 2006, scholars and researchers have given crowdsourcing a number of definitions, and a collection of these definitions is listed in the following in chronological order.

According to Howe (2006), “a web based business pattern, which makes best use of the individuals on the Internet, through open call, and finally gets innovative solutions” (p.1).

According to Brabham (2008), “... an online, distributed problem-solving and production model already in use by for profit organizations” (p.75).

According to Brabham in (2008), “... a strategic model to attract an interested, motivated crowd of individuals capable of providing solutions superior in quality and quantity to those that even traditional forms of business can” (p.79).

According to Chanal and Caron-Fasan (2008), "... the opening of the innovation process of a firm to integrate numerous and disseminated outside competencies through web facilities. These competences can be those of individuals or existing organized communities" (p.5).

> According to Kleeman (2008), ... a profit-oriented form outsources specific tasks essential for the making or sale of its product to the general public (the crowd) in the form of an open call over the internet, with the intention of animating individuals to make a contribution to the firm's production process for free or significantly less than that contribution is worth to the firm. (p.6)

According to Palantino and Vojnovic (2009), "... methods of soliciting solutions to tasks via open calls to large-scale communities" (p.1).

According to Vukovic (2009), "... new online distributed problem-solving and production model in which networked people collaborate to complete a task" (p.1).

According to Vukovic (2009), "... a new online distributed production model in which people collaborate and may be awarded to complete task" (p.539).

According to Whitla (2009), "... a process of outsourcing of activities by a firm to an online community or crowd in the form of an 'open call" (p.15).

According to Heer and Bostock (2010), "... a relatively new phenomenon in which web workers complete one or more small tasks, often for micro-payments on the order of $0.01 to $0.10 per task" (p.1).

According to Buecheler (2010), "... one way for a firm to access external knowledge" (p.1).

According to La Vecchia and Cisternino (2010), "... a tool for addressing problems in organizations and business" (p.425).

According to Ling (2010), "... a new innovation business model through the internet" (p.1).

> According to Mazzola and Distefano (2010), ... an intentional mobilization, through Web 2.0, of creative and innovative ideas or stimuli, to solve a problem, where voluntary users are included by a firm within the internal problem- solving process, not necessarily aimed to increase profit or to create product or market innovations, but in general, to solve a specific problem. (p.3)

According to Alonso and Lease (2011), "... the outsourcing of tasks to a large group of people instead of assigning such tasks to an in- house employee or contractor" (p.1).

According to Doan (2011), "... a general-purpose problem-solving method" (p.2).

According to Heymann and Garcia-Molina (2011), "... getting one or more remote internet users to perform work via a marketplace" (p.1).

According to Kazai (2011), "... an open call for contributions from members of the crowd to solve a problem or carry out human intelligence tasks, often in exchange for micro-payments, social recognition or entertainment value" (p.1).

According to Wexler (2011), "... focal entity's use of an enthusiastic crowd or loosely bound public to provide solutions to problems" (p.11).

According to Poetz and Schreier (2012), "... outsource the phase of idea generation to a potentially large and unknown population in the form of an open call" (p.4).

> According to Estellés-Arolas and González- Ladrón-de- Guevara (2012), a type of participative online activity in which an individual, an institution, a non-profit organization, or company proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task. The undertaking of the task, of variable complexity and modularity, and in which the crowd should participate bringing their work, money, knowledge and/or experience, always entails mutual benefit. The user will receive the satisfaction of a given type of need, be it economic, social recognition, self-esteem, or the development of individual skills, while the crowdsourcer will obtain and utilize to their advantage what the user has brought to the venture, whose form will depend on the type of activity undertaken. (p.197)

As it can be seen, researchers defined crowdsourcing based on different point of views and thus there is not an agreed definition. In this research, crowdsourcing is seen as a type of participative online activity in which an individual or an organization proposes to a group of individuals or organizations of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task. The undertaking of the task, of variable complexity and modularity, and in which the crowd should participate bringing their work, money, knowledge and/or experience, always entails mutual benefit. The user will receive the satisfaction of a given type of need, be it economic, social recognition, self-

esteem, or the development of individual skills, while the crowdsourcers will obtain and utilize to their advantage what the user has brought to the venture, whose form will depend on the type of activity undertaken.

2.3 The Importance of Crowdsourcing

The Internet is now a melting pot of user-generated content from blogs to Wikipedia entries to YouTube videos. The distinction between producer and consumer is no longer such a prevalent distinction as everyone is equipped with the tools needed to create as well as consume. As a business strategy, soliciting customer input is not new, and open source software has proven the productivity possible through a large group of individuals. Crowdsourcing is a powerful business-marketing tool as it allows an organization to leverage the creativity and resources of its own audience in promoting and growing the company for free.

Figure 2.1 : Importance of crowdsourcing.

From designing marketing campaigns, communication, collaboration to researching new products to solving difficult business roadblocks, an organization's consumers can likely provide important guidance and answers (see Figure 2.1). Moreover, best of all, all the consumer wants in return for their opinion and effort is some recognition or even a simple reward [3].

2.4 Advantages and Disadvantages of Crowdsourcing

Important advantage of crowdsourcing is that it provides immediate attention to and staffing for a current business need. Although crowdsourcing is often compared to outsourcing, it is altogether a different concept. When outsourcing, a company must make hiring decisions, allocate training resources and perhaps supplement a benefits package. With crowdsourcing, the forum – by definition – is open and voluntary. This provides lower overhead costs on a project and more agility in the problem-solving process.

Crowdsourcing increases the productivity of a company while minimizing labor expenses. The Internet is a time-proven strategy for soliciting feedback from an active and passionate consumer base. Customers today want to be involved in the companies they buy from, which makes crowdsourcing an incredibly effective tool.

A variety of websites allows companies to post their job, or challenge, and a number of people typically begin working on the assignment. Simple tasks such as providing feedback on a website layout and rating its user-friendly features, or describing merchandise for an online catalogue are common uses of crowdsourcing. Tasks that require highly sophisticated knowledge and intense time management can become a logistical burden and might not be ideal for crowdsourcing. Because participants are often in competition with one another for the work, there may not be great communication among participants without significant planning on the part of the organization providing the work. In addition, workers do not sign contracts so they may leave a project at any given time.

Despite these obstacles, crowdsourcing is an innovative new use of collaborative and creative talents and that potential is just uncovered. Planning and structuring a project with specific requirements and benchmark goals can lead to a more effective outcome [4].

Although, crowdsourcing offers so many benefits, then why has not it been more widely adopted?

Firstly, one of the main disadvantages of crowdsourcing is the quality of the work in general. The skill-level of the crowd is expectedly lower than that of the professionals and employees traditionally dedicated to the task. In addition, unlike supervised employees, the crowd usually experiences less pressure to perform high

quality work. The company may thus either receive low quality results or have to spend time reviewing the work. For instance, Frito-Lay invited Internet users to help craft its new advertisement through Yahoo. The winning submission was eventually shown during the 2007 super-bowl. Although the crowdsourcing was considered successful and resulted in an acceptable submission, it's hard to estimate the extra time and effort spent by the company in reviewing all these submissions. On the other hand, if Frito-Lay outsourced the task to an advertising agency, then it might have more control over the process and more confidence that an acceptable advertisement would be produced at the end.

Note that companies need not always review the work produced by the crowd themselves. Nowadays companies often ask the crowd to review the quality of their peers' work as well, and then only review the highly rated ones at the end. This is indeed much more efficient but then again the company faces the same issue of whether it can trust the crowd in helping it performs the task, in this case reviewing the quality of the work rather than producing the work itself.

The second challenge is that given the relatively low monetary rewards, how companies can encourage the crowd to help. To overcome this challenge, companies often tap into the crowd's internal motivations instead, like their craving for attention and entertainment. Companies often accomplish this by carefully building a community for the intended crowd. The community not only helps retain the crowd, but also serves as a platform whereby the crowd can satisfy their desires for attention and appreciation. However, with all the social networking and community sites out there, competing for the crowd's attention is increasingly harder. The company thus may have to invest extra resources and efforts to build the community beforehand while not being able to reap any results immediately.

The third challenge is that while the crowd may perform quick, short-term tasks effectively, they simply cannot depend on for long-term projects. There are several reasons. First of all, the crowd typically has no obligation to the company, and so is free to perform as much work as they please. It is rather infeasible to hope that the same person would be motivated to work for an extended period of time unless the reward for the task is extremely appealing. Secondly, while a long-term project can be broken down into multiple smaller tasks to make it easier to assign to the crowd, often times there are dependencies or common components to the tasks. The

company may face difficulty in coordinating the crowd in such cases. Thirdly, tasks are rarely isolated but often related to the existing systems or require some internal knowledge. Not only will training the crowd present a difficulty, but any knowledge and skill increase acquired from the task will be lost afterwards. Fourthly, while employees are expected to be present at work or at least easy to contact, the same is not true for the crowd at all. All these reasons impose limits on the sort of activities that can be effectively crowdsourced [5].

3. CROWDSOURCING APPLICATIONS

Crowdsourcing is a form of outsourcing tasks (jobs) to the crowd by means of open calls via related platforms (using web or mobile applications). As it was mention in earlier chapter, Crowdsourcing is one of the most effective and widely used approaches to capture ideas (get help) from the crowd. Although technology is improving day by day, there are still different types of problems that cannot computerize easily (like image annotation, image classification, text annotation, pattern recognition, software development, environmental and medical issues, etc.). These tasks are difficult to solve for computers, additional human work is needed. Concept of crowdsourcing is being used in order to deal with these problems. Traditionally, companies to hired professionals individuals to accomplish such task or outsource tasks to other companies. Crowdsourcing replaced this methodology with distributed task solving model. In crowdsourcing model, it is important to emphasize that the distribution of task is not limited to experts or preselected candidates [6]. First of all, lets take a look at the crowdsourcing traditional process overview.

3.1 Crowdsourcing Traditional Process Overview

Analyzing many works done before, is identified that crowdsourcing traditionally is comprised of four parts (see Figure 3.1). These four parts, or called four pillars [7]: Crowdsourcer, Crowd, Task and crowdsourcing Platform. In crowdsourcing applications crowds can work simultaneously on a given task, project, etc. The crowdsourcing platform exhibits a list of available tasks, associating with reward and time period, that are presented by requesters (crowdsourcer); and during the period, workers compete to provide the best submission. Meanwhile, a worker (crowd) selects a task from the task list and completes the task because the worker wants to earn the associated reward. At the end of the period, subsets of submissions are selected, and the corresponding workers are granted the reward by the requesters. In

addition to monetary reward, a worker gains credibility when their task accepted by the requester.

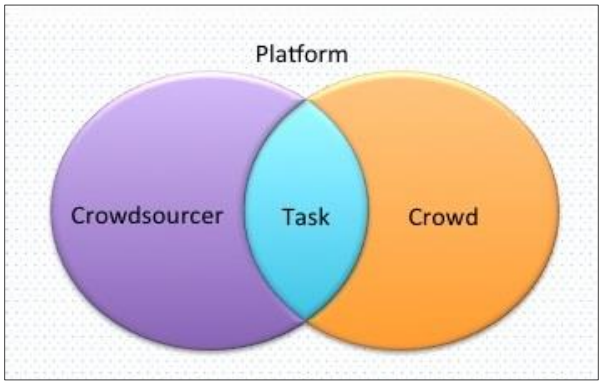

Figure 3.1 : Components of crowdsourcing.

For the Crowdsourcer (company, researcher, etc.) the benefit is substantial. It can externalize the risk of failure and it only pays for products or services that meet its expectations. Roughly, this is the concept of work behind most crowdsourcing systems.

Another key characteristic of crowdsourcing processes is whether the crowd's contribution is: participatory or opportunistic.

- Classical crowdsourcing services on the Web are participatory because they require users' active participation.
- Opportunistic crowdsourcing is data generated from sensors and mostly computations that are automatically performed by the crowd's devices - for example, trajectory matching and positional triangulation [8].

We can classify crowdsourcing applications into either of Web-based applications (desktop) or new applications (mobile). Web and mobile-based crowsourcing applications are presented in next sections.

3.2 Web Based Crowdsourcing Applications

We have studied crowdsourcing applications developed in various fields. There is a wide range of crowdsourcing applications on various platforms. Some crowdsourcing platforms are specialized on specific problems, and some are micro-task based web platforms.

3.2.1 Amazon's Mechanical Turk

MTurk is an Internet marketplace in which companies and computer programmers outsource simple tasks, and workers are free to choose which ones they want to perform. Workers are paid based on their performance. First, the goal is usually clear and requires a low level of participation. It is in the principal's interest to communicate and specify the goal. Second, the task has a low level of variety, specificity, and identification. The skills required to complete it are typically trivial for the community to which it is addressed. Figure 3.2 shows view of web site.

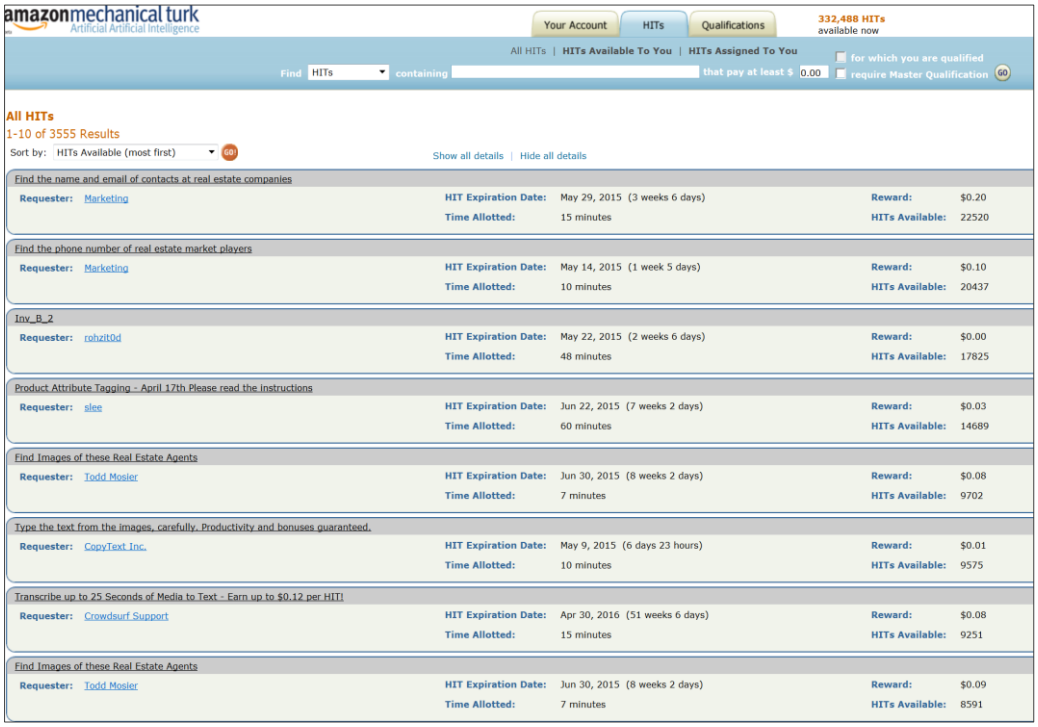

Figure 3.2: Amazon Mechanical Turk.

The social structure represents a typical principal-agent problem. The crowdsourcer (a principal) lists a task that one or many crowd user in the crowd can select. Although we can define the social structure as hierarchical, contributors are autonomous, essentially anonymous, and thus cannot benefit from building a reputation system using the platform. Finally, the nature of the good is private, in that the principal will appropriate it.

Because the tasks are simple and the performance is measurable, the initiator will pay for each task (pay per piece) performed and will invest few resources to communicate the goal. In addition, because the contributors are anonymous and the

level of payment is low, the turnover is high. This might affect result quality and preclude the crowd's participation in more complex tasks [1].

3.2.2 Threadless

Another interesting case is the Threadless crowdsourcing t-shirt company (www.threadless.com). On this platform, consumers are part of a community that designs t-shirts and votes on them. Any design with enough votes is offered in the site's store, and the designer gets a big payout (US$2,000). Here, people participate both for the monetary prizes and to demonstrate their skills to a community or an employer. Prizes are frequently used in situations where creativity is required.

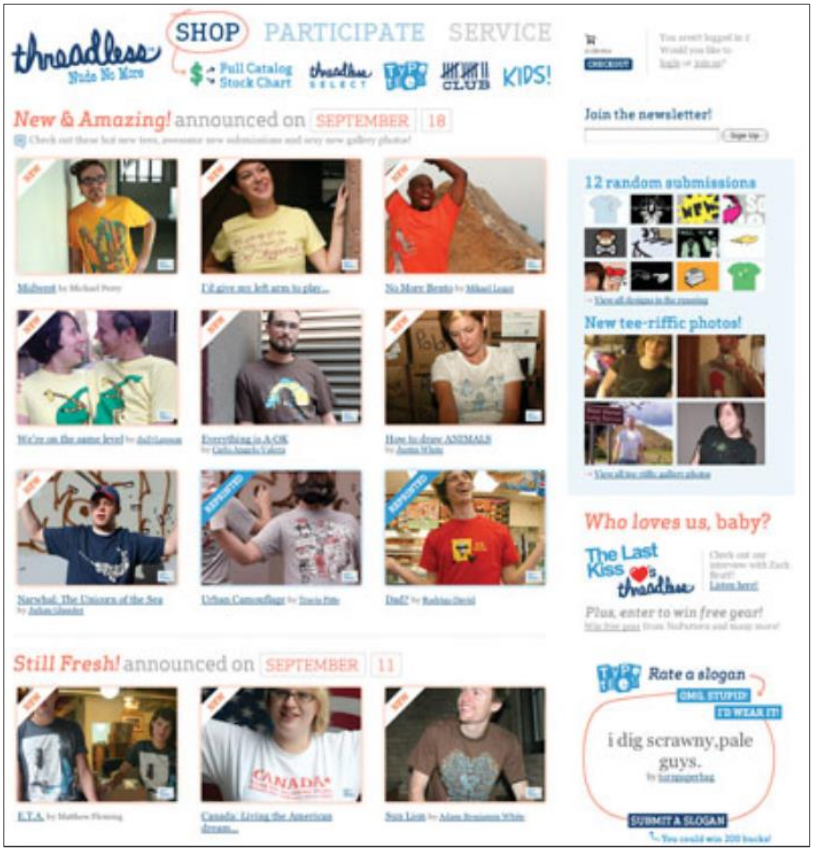

Figure 3.3: Threadless (t-shirt design).

The goal is usually clear and requires high participation. Communication about the goal is both high-level and specific. The task has high variety, specificity, and identification, and requires highly specific skills. The social structure is partly hierarchical but with some "democratic" features (for instance, users can vote on others' projects and contribute their own designs). Figure 3.3 shows view of web site.

Participants in Threadless have stated that they like the idea of community but that they can also make money, develop their creative skills, or take up freelance work [9].

3.2.3 Galaxy Zoo and Moon Zoo

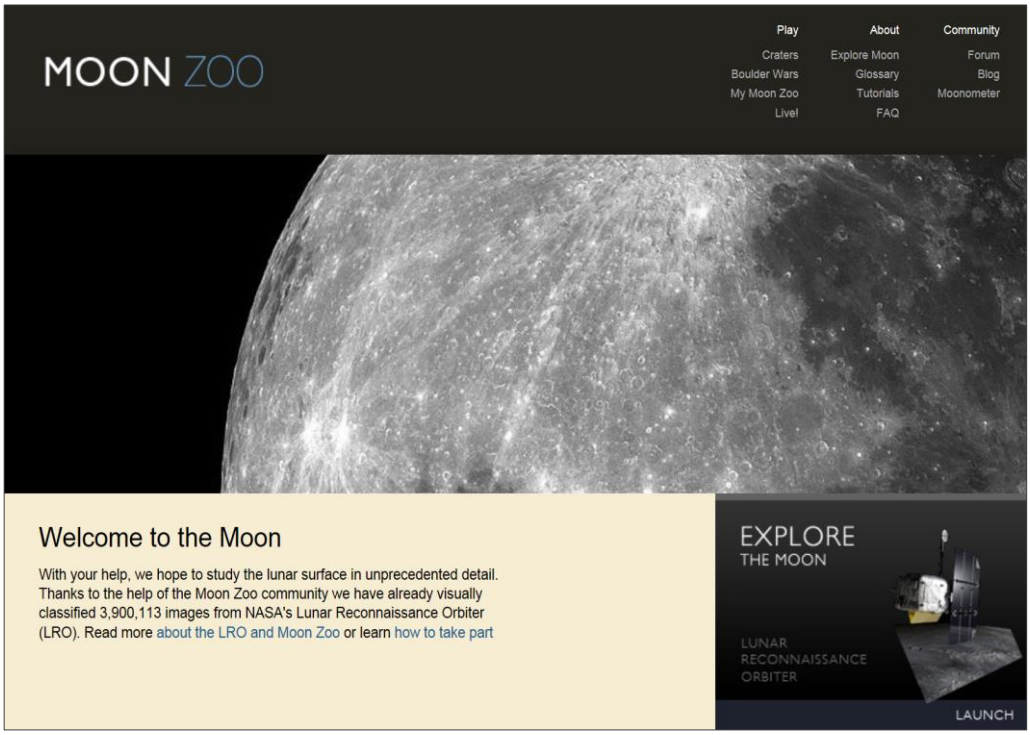

Figure 3.4 : Moon Zoo.

Some crowdsourcing platforms thus divide the complex task into simple activities to enable the crowd to perform without extensive training. These activities require a large investment from the crowdsourcer to decompose the task and design a user-friendly platform. The problem is less critical if a huge number of contributors produce the good, the task is simple and requires common skills, or the crowdsourcing is related to a cause. Figure 3.4 shows view of web site. The Galaxy Zoo (www.galaxyzoo. org) and Moon Zoo (www.moonzoo.org) projects are good examples of this type of situation. Galaxy Zoo is a crowdsourcing project that aims to visually classify images of galaxies drawn from NASA's Hubble Space Telescope archive. Moon Zoo does the same for images of moon craters from NASA's Lunar Reconnaissance Orbiter [1].

3.2.4 iStockphoto

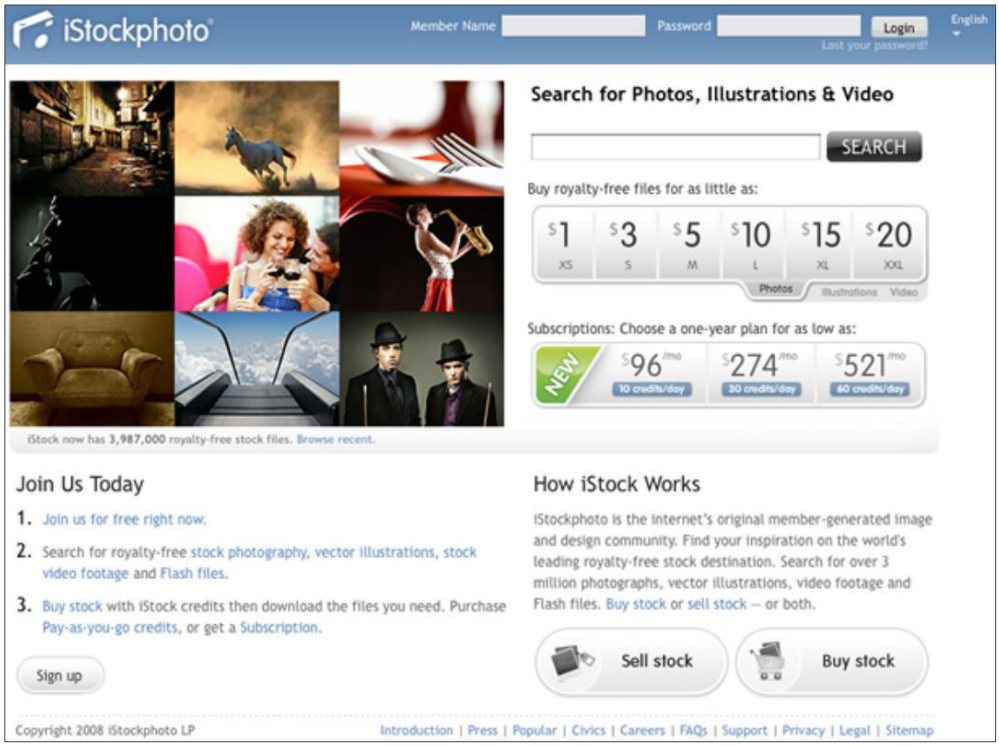

Figure 3.5 : iStockphoto web page.

iStockphoto.com is a web-based company that sells royalty-free stock photography, animations, and video clips. Calgary, Alberta-based iStockphoto was launched in February 2000, founded by Bruce Livingstone, who 'conceived the iStockphoto engine'. Figure 3.5 shows view of web site. To become a photographer for iStockphoto, one must fill out an online form, submit proof of identification, and submit three photographs for judging by the iStockphoto staff. If the photographs are technically sound, regardless of their content, applicants are typically admitted as photographers to the website. From that point, photographers may submit their photographs to the website to be stored in the databases under keywords. Clients seeking stock images – for use on websites, in brochures, in business presentations and so on – purchase credits (US $1 per credit) and start buying the stock images they want. Typical sizes and qualities of photographs can be purchased, royalty-free, from between one and five credits, with high resolution photographs, oversized images, and some longer video clips costing as many as 50 credits.

Photographers receive 20 per cent of the purchase price any time one of their images is downloaded (Frequently Asked Questions, n.d.), and some photographers, who become more involved members of the online community and typically end up

donating their talents for screening applicants and maintaining the database, can begin to earn exclusive contracts with iStockphoto and get 40 per cent of the price of their sold work. As long as photographs are in focus, free of dust specks and so forth, they will be accepted to the database, meaning anyone able to operate a camera can potentially earn money as a stock photographer. Like Threadless, iStockphoto's community is composed of both amateurs and working professionals in the field.

3.2.5 InnoCentive

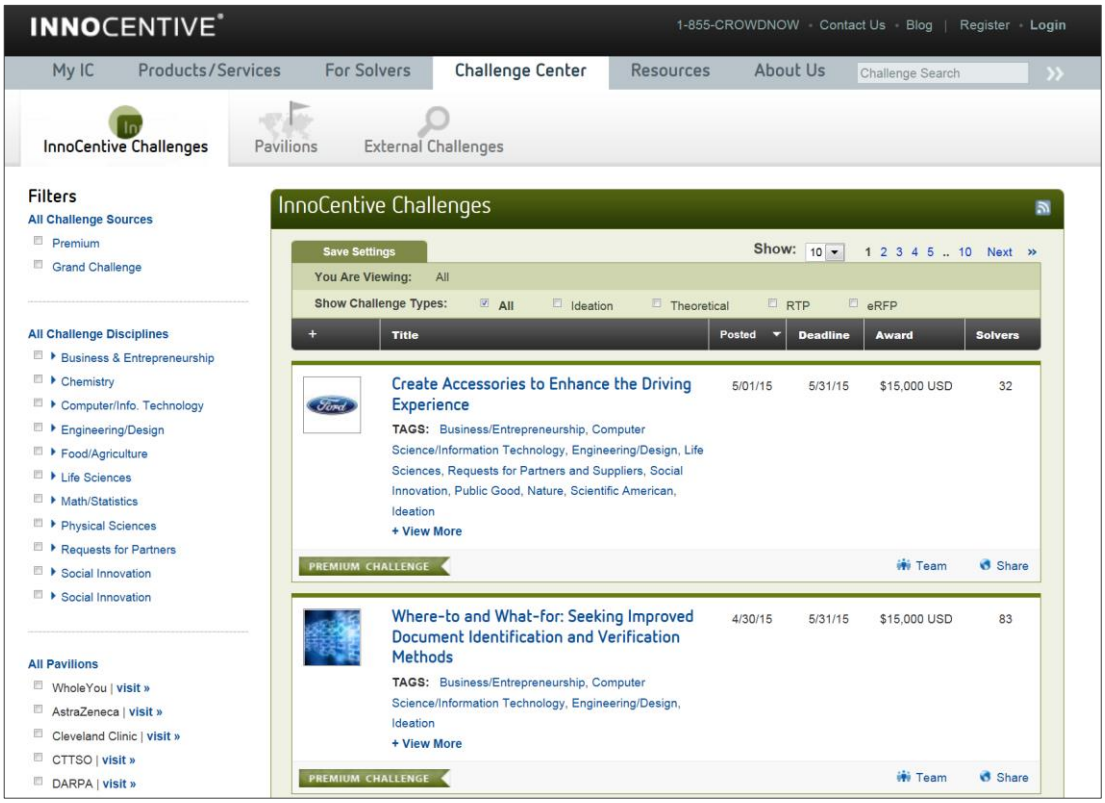

Figure 3.6 : InnoCentive web platform.

Crowdsourcing is not limited to the creative and design industries. Corporate research and development (R&D) for scientific problems is taking place in a crowdsourced way at InnoCentive.com. Launched in 2001 with funding from pharmaceutical giant Eli Lilly. Andover, Massachusetts-based InnoCentive 'enables scientists to receive professional recognition and financial award for solving R&D challenges', while it simultaneously 'enables companies to tap into the talents of a global scientific community for innovative solutions to tough R&D problems'. Seeker companies, which include 'Boeing, DuPont, and Proctor and Gamble' (Howe, 2006 p: 22), post their most difficult R&D challenges to the InnoCentive solvers under the broad categories of Life Sciences and Chemistry and Applied Sciences.

The crowd of solvers can then submit solutions through the web, which go under review by the seeker, which remains anonymous at least during the open phase. If a solution meets the technical requirements for the challenge, which about half of the time only requires written theoretical and methodological proposals, the seeker company awards a cash prize that they determine up front. Awards range from US$10,000 to $100,000 per challenge, though a current challenge, open through November 2008, offers US$1 million to a solution actually put into practice that identifies a biomarker for measuring disease progression in ALS (Amyotrophic Lateral Sclerosis). Figure 3.6 shows view of web site.

Potential solvers need only to register for free at InnoCentive, supplying contact information and checking off categories for degrees earned, areas of research interest and so on, though each of these questions required for registration includes an 'other' option, meaning solvers need not be professional scientists or scholars. Submitting solutions is simple, also, requiring only the uploading of a word-processed solution written into a downloadable template in most cases. InnoCentive 'broadcasts scientific challenges to over 80,000 independent scientists from over 150 countries' [10].

3.2.6 TopCoder

Founded in 2001, TopCoder is a company that specializes in holding programming competitions. Every couple of weeks, young talented programmers in the world would compete in TopCoder's SRM (Single Round Match), a 2-hour competition that tests the programming and debugging skills of the contestants, as well as their knowledge of algorithms. There are other types of competitions as well, all related to software development and testing. By holding these competitions, TopCoder is able to build a community consisting of 300,000 members, and generating $19 million of revenue by 2007. TopCoder has several ways to make money with this community. One of them is by helping companies exploit this pool of talent through crowdsourcing their projects. For instance, when a third party wants to crowdsource a software component, TopCoder can first hold a design competition for that component.

Figure 3.7 : TopCoder web platform.

Contestants can then submit their entries and the best submission will be selected and the winner will be rewarded. there might be a separate development competition to implement the chosen design. Again, the best submission will be selected. There are other types of competitions as well, including software specification, architecture, assembly and testing. Since these competitions actually represent different stages of software development, TopCoder can help a company crowdsource a project phase simply by holding a corresponding competition for it. By breaking up a task into smaller pieces, the crowdsourcing can be carried out much more effectively by allowing the contestants to perform what they are best at. For instance, the best design might be implemented by someone else, whose implementation will be tested by yet another person. To make the website both a fun place to compete and learn, TopCoder also pays members to write various programming articles so that members can learn from each other. The reward is typically less than $500 for each article, which is a decent amount considering that many members are still college students. Figure 3.7 shows view of web site. The articles are also a great way for members to establish their status and gain recognition in the community [5].

3.3 Crowdsourcing With Sensing

Some crowdsourcing applications need sensing capabilities of devices to generate data. Wireless, 3G, LTE (Long Term Evolution) new generation Internet

connectivity have made mobile applications to burst and allow people to work not only on PC's (Personal Computer) but also in any location by using their smart devices.

3.3.1 Sensors

Today's smartphone not only serves as the key computing and communication mobile device of choice, but it also comes with a rich set of embedded sensors, such as an accelerometer, digital compass, gyroscope, GPS (Global Positioning System), microphone, and camera. See Fig. 3.8. Collectively, these sensors are enabling new applications across a wide variety of domains, such as healthcare [11], social networks [12], safety, environmental monitoring [13], and transportation [14], [15], and give rise to a new area of research called mobile phone sensing.

Figure 3.8 shows the suite of sensors found in the Apple iPhone. The phone's sensors include a gyroscope, compass, accelerometer, proximity sensor, and ambient light sensor, as well as other more conventional devices that can be used to sense such as front and back facing cameras, a microphone, GPS and WiFi (Wireless Fidelity), and Bluetooth radios. Many of the newer sensors are added to support the user interface (e.g., the accelerometer) or augment location-based services (e.g., the digital compass).

The proximity and light sensors allow the phone to perform simple forms of context recognition associated with the user interface. The proximity sensor detects, for example, when the user holds the phone to her face to speak. In this case, the touchscreen and keys are disabled, preventing them from accidentally being pressed as well as saving power because the screen is turned off. Light sensors are used to adjust the brightness of the screen. The GPS, which allows the phone to localize itself, enables new location-based applications such as local search, mobile social networks, and navigation. The compass and gyroscope represent an extension of location, providing the phone with increased awareness of its position in relation to the physical world (e.g., its direction and orientation) enhancing location-based applications.

Not only are these sensors useful in driving the user interface and providing location-based services; they also represent a significant opportunity to gather data about people and their environments. For example, accelerometer data is capable of

characterizing the physical movements of the user carrying the phone [12]. Distinct patterns within the accelerometer data can be exploited to automatically recognize different activities (e.g., running, walking, and standing). The camera and microphone are powerful sensors. These are probably the most ubiquitous sensors on the planet. By continuously collecting audio from the phone's microphone, for example, it is possible to classify a diverse set of distinctive sounds associated with a particular context or activity in a person's life, such as using an ATM (Automatic Teller Machine), being in a particular coffee shop, having a conversation, listening to music, making coffee, and driving [16]. The camera on the phone can be used for many things including traditional tasks such as photo blogging to more specialized sensing activities such as tracking the user's eye movement across the phone's display as a means to activate applications using the camera mounted on the front of the phone [17]. The combination of accelerometer data and a stream of location estimates from the GPS can recognize the mode of transportation of a user, such as using a bike or car or taking a bus or the subway [13].

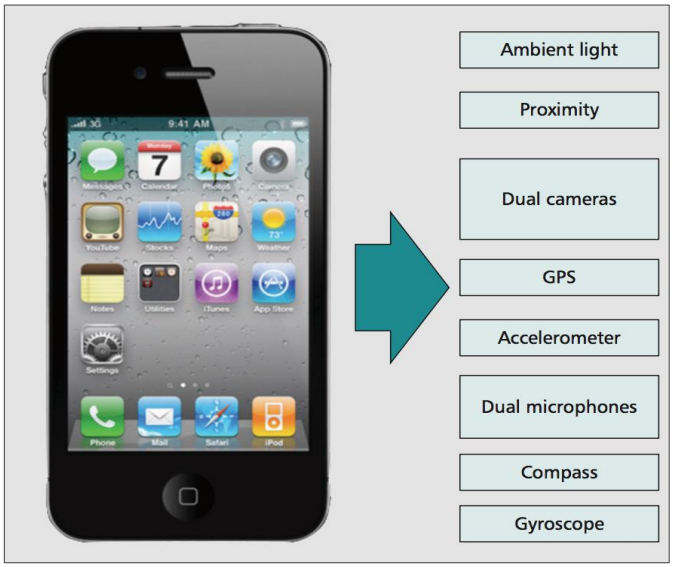

Figure 3.8 : Sensors of smartphones.

3.3.2 Smartphone applications

We can classify smartphone crowdsourcing applications into either extensions of Web-based applications or new applications. The former class expands to users who do not have access to a conventional workstation and adds the dimension of real-time location-based information to the service. Examples include Gigwalk

(www.gigwalk.com), Jana (www.jana.com), and work by Jonathan Ledlie and his colleagues.

The new applications offer functionalities such as crowdsourced traffic monitoring, as with Waze (www.waze.com); road-traffic delay estimation, as in VTrack; the construction of fine-grained noise maps using uploaded data captured by users' smartphone microphones (Ear-Phone and NoiseTube); the identification of holes in streets by letting users share vibration and location data their smartphones capture (PotHole); Location-based games aimed at collecting geo-spatial data (such as City-Explorer); Collaborative traffic signal schedule advisories (SignalGuru); and real-time, fine grained indoor localization services that exploit the radio signal strength of Wi-Fi access points (Airplace).

Another key characteristic of mobile crowdsourcing is whether the crowd's contribution is participatory or opportunistic. Typically, users perform computations or generate data as input for participatory crowdsourcing; the input for opportunistic crowdsourcing is data generated from sensors and computations that are automatically performed by the crowd's devices — for example, trajectory matching and positional triangulation. Classical crowdsourcing services on the Web are participatory because they require users' active participation. The second category's crowdsourcing tasks are transparent to users because they usually run in the background using sensors to collect environmental readings.

Further classifications can be adapted from crowdsourcing taxonomies proposed by David Geiger and his colleagues and by Alexander Quinn and Benjamin Bederson. Both studies recognize that the input's value can lie either in the individual or the collective contribution, where "the crowdsourcing system strives to benefit from each contribution in isolation or from an emerging property resulting from the system of stimuli," respectively. Furthermore, Geiger and colleagues divide applications by contribution quality, which can be homogeneous or heterogeneous. In the former, each contribution has the same weight, whereas in the latter, each contribution is evaluated and can be compared to, compete against, or complete other contributions.

Table 3.1 shows taxonomy of existing mobile crowdsourcing applications. The "Sensors" column shows which sensors the application is using. A separate

"Location" column is dedicated to the sensors that offer location awareness and shows that most crowdsourcing applications use this feature.

Location-dependent crowdsourcing applications can further benefit from adding a temporal dimension to location data to exploit trajectory-related information. They can also benefit from interrelations between location data, such as proximity information [16].

Table 3.1: Taxonomy of mobile crowdsourcing applications.

Applications	Web extend	Involvement	Data wisdom	Contribution quality	Incentives	Human skill	Sensors	Location
Gigwalk.com	Y	Participatory	Individual	Heterogeneous	Monetary	Labor	Camera	Y
Jana.com	Y	Participatory	Individual	Heterogeneous	Monetary	Visual	X	Y
Crowd Translator	Y	Participatory	Collective	Homogeneous	Service	Visual	X	X
Waze	X	Both	Collective	Homogeneous	Ethical/service	Visual	Camera	Y
City Exlporer	X	Participatory	Collective	Homogeneous	Entertainment	Visual	Camera	Y
Vtrack	X	Opurtunistic	Collective	Homogeneous	Ethical/service	X	X	Y
Signal Guru	X	Opurtunistic	Collective	Homogeneous	Ethical/service	X	Camera	Y
Ear-Phone	X	Opurtunistic	Collective	Homogeneous	Ethical	X	Audio	Y
NoiseTube	X	Opurtunistic	Collective	Homogeneous	Ethical	X	Audio	Y
PotHole	X	Opurtunistic	Collective	Homogeneous	Ethical	X	Vibration	Y
AirPlace	X	Opurtunistic	Collective	Homogeneous	Service	X	X	Y
SmartTrace	X	Opurtunistic	Collective	Homogeneous	Service	X	X	Y
Crowdcast	X	Opurtunistic	Collective	Homogeneous	Service	X	X	Y
SmartP2P	X	Opurtunistic	Collective	Homogeneous	Service	X	X	Y

3.3.3 SmartTrace+

One example of smartphone crowdsourcing is to ask a crowd of smartphone users to help identify mobility patterns or a given trajectory's popularity. Such a contribution can be utilized in large-scale urban and transit planning, transit rider information applications (www. tiramisutransit.com), shared-ride applications (www.avego.com and www.relayrides.com), social networking applications on smartphones, habitant monitoring, and so on.

Consider a transit authority that plans its bus routes and wants to know whether a specific route is taken by at least k users between 7:00 a.m. and 8:00 a.m. In such a scenario, the transit authority asks a crowd of users in a target area to participate with their local trace history through an open call. Users can opportunistically participate in the query's resolution without disclosing their traces to the authority for monetary benefit or for intellectual satisfaction. The SmartTrace+ project (http://smarttrace.

cs.ucy.ac.cy) enables trace similarity search among smartphone users and optimizes queries with respect to response time and energy consumption (see Figure 3.9). More importantly, SmartTrace+ is privacy-aware: it doesn't share user trajectories with the authority, but rather returns only matching scores.

At a high level, the SmartTrace+ GUI can

- record traces on local storage and plot those on the screen for the outdoor case,
- configure various logging and querying features,
- connect to a SmartTrace+ server and query the traces stored on other connected nodes, and
- switch between online and offline mode to change between experimentation and real operation.

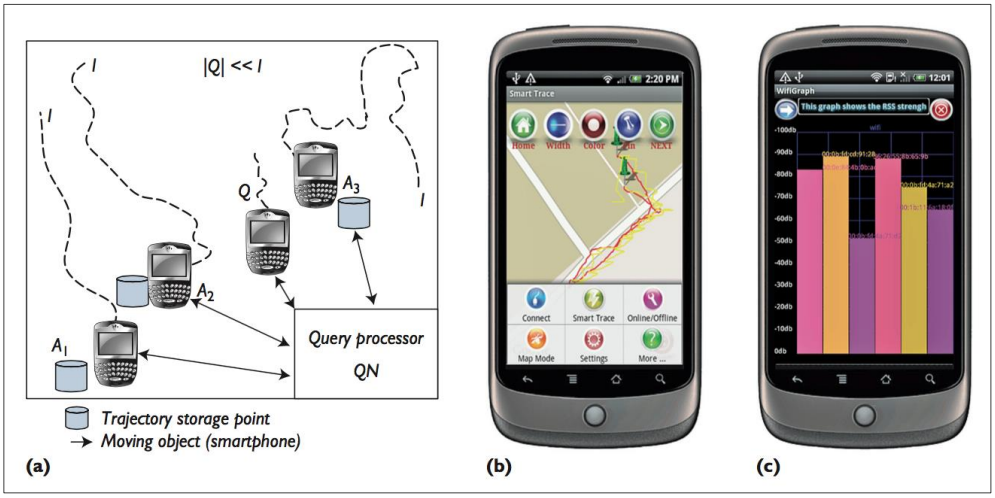

Figure 3.9 : SmartTrace +.

The SmartTrace+ project enables trace similarity search among smartphone users. It answers queries of the form "Report the users that move similar to Q," where Q is some query trace. It optimizes such queries with respect to response time and energy consumption on the smartphones, without sharing users' personal trajectories with the query processor. It also rewards clients. (a) The SmartTrace+ system model. (b) Fig. 3.9. A screenshot from the SmartTrace+ client for outdoor environments with GPS. (c) A screenshot from the SmartTrace+ client for indoor environments showing radio signal strength [8].

3.3.4 The CenceMe

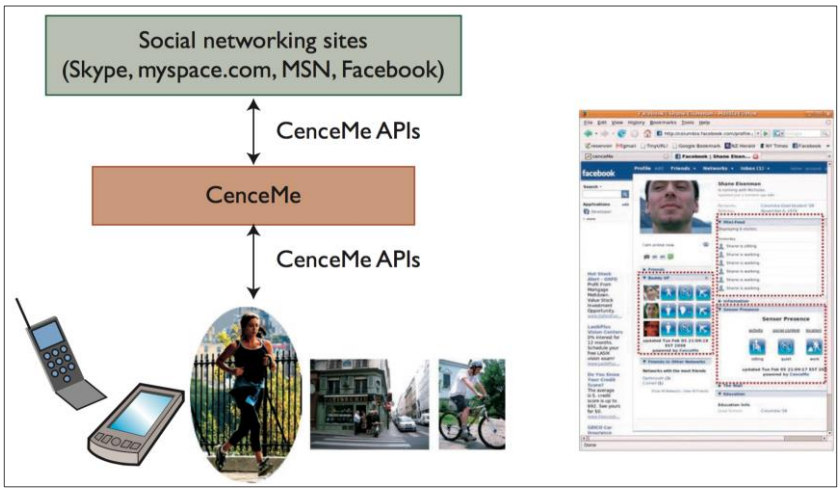

Figure 3.10 : The CenceMe.

CenceMe distills a user's sensing presence from samples taken from sensors embedded in personal mobile devices, sports equipment (such as running shoes or a bicycle), and the civic infrastructure (see Figure 3.10). Users can share sensing presence with their friends through popular social networking applications. There are widgets build for Facebook that allow expression of sensing presence through the friends list, the mini- feed, and a dedicated Sensor Presence display.

3.3.5 BikeNet

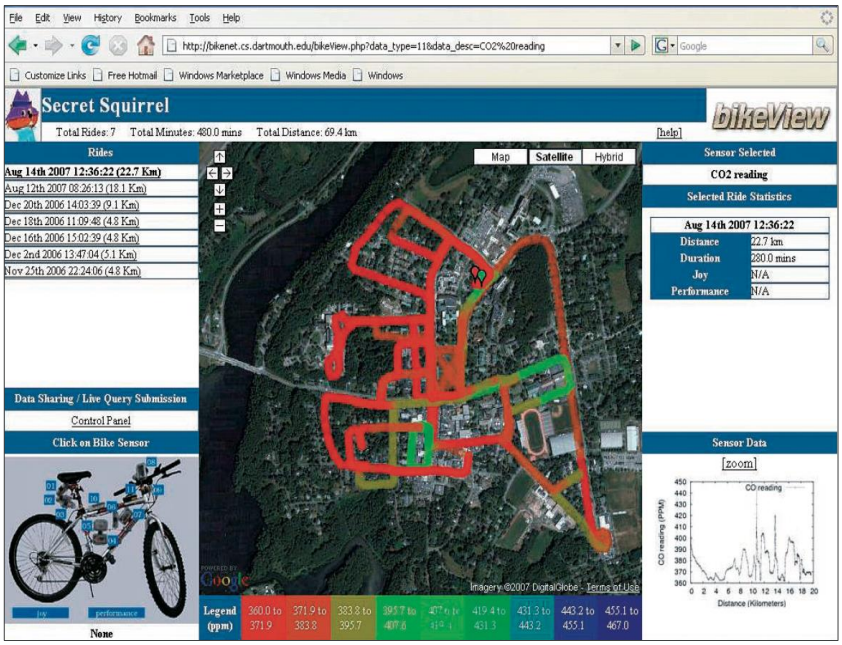

Figure 3.11 : The BikeView application.

BikeNet is a recreational application that contains elements of personal, social, and public sensing. There's substantial interest in the mainstream recreational cycling community in collecting data quantifying various aspects of the cycling experience, mirroring the broader interest in fitness metrics among exercise enthusiasts and other health-conscious individuals.

The BikeNet application measures several metrics to give a holistic picture of the cyclist experience: current speed, average speed, distance traveled, calories burned, path incline, heart rate, CO_2 (carbon dioxide) level, car density surrounding the cyclist (see Figure 3.11). The portal provides personal access to archived cycling data, which can be socially shared with cyclists or used to support a public sensing initiative. This CO_2 map is the result of multiple users' data merged to form a complete map of Hanover, New Hampshire [18].

Below are listed some other popular applications according to their field of work. Clickworker [19] on text creation and data categorization, Humangird [20] specialized on data analysis. Platforms like vWorker [21], CrowdFlower [22], Odesk [23], Microworkers [24], and ShortTask [25] are those where employers submit individually designed tasks. Atizio [26] innovative concepts. Wilogo [27] employed to design logos, onclickdesign.com is used for graphical design. Mobile applications like TaskRabbit [28], EasyShift [29], Gigwalk [30], MobileWorks [31], OpenStreetMap [32] and mClerk [33] are popular examples of crowdsourcing applications.

4. REFERENCE MODEL PROPOSAL AS A SERVICE

4.1 Previous Work on Reference Models

On previous crowdsourcing model works, researchers analysed the literature and deduces taxonomy of crowdsourcing. Generally, the taxonomy presents components of crowdsourcing process. These works have contributed on understanding attributes and components of so-called pillars of crowdsourcing process, which are the crowdsourcer, the crowd, the task and the platform. On "The Four Pillars of Crowdsourcing: a Reference Model" [7] authors were concentrated to classify components and features of components mentioned. Although presenting four components of crowdsourcing process, authors have not mention some other key component like Task Manager, Evaluation, and Reputation etc. In addition athuors have not clearly outline interacitons between components. Currently there are hundreds of applications working with the principles of crowdsourcing. Most of these applications are designed in ad-hoc matter of realizing an aim. This can be a business model, finding solution for biomedical problem, retrieving data from sensors, or platforms where people can earn money by participating different type of tasks. Analysing separately applications is valuable for identifying new components or features in crowdsourcing process, but having researches on generealizing crowdsourcing process in a model that represent applications realized untill now will accelerate growth of crowdsourcing.

4.2 The Introduced Reference Model

In this research, we tried to develop a general reference model for crowdsourcing as a service. A crowdsourcing system has four primary components, namely the crowd (crowdsourcing providers), crowdsourcing tasks, the platform and the crowdsourcers (i.e., end users of crowdsourced data). Besides these, a crowdsourcing system hosts the following sub-components that run over the platform:

i) The task manager

ii) evaluation,
iii) user ranking
iv) incentives.

The proposed reference model along with the four primary players and the computational components that are hosted in a cloud platform are illustrated in Fig. 4.1 It is worthwhile noting that the crowdsourcing service providers do not necessarily interact with the platform directly but they can use a data publisher layer (e.g., social media accounts) to communicate their crowdsourced data [34].

As seen in the figure, we partition the crowdsourcing process into four phases as follows:

i) registry and task generation,
ii) task distribution,
iii) evaluation
iv) ranking-payment.

4.2.1 First phase: registry and task creation

Registry and task generation phase involves the crowdsourcing providers and end users. Both providers and end users register the cloud platform to receive crowdsourcing as a service.

This phase contains three important activities of Crowdsourcer and Crowd users:

- Registration to platform
- Task creation and
- Publishing tasks in the form of Open Call

As it can be seen, at first phase Crowdsourcer and crowd users will register to the platform. User registration is important and all actions taken by each user (Crowdsourcer or a Crowd) are tracked and logged using user id. There are two types of registration.

- Valid identity registration
- Anonymous registration

In the valid identity registration, crowd users are registered to the platform with their real identities. There are different types of identity validation techniques. Platform pushes crowd to provide real identity information or crowds can be registered using other information providers like Twitter, Facebook and other social media official accounts [35].

In the anonymous registration, crowd users don't share their identity information. In this type of applications where crowd users are unknown, work is done in the matter of willingness or for a charity issue. Mostly users are tracked by a nickname or other invalid identification. Being unknown does not prevent crowd user working for a task, but in order to receive payment user should share identity with the platform. It is worthwhile noting that users that are anonymous registered will not be rewarded for providing their crowdsourced data as a service to the platform.

After Crowdsourcer and Crowd register to platform they start their activities. Crowdsourcer create tasks. These tasks are published in form of open calls to crowd. Created tasks are stored in the Task Storage module with an ID (Identity), Task Type and other parameters. These parameters are used by Task Manager to decide which tasks are open for all, which ones can be assigned to specific crowd users (based on location, gender, etc.).

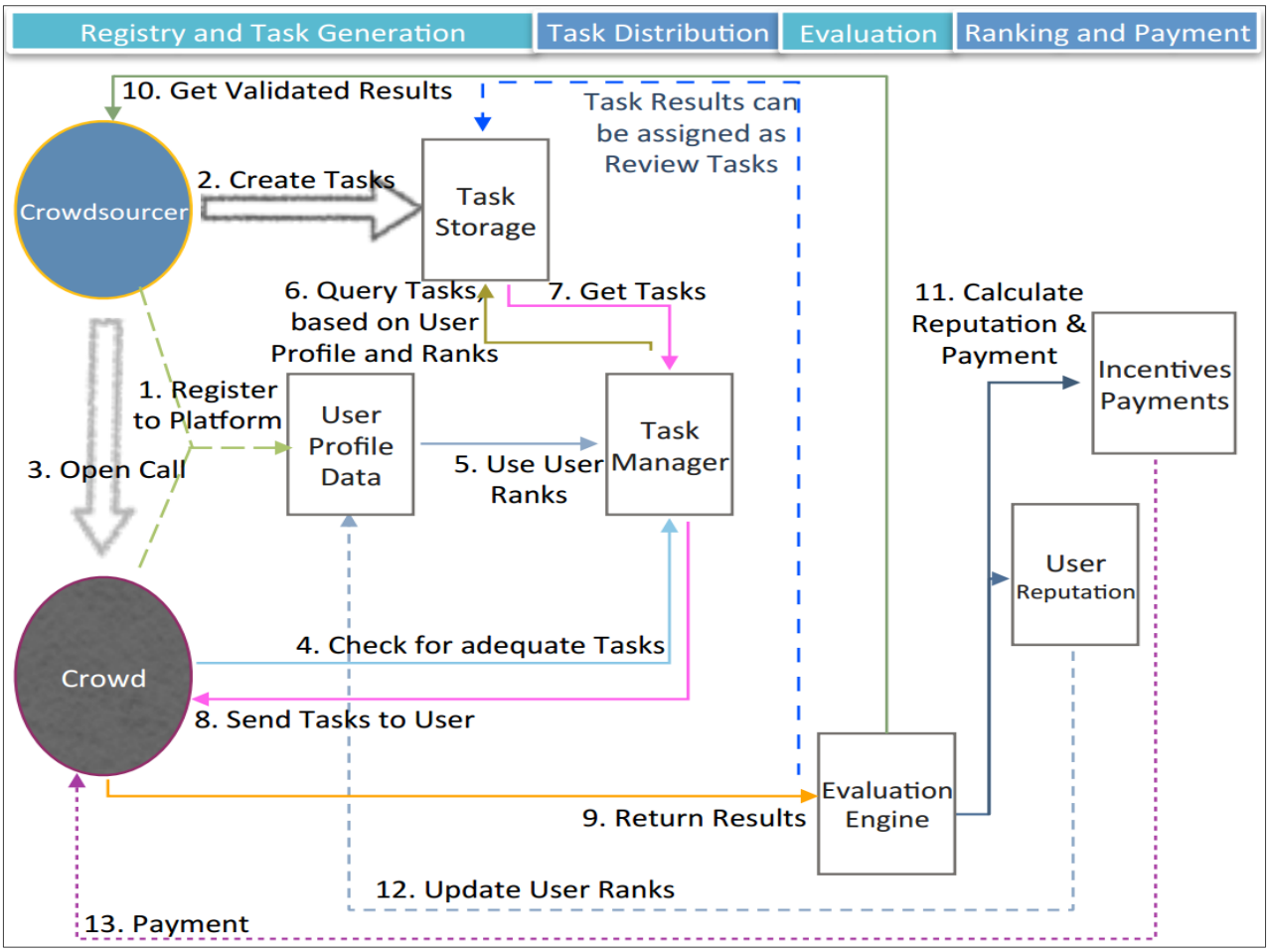

Figure 4.1: A reference model proposal for crowdsourcing as service.

4.2.2 Second phase: task distribution

In this phase Task Manager controls and distributes tasks to proper crowd users in most efficient. Although mostly many tasks are open for all crowd users to participate, but some crowdsourcing tasks can only be open for specific individuals, (considering sensors, groups or locations, etc.). Having many selection criteria to choose crowd users, task distribution becomes a complicated issue. Crowd user's profile attributes rank (which shows the percentage of successfully finished tasks by the user), task type, are some of the parameters that task manager uses to decide which task can be assigned or recommended to a crowd user. Task distribution phase involves matching the crowdsourcing service providers with tasks. Service provider-task matching can be performed based on various criteria. In [36], crowdsourcing is applied to a participatory sensing application via mobile phones where service providers are selected based on their location, reputations, and residual battery power. Furthermore in [37], sensing costs of the users are used as input parameters of the selection process since users are selected based on an auction in

which they are guaranteed that they will be rewarded no less than their sensing costs. Task manager explained in detailed in next Sections.

4.2.3 Third phase: evaluation

Evaluation phase evaluates the "value" of the crowdsourced data. Furthermore, anomalous crowdsourced data regarding a particular task is discarded while data is being aggregated, and the corresponding crowdsourcing service provider is marked as an outlier whose likelihood of recruitment is potentially to be degraded in the following task distributions. Data generated by crowd users are studied and evaluated by the corresponding evaluation engine. We can see in many crowdsourcing applications that the system introduces the evaluation process as another task to the platform. Although such an evaluation has advantages, it also needs a final control. In addition to evaluation type mentioned above there are different types of evaluation algorithms used by crowdsourcing applications. A valid evaluation of results returned by crowd users is important for the motivation of crowd users.

4.2.4 Fourth phase: ranking and payment

The last phase is crowd ranking and incentive (motivation/payment) phase. Ranking and rewarding phase is closely related to user incentives and trustworthiness. As crowdsourcing as a service requires effective incentives to enable participation from large number of users, rewarding mechanism have to be set [38]. Furthermore, in order to ensure trustworthiness of crowdsourced data to the end user, the crowdsourcing providers should be either ranked or a reputation database has to be maintained.

Motivation has effective role on increasing the performance of crowd users. Therefore, evaluation of results and the fair judgment are important in incentives that will be used after task completed. Payment is done after confirmation that task is completed successfully. Crowd rank will be updated by calculating the percentage of successfully finished tasks. Crowd users with higher ranks have more chance to get a task than the other users with low ranks. Ranks are used to decide payment rate. Incentives depend on task type, complexity and time needed to complete. Although most of micro task payments differ between some cents and a few dollars, in complicated or creative tasks payments go up to thousands of dollars.

Depending on platform policies, ranks may not be shared with the crowd users because this may have negative effect on crowd users. Meaning that low rank users may leave account and open new a one.

4.3 Entities of Crowdsourcing Model

In this section the component of crowdsourcing model, which are Crowdsourcer, Crowd, Task, Task Manager, Evaluation, and Platform are described in detail.

4.3.1 Crowdsourcer

In Crowdsourcer is the initiator of task process in crowdsourcing, by creating tasks feeding the platform, and creates new opportunities for crowd users. Crowdsourcer that join crowdsourcing platform can be people from different regions, different nationalities, gender, religion, and with different purpose. A Crowdsourcer might be an individual, an institution, a non-profit organization, or a company that seeks completion of a task through the crowd. Crowdsourcer should be a validated user whose identity is known or should have confirmed deposition of payment in the crowdsourcing application platform. Crowdsourcer has responsibilities and its own objectives, which are as follows: creation of tasks, constructing task size, designing content, broke the job into micro tasks. Creating solvable tasks, and defining incentives (payment, reputation etc.) is duty of Crowdsourcer. Usually Crowdsourcer selects predefined task types while defining task. However, tasks that do not fit in any of the categories can be submitted as “Others”.

Crowdsourcer’s prime objective is to maximize the expected number of tasks performed without exceeding the budget. A complementary objective is to minimize the expected payments for a given number of tasks.

An incentive is something that stimulates crowds to work. Crowdsourcer should act ethically and carefully evaluate the work done by crowd. Unfairly rejected work, slow payment, and payments that do not fairly reflect the work performed for the given tasks are problems that face crowd after completing their tasks.

Privacy of crowds is also an important concern of Crowdsourcer in the applications where opinion of crowd is directly reflected, Crowdsourcer should

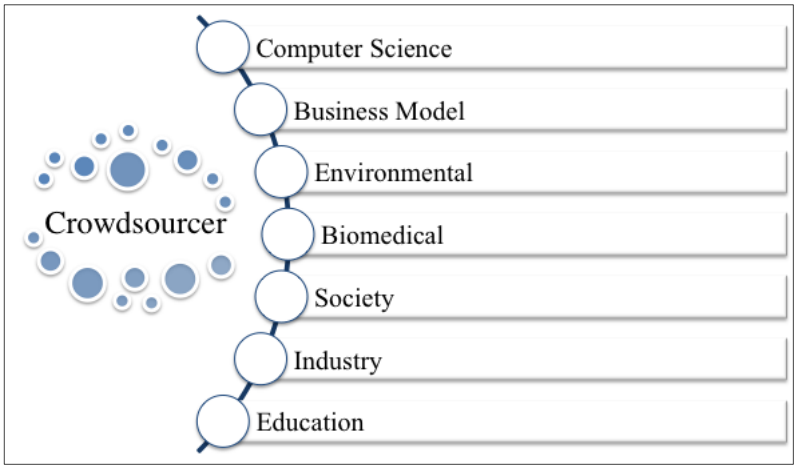

Figure 4.2 : Crowdsourcer's fields of working.

act ethically and not reveal private information.Crowdsourcing have opened new opportunity to all crowdsourcers to share their work that needs to be done by human work power. Mostly crowdsourcers tend to divide tasks into micro-task to keep low cost payments and making tasks easier to be solved by crowds. Computer Vision, Natural Language Processing, and Human Computer Interaction, Business and Marketing, Biomedical Engineering, Environment Issues, Education are some fields attracting Crowdsourcer and Crowds. Some of works done in different fields mentioned are described below (see Fig. 4.2.). Micro-task crowdsourcing, where workers are paid small amounts to complete simple, short tasks, may be one technique to alleviate some of the development difficulties in biomedical ontologies [39]. Medical image retrieval and medical image classification have been used as a way of improving access to visual medical information in clinical domains [40]. Firms use to get input and advice on their own product development efforts from existing end-users, and experts who may be able to solve a certain scientific or design problem . There is also a rapid growth of crowdsourcing applications for disaster relief [41].

Crowdsourcing calls attract different categories of potential workers. With spread of Internet and mobile smart device usage new working power force has been revealed. Greater work potential of any category will attract more crowdsourcersto share their work through crowdsourcing platforms. As an example, beside simple micro tasks, recently we see a trend of complicated tasks being crowdsourced as well. These complicated tasks have great impact in attracting qualified professionals to be part of crowdsourcing application. Many organizations have

already turned to crowdsource theirsoftware development projects. TopCoder [5], InnoCentive [42] are two good examples of this type.

4.3.2 Crowd

"Crowd" means a "Large number of people gathered together in a disorganized or unruly way" according to [43]. However, considering the daily life equipment we are commonly using, we think "Crowd" mean "Large number of people equipped with smart devices (or sensors itself) which can be disorganized or organized in communities (groups)". As well as Crowdsourcer, crowds also have features and their own characteristics. A common feature with Crowdsourcer is that both can be from different locations and participate in crowdsourcing system through a web platform using a mobile smart device or computer depending on task they will perform. Crowds join to system willingly, for a profit / incentive.

Crowds identity can be invalidated, which means crowd user may keep anonymity to Crowdsourcer as well as Crowdsourcer may be unknown to crowd user.

Crowds participate to crowdsourcing applications for many reasons. Mostly crowd users think to complete tasks properly and get paid (or other incentives) for work done. On the other hand, some crowd users try to earn money in easily by finding weaknesses of the system and cheating by generating results for tasks assigned to her/him and trying to get payment. This type of work is done more in less complicated tasks where unreliable users write programs to generate the invalid data and send these as they are collected naturally. In Evaluation stage, such as cheating should be identified.

Crowds can work individually as well as in-groups, form communities. Working in groups brings crowd users ability to transfer tasks to each other within the group.

Crowd users should be free in resigning from tasks any time. A short period of time allowed to crowd user to understand task. If user figures out that he/she does not have ability to complete the task it drops. In this case, no penalty is given toward crowd user; otherwise penalty is applied to the crowd. In case of public tasks that are open to everyone, Crowdsourcer may set maximum number of tasks that can be finished by Crowd. When maximum number is reached the task is closed, Crowd users that have been working but have not delivered results are informed and not

paid. Such policies are defined, shared with crowd users and seek for acceptance from crowd users before start working in the platform.

4.3.2.1 Crowd participation

All crowdsourcing platforms require the active participation of users in the solution of the assigned task. There are some works done in order to increase participation of crowd. As part of their work, Quinn and Bederson study incentives [42] used for the crowd, some of which can include: altruism, enjoyment, reputation, payment, and entertainment. Effectiveness of incentives is high. It is seen that willingness of people to participate is a key factor for the success of applications. The widespread availability of smartphones and tablets featuring geo-location and other sensing capabilities (e.g., proximity, ambient light, accelerometer, camera, microphone, etc.) are providing new means for opportunistic crowdsourcing frameworks.

The expected participation rate is a key factor when designing a crowdsourcing system. In repetitive tasks (like voting), there is minimum participation threshold for reaching the expected results. Low rate of participation decreases accuracy of results. A system with a lower expected participation rate will place a higher burden on individual participants. Hence, by designing more interfaces that are elaborate the accuracy of the collected data can be improved. In applications where Crowdsourcer ask Crowd about any kind of design or solution for a specific task, Crowdsourcer will try to increase participation without setting any (max, min) submission threshold limit. High number of submitted work means more options for Crowdsourcer to reach result. Quantity does not always mean higher quality. It complicates the evaluation process and the related cost also. High number of participation is an advantage for web based crowdsourcing applications. However, in case of mobile devices too much data sharing and processing causes energy consumption, which can be concern for crowd users. Our aim should be try to keep high participation but optimize required data collected from crowd users especially when we have to work with mobile devices.

Many previous works showed that user interfaces can affect the behavior of crowdsourcing workers. By analyzing the waiting time for the posted tasks on MTurk, Ipeirotis [44] found that workers are limited by the current user interface and complete tasks by picking the tasks available through one of the existing sorting

criteria. In addition to user interfaces, other factors affecting the behavior of crowdsourcing workers were found in the literature. Grady and Lease [41] investigated human factors involved in designing effective tasks on MTurk for document relevant assessment. They found that many of the same workers completed tasks in multiple batches, compromising the experimental control and likely introducing effects of training or fatigue. However, MTurk cannot prevent this happens. It is necessary to ensure each experiment involves a different set of workers in order to increase the output accuracy. Besides MTurk, other crowdsourcing websites were studied in literature. In 2008, Yang et al. [45] observed several characteristics in workers' activity over time on one of the biggest crowdsourcing websites in China, Taskcn.com. It found that most workers become inactive after only a few submissions, while others keep attempting tasks. They tend to select tasks where they are competing against fewer opponents to increase their chances of winning; or they tend to select tasks with higher expected rewards. Instead of public crowdsourcing, a firm can outsource tasks to its employees rather than assign tasks to specified employees. Based on quantifiable effort-level metrics, Stewart et al. [46] proposed a SCOUT (S)uper, (C)ontributor, and (OUT)lier) model for describing user participation inside the enterprise (within a company's firewall) and showed that it is possible to achieve a more equitable distribution.

The following model (Figure 4.3) can be derived from the proposed relationships between the different variables and Participation.

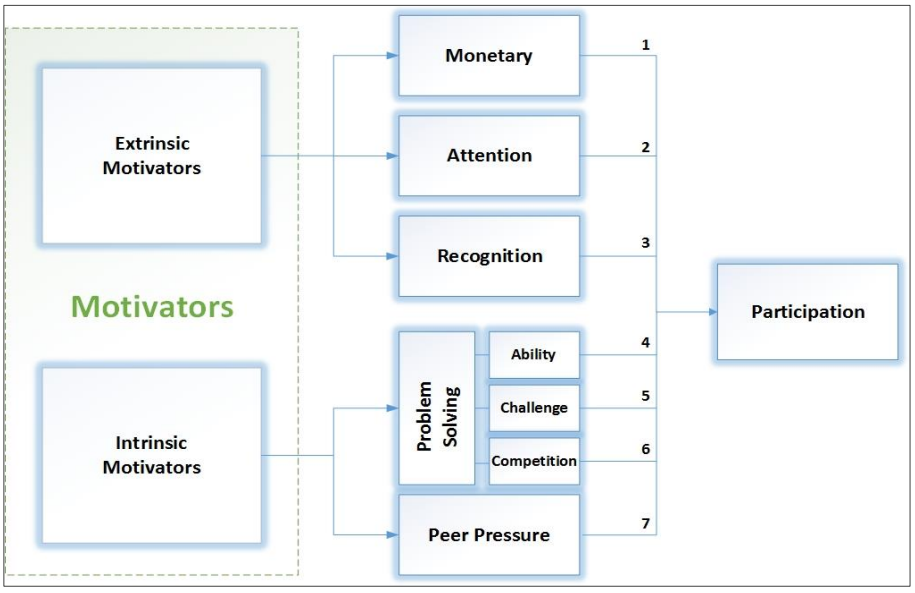

Figure 4.3 : Proposed model. Own, 2012.

1: Monetary Awards have a positive effect on Crowd-Participation

2: Attention has a positive effect on Crowd-Participation.

3: Recognition has a positive effect on Crowd-Participation.

4: The perceived Ability to solve a challenging task has a positive effect on CrowdParticipation.

5: The Challenge of solving a problem has a positive effect on Crowd Participation.

6: The ability to win a Competition has a positive effect on Crowd-Participation.

7: Peer Pressure has a positive effect on Crowd-Participation

In order to find drivers for Participation for the context of an explicit Crowdsourcing platform, drivers for other contexts can be located and tested for the explicit Crowdsourcing specific context. Firstly, a distinction should be made between extrinsic and intrinsic motivators. It is important to include intrinsic motivational factors and not only look at extrinsic factors.

4.3.2.2 Incentives of crowd

Platforms such as Wikipedia, OpenStreetMap [32], and others examples (entertainment, educational opportunities, information [47], and altruism have gained great success. However, it is difficult to setup non-monetary incentive schemes. Fun and enjoyment are the two dominant intrinsic motivational factors present in online platforms [20], [8], [30]. Some crowds may take part in crowdsourcing activities in order to gain peer recognition [48] or public recognition [49]. Reward and recognition are worthwhile factors. Considering commercial usage of crowdsourcing variety of crowdsourcing platforms has been developed.

Companies and other institutions implement crowdsourcing applications to solve their problems using cheaper workforce, quickly and efficiently. Companies must consider Crowd's motivation and willingness to guarantee a satisfying level of participation and good-quality results. Understanding and analyzing the incentives that encourages users to contribute on crowdsourcing applications is crucial. Therefore, majority of crowdsourcing tasks are performed in exchange for payments. In such applications, implementing a campaign successfully requires pricing and allocating tasks effectively.

Classifications of motivations

Motivations are mainly classifies into groups: extrinsic and intrinsic. Extrinsic motivation further classified into three groups:

1) Financial
2) Social, and
3) Organizational

While intrinsic motivation cannot be classified further as the intrinsic motivation originate from a particular task itself.

However, there may have different classifications based other certain categories. Financial motivation includes, among other, benefits, cash, dissatisfaction, job opportunities, personal need, problem, pressure and revenue, etc. People socially bonded and social motivations are prevailing in online platform. Moreover, social motivational factors are very useful as it does not require financial issues. The prominent social motivators are obligation, peer recognition, status, reputation, power, skill development, experience, knowledge gathering, social bonds social interaction, networking, collaboration, ego, publicity privilege attainment, frustration, helpfulness, etc. Apart from financial and social motivators, people are motivated for organizational purpose as in each organization people need to consider their responsibilities, prestige, career, recruitment. Furthermore, employees need to market themselves. Table 4.1 provides intrinsic motivational factors unclassified as there are no classifications of them. Intrinsic motivators of wide range: charity, enjoyment, pleasure, self-satisfaction, desire to solve, competence, fun, autonomy, self-determination, identification, altruism, belonging, community drives, pride, pastime, learning, self-fulfillment, getting.

Table 4.1 : Motivators.

Financial Motivators	Social Motivators	Organizational Motivators
Benefits Cash Dissatisfaction Job opportunities Personal need Problem pressure Revenue	Obligation Collaboration Ego Experience Frustration Knowledge gathering Networking Peer recognition Power Privilege attainment Publicity Reputation Skill development Social bonds Social interaction Status	Career development Marketing oneself Professional prestige Recruitment Responsibilities
Extrinsic Motivators		
Charity Competence Desire to solve Enjoyment Fun Pleasure Self- satisfaction	Altruism Autonomy Belongingness Community Drives Identification Self-determination	Getting Ideas Ideas Comes True Learning Pastime Pride Self-fulfillment

Classifications of incentives

Incentives and motivators are highly inter-related. The incentives should be appropriate to meet what motivate people to do something. We have categorized incentives in similar fashion as we did for motivators. Hence, it gives a well illustration of comparative positions of motivators and incentives.

The salient incentives in financial categories are payment, premium, license, free product, free service, bonus, coupons, right to use product, right to use service, deals with lower cost, etc. On the other hand, social incentives are not many as we see in the Table 4.2 Social incentives include, among others, award, prize, honor, pride for development, trustworthiness, etc. while organizational incentives mainly are extra right, extra privilege, career opportunities, accessibility to information, etc.

Table 4.2 : Incentives.

Financial Incentives	Social Incentives	Organizational Incentives
Bonus Coupons Deals with lower cost Free product Free Service License Payment Premium Right to use product Right to use service	Award Honor Pride for development Prize Trustworthiness	Accessibility to information Career opportunities Extra Privilege Extra Right
Intrinsic Incentives **An activity itself**		

Pricing

One of the most important challenges for task requesters on crowdsourcing markets like Amazon Mechanical Turk (AMT) is to properly price and schedule their tasks (or "HITs," which stands for "Human Intelligence Tasks"). Improper pricing or scheduling often results in task starvation and loss of capital on these markets. For example, it is believed that workers have an expected hourly wage in mind and they tend to not accept underpriced tasks that need more time per unit reward than what they have in mind. Tasks that are not accepted stay in the system (they are often called "starved HITs"). Starved HITs may be canceled or reposted by the requester resulting in expenditure of more time and money than planned for the task. Overpriced tasks are also undesirable since requesters can invest excess capital in quality assurance for the data that they have collected.

Applications with payment mechanism have more chances to succeed. Setting a price to task is not such an easy task. Crowdsourcing applications are becoming popular, having higher payment rate per task comparing to other competitor platforms is important to attract Crowd. However, [2] found that increased financial incentives increase the quantity, but not the quality, of work performed by participants, where the difference appears to be due to an "anchoring" effect: workers who were paid more also perceived the value of their work to be greater, and thus were no more motivated than workers paid less. In contrast with compensation levels, we find the details of the compensation scheme do matter - specifically, a "quota" system results in better work for less pay than an equivalent "piece rate" system. Although

counterintuitive, these findings are consistent with previous laboratory studies, and may have real-world analogs as well.

4.3.2.3 Reliable crowds

Crowd users try to finish tasks fast and careless or on more complicated ones they are incapable to finish tasks and they submit invalid results. Crowd users try to maximize their financial gains (by producing generic answers rather than actually working properly on the task). Currently, cheat-detection techniques are either based on control questions, which are evaluated automatically or rely on manual checking by the requester. Eickhoff and de Vries inspected the commonly observed methods of malicious crowdsourcing workers, such as task-dependent evaluation, interface-dependent evaluation and audience-dependent evaluation. Based on experimental results, they concluded that malicious workers are less frequently encountered in novel tasks that involve a degree of creativity and abstraction, and prior crowd filtering can greatly reduce the number of malicious workers.

Although distributed classification in sensor networks and in crowdsourcing are structurally similar, an important difference is the anonymity of crowds. Since the crowd users are anonymous, we cannot identify the specific reliability of a specific worker as could potentially be done with a sensor. Hence, we assume that each worker in the crowd has an associated reliability P. Initially all crowds are assumed to be reliable, after submitting tasks, within evaluation process crowd users should be classified as reliable or unreliable.

4.3.3 Task

Task design is the sole responsibility of the Crowdsourcer (requester). Crowdsourcer serves task to crowd as work to be done in return of inceptive. A crowdsourced task may take different forms (example form of a problem, an innovation model, a data collection issue, or a fundraising scheme). As it has been mentioned, tasks differ on the size and complexity. Breaking complicated tasks into micro tasks will speed up time of crowd to finish work. Clarity, modularity and complexity of task have important affect on performance of crowd users. Complicated tasks are less understandable, this makes crowd users to be impatient to complete task. Also less complicated but big in size, less modularity makes crowd users to be bored and drop task without completing it.

Every task created has a unique id number, which makes possible to track it in every stage of crowdsourcing platform. Tasks contain information of responsible person and other required information requested from crowdsourcing platform. Table 4.3 shows attributes of tasks that should be shared to crowd users.

Table 4.3 : Public task attributes of task.

Domain	Description
TASKID	A unique task id number generated by the system
TASKTYPE	Type of task
CDATE/CTIME	Task create date/ Task create time
EDATE/ETIME	Task Expire date / Task Expire time
LIMITATIONS	Crowdsourcer may put criteria on tasks
LOCATION	Task uses location information
INCENTIVE	Monetary Payment / Other incentive from
TASK_DOC	Crowdsourcer attaches any type of document related to task
DESCRIPTION	Task description
REQ_PROOF	Required Proof from Crowd user
RESULT_DOC	Any type document attached by Crowd user containing work done

Task type domain (TASKTYPE) is attribute that defines borders of task. In many crowdsourcing applications crowd users are required to select type of task they wanted to work before started. Having a web based or mobile task type widens number of tasks that can be served to crowds. Features of the tasks usually determine the complexity of the whole system. Table 4.4 shows task types grouped into six categories.

Table 4.4 : Crowdsourcing task types.

Task Types
▪ Voting
▪ Information Sharing
▪ Creative Work - Designs
▪ Entertainment & Voluntary
▪ Complicated Work
▪ Sensed Data sharing

4.3.3.1 Voting

Usually tasks require a crowdsourcing worker to select his answer from a number of choices. Voting can be used as a tool to evaluate the correctness of an answer from the crowd.

An example of popular crowdsourcing websites is Amazon Mechanical Turk (or MTurk) [50]. A large number of applications or experiments were conducted in Amazon's MTurk site. It can support a large number of voting tasks. These voting tasks require a crowdsourcing worker to select his answer from a number of choices. The answer that the majority selected is considered correct. Voting can be used as a tool to evaluate the correctness of an answer from the crowd. Some examples are shown below:

- Geometric reasoning tasks - The ability to interpret and reason about shapes is a specific human capability that has proven difficult to reproduce algorithmically. Some work was proposed to solve the problem of geometric reasoning on MTurk [51], [52].
- Named entity annotation - Named entity recognitionis used to identify and categorize textual references to objects in the world, such as persons and organizations. MTurk is a very promising tool for annotating large-scale corpora, such as Arabic nicknames, Twitter data, large email datasets and medical named entities [53], [54].
- Opinions - Opinions are subjective preferences. Gath- ering opinions from the crowd can be achieved easily in a crowdsourcing system. Mellebeek et al. [55] used the crowdsourcing paradigm to classify Spanish consumer comments. They demonstrated that non-expert MTurk an- notations outperformed expert annotations using a varity of classifiers.
- Commonsense - Obviously, humans can poss common- sense knowledge about the world, but computer programs cannot. Many studies focused on collecting commonsense knowledge in MTurk [56], [57].
- Relevance evaluation - Humans have to read every document in a corpus to determine its relevance to a set of test queries. Alonso et al. proposed crowdsourcing for relevance evaluation, so that each crowdsourcing work performs a small evaluation task [58].
- Natural language annotation - Natural language annota- tion is a task that is easy for humans but currently difficult for automated processes. Recently, researchers' investi- gated MTurk as a source of non-expert natural language annotation, which is a cheap and quick alternative to expert annotations [59], Akkaya et al. showed that crowdsourcing for subjectivity word sense

annotation is reliable. Callison-Burch and Dredze [60] demonstrated their success on creating data for speech and language applications with a very low cost. Gao and Vogel [61] proved that crowdsourcing workers outperformed experts on word alignment tasks in terms of alignment error rate. Jha et al. [62] showed that it is possible to build up an accurate prepositional phrase attachment corpus by crowdsourcing workers. Parent and Eskenazi [62] demonstrated a way to cluster a task of dictionary definitions in MTurk.

- Spam identification - Junk email cannot be determined without the task of understanding content by humans. Some anti-spam mechanisms such as Vipul's Razor use human votes to determine if a given email is spam.

4.3.3.2 Information sharing

Websites can help to share information easily among Internet users. Some crowdsourcing systems aim to share various types of information among the crowd. For monitoring noise pollution, Maisonneuve designed a system called NoiseTube, which enables citizens to measure their personal exposure to noise in their everyday environment by using GPS equipped mobile phones as noise sensors [65]. The geo-localised measures and user-generated meta-data can be automatically sent and shared online with the public to contribute to the col- lective noise mapping of cities. Moreover, Choffnes et al. [66] utilized the crowdsoured contributions to monitor service-level network events and studied the impacts of network events on services in the view of end users. Furthermore, a lot of popular information sharing systems were launched on the Internet as shown in the following:

- Wikipedia are online encyclopedias that are written by Internet users, and the writing is distributed in that essentially almost anyone can contribute to the Wiki.
- Yahoo! Answers is a general question-answering forum to provide automated collection of human reviewed data at Internet-scale. These human-reviewed data are often required by enterprise and web data processing.
- Yahoo! Suggestion Board is an Internet-scale feedback and suggestion system.
- The website 43Things also collects goals from users, and in turn it provides a way for user to find other users who have the same goals, even if they are un-

common.

- Yahoo's flickr is a popular photo-sharing site and pro- vides a mechanism for users to caption their photos. These captions are already being used as alternative text.
- del.icio.us is a social bookmark site on the Internet developed by Golder and Huberman [67].

4.3.3.3 Entertaiment (Gaming)

By taking advantage of people's desire to be entertained, problems can be solved efficiently by online game players. The concept of "Social Game" was pioneered by Luis Von Ahn and his colleagues, who created games with a purpose [68]. The games produce useful metadata as a by-product. By taking advantage of people's desire to be entertained, problems can be solved efficiently by online game players. The online ESP Game [69] was the first human computation system, and it was subsequently adopted as the Google Image Labeler8. Its objective is to collect labels for images on Web. In addition to image annotation, the Peekaboom system [70] can help determine the location of objects in images, and the Squigl system provides complete outlines of the objects in an image. Besides, Phetch [71], provides image descriptions that improve web accessibility and image searches, while the Matchin system [72] helps image search engines rank images based on which ones are the most appealing. The concept of the ESP Game has been applied to other problems. For instance, the TagATune system, MajorMiner and The Listen Game provide annotation for sounds and music which can improve audio searches. The Verbosity system and the Common Consensus system collect commonsense knowledge that is valuable for commonsense reasoning and enhancing the design of interactive user interfaces. Several GWAP (Games with a Purpose) based geospatial tagging systems have been proposed in recent years, such as MobiMission, Gopher game and CityExplorer [46]. To simplify the way of designing a social game for a specific problem, Chan et al. [73] presented a formal framework for designing social games in general.

4.3.3.4 Creative work - designs

The role of human in creativity cannot be replaced by any advanced technologies. The creative tasks, such as drawing and coding, can only be done by humans. As a

result, some researchers seeked for crowdsourcing workers to do some creative tasks to reduce the production costs. An example is the Sheep Market. The Sheep Market is a web-based artwork to implicate thousands of online workers in the creation of a massive database of drawings. It is a collection of 10,000 sheeps created by MTurk workers, and each worker was paid US$0.02 to draw a sheep facing left [74]. Another example is Threadless. Threadless is a platform of collecting graphic t- shirt designs created by the communicty. Although technological advances rapidly nowadays, humans can innovate creative ideas in a product design process but computers cannot. It has no clue about how to solve a specific problem for developing a new product. Different individuals may create different ideas such as designing a T-shirt [10].

4.3.3.5 Complicated work

Coding may be shown as example, some companies seek for crowdsourcing workers to develop products using crowdsourcing tasks to reduce the production costs. Leimeister et al. proposed to crowdsource software development tasks as ideas competitions to motivate more users to support and participate [75]. Nowadays, scientists are being confronted by increasingly complex problems, but current technology unable to provide solutions. Some crowdsourcing systems were designed to solve these problems. Foldit is a revolutionary new computer game that allows players to assist in predicting protein structures, an important area of biochemistry that seeks to find cures for diseases, by taking advantage of humans' puzzle-solving intuitions.

A crowdsourcing system typically supports only simple, independent tasks, such as labeling an image or judging the relevance of a search result. Some works proposed an idea of coordination among many individuals to complete more complex human computation tasks [76]. Little et al. presented TurKit, which is a toolkit for exploring human computation algorithms on MTurk [77]. TurKit allows users to write algorithms in a straight-forward imperative programming style, abstracting MTurk as a function call. Rather than solving many small, unrelated tasks partitioned into individual HITs, TurKit presented the notion that a single task, such as sorting or editing text, might require multiple coordinated HITs, and offers a persistence layer that makes it simple to iteratively develop such tasks without incurring excessive

HIT costs. Kittur et al. presented CrowdForge, a general purpose framework for micro-task markets that provides a scaffolding for more complex human computation tasks which require coordination among many individuals, such as writing an article. CrowdForge abstracts away many of the programming details of creating and managing subtasks by treating partition/map/reduce steps as the basic building blocks for distributed process flows, enabling complex tasks to be broken up systematically and dynamically into sequential and parallelizable subtasks [76].

4.3.3.6 Sensed data sharing

Tasks are mostly required to be solved using devices with sensing capabilities like smartphones. Given examples in crowdsourcing applications chapter can show varaity of applications can be build using sensing capabilities of devices. Table 4.5 shows sample tasks related to crowdsourcing platforms.

Table 4.5 : Sample tasks.

Voting, Reviewing, Evaluation,
Creative Style Designs, Design surveys
Coding, Algorithms, Project Documentation,
Micro-tasks like Blog Comments, Video sharing, Re-tweet, Downloading apps, Site memberships
Tagging of images, and Videos, Annotations
Noise sensing, Location sharing based tasks, Photo sharing etc.
Translation, Search relevance, Grammar check, Syntax annotation

4.3.4 Task manager

Task manager or the task distributer is gateway for tasks to meet crowd users. Task manager is one of the processing units of crowdsourcing application. Task manager have distinct features, we can list those as ones below:

a) Display Tasks
b) Task Recommendation
c) Worker Recommendation

Display task: Task Manager filters tasks based on profile information of crowd users. Crowd users that are interested working on creative work Task Manager, should filter tasks and not display list of tasks that have to do with voting. For public tasks that have no limitation set, tasks manager displays list of tasks based on search request from crowd user. Tasks created may not be open to all crowd users to work.

As mentioned depending on task context and admission criteria, Crowdsourcer may ask from only specific profession type, specific regions or gender to work on task. Task manager should handle such controls.

Task recommendation: Recommending proper task make crowd users happier and increase motivation to complete task. Today, there are lots of recommendation systems, most of those use basically information provided by users and surfs users do on web. In crowdsourcing applications usage of recommendation method inside task manager will improve entire application performance. We can define generally in four types of recommendation:

- User Profile Information
- Internal User Assessments
- Crowd Community Recommendation
- Crowd User Feedback

User Profile Information: is about users interests, data shared in social media etc. typically obtained in a structured manner as part of a sign-up process.

Internal User Assessments: means that Task Manager based on ranks of users verifies on which tasks they are good and on which ones they like working most. Obtaining information of the user is the key to modeling task recommendation. Profile information and feedbacks from crowd users are important.

Crowd Community Recommendation: Crowd users should be able to recommend task to each other's. In this way cooperation between users increases productivity of application.

Crowd User Feedback: Explicit feedback the data required for learning user preferences reliably is explicit ratings provided by the worker. Implicit feedback refers to information acquired through understanding the user behavior in the platform.

Worker recommendation: Is effective specifically when Crowdsourcer could not find adequate worker for an expertise needed task. Task Manager searches crowd

users for specific expertise needed and recommends to Crowdsourcer.

4.3.5 Evaluation

Evaluation process differs depending on number of participants and volume of data submitted from crowd. Evaluating data gathered from sensors is easier than evaluating data generated by an action of human. Ensuring high quality of results is hard, because of crowd users (workers) has different level of capabilities needed to complete the offered tasks. Cheaters try to get paid as easily as possible, sending worthless responses which makes evaluation process one of the most challenging issues in crowdsourcing applications. There have been a lot of work on evaluation but this issue is still open for new researches. Mostly, responds collected from crowd users does not have unique return structures. For example on voting type applications setting return structure is much easier than on creative work applications. Having binary questions (e.g., Yes/No) or multiple-choice questions (e.g., five-point ratings) are typical examples where return structure are defined. For tasks with unstructured response formats such as article writing and logo design tasks makes real challenge comparison of results with a reference solution. Ipeirotis [78] reported that five of the top twelve Mechanical Turk requesters of the crowdsourcing tasks fall into this category. However, it is not realistic to check all work manually if their volume is large. In Fig. 4.1, we can see that evaluation tasks can be assigned to crowd users to make judgment (reviews). These juries, which are selected crowd users, should be reliable. To assure this Crowdsourcer can prepare datasets with questions where each worker is required to qualify before started reviewing tasks. Qualification of the crowd user is examined through completing several tasks (e.g., questions) selected from the dataset. This method is used but needs extra effort of preparing datasets. Statistical aggregation techniques can also be applied.

Redundancy is also used to ensure work quality. Assigning a single task to multiple crowd workers, and aggregate their responses by applying majority voting.

In a crowdsourcing system, a requester has to decide how to break down a task into several small tasks. A central challenge in crowdsourcing systems is how a task should be designed to induce good output from workers. Several studies performed comprehensive experiments using real datasets to study the impacts of user behavior on the quality of human reviewed data. Mason and Watts [79] showed that increased financial incentives increase the quantity, but not the quality, of work performed by

crowdsourcing workers. It is necessary to derive a set of design principles for tasks on crowdsourcing systems to guarantee the output quality of workers.

4.3.5.1 Quality control approches

Researchers and practitioners have proposed several quality-control approaches that fall under the afore-mentioned quality dimensions and factors. We broadly classify existing approaches into two categories: design-time (see Table 4.6) and run- time (see Table 4.7). These two categories are not mutually exclusive. A task can employ both approaches to maximize the possibility of receiving high-quality outcomes.

At design time, the requesters can leverage techniques for preparing a well-designed task and just allow a suitable crowd to contribute to the task. Although these techniques increase the possibility of receiving high-quality contributions from the crowd, there is still a need to control the quality of contributions at runtime. Even high-quality workers might submit low-quality contributions because of mistakes or misunderstanding. Therefore, requesters must still put in place runtime quality control approaches when the task is running as well as when the crowd contributions are being collected and probably aggregated to build the final task answer.

Table 4.6 : Existing quality-control design-time approaches.

Quality-control approach	Subcategories	Description	Sample application
Effective task preparation	Defensive design	Provides an unambiguous description of the task; task design is defensive — that is, cheating isn't easier than doing the task; defines evaluation and compensation criteria	References 1,3,6,12
Worker selection	Open to all	Allows everybody to contribute to the task	ESP Game, Thredless.com
	Reputation-based	Lets only workers with prespecified reputation levels contribute to the task	MTurk, Stack Overflow, 4
	Credential-based	Allows only workers with prespecified credentials to do the task	Wikipedia, Stack Overflow, 4

Table 4.7 : Existing quality-control runtime approaches.

Quality-control approach	Description	Sample application
Output agreement	If workers independently and simultaneously provide the same description for an input, they are deemed correct.	ESP Game
Input agreement	Independent workers receive an input and describe it to each other.If they all decided that it's a same input, it's accepted as a quality answer.	Tag-A-Tune
Ground truth	Compares answers with a gold standard, such as known answers or common sense facts to check the quality.	CrowdFlower, MTurk
Majority consensus	The judgment of a majority of reviewers on the contribution's quality is accepted as its real quality.	TurKit, Threadless.com, MTurk
Contributor evaluation	Assesses a contribution based on the contributor's quality.	Wikipedia, Stack Overflow, MTurk
Real-time support	Provides shepherding and support to workers in real time to help them increase contribution quality.	Reference 12
Workflow management	Designs a suitable workflow for a complex task; workflow is monitored to control quality, cost, and so on, on the fly.	CrowdForge

4.3.5.2 Image annotation

Using Amazon Mechanical Turk as an example, Snow et al. compared the quality of non-expert annotations and existing gold standard labels for natural language tasks provided by expert labelers [80]. The results demonstrated that it is required to collect an average of 4 non-expert labels per item in order to emulate expert-level label quality, and that the annotation quality can be improved significantly after applying bias correction techniques. Sheng et al. proposed an analysis to model the data quality using repeated labeling with a cost. They found that, with repeated labeling, it is possible to improve the data quality at low cost, especially when labels are noisy. Moreover, when the cost of processing the unlabeled data is not free, repeated labeling is preferable in that it is effective and robust in providing labels of good quality [81]. In 2010, Nowak and Rüger conducted a study about inter-annotator agreement for multi-label image annotation. Although they did not answer the question how many annotation sets of non-experts are necessary to obtain comparable results to expert annotators, they evidenced that different annotators

judge the same data and the inter-annotator agreement among different annotators can ensure the quality [82].

4.3.5.3 Text annotation

Rashtchian et al. [83] found that the use of a qualification test provides the highest improvement of quality of linguistic data collected in MTurk. Hsueh et al. considered the difficult problem of classifying sentiment in political blog snippets. They identified and confirmed the utility of the three selection criteria for high-quality annotations in MTurk: noise level, sentiment ambiguity, and lexical uncer- tainty. In fact, label quality is affected by cognitive awareness of human knowledge [84]. Feng et al. carried out experiments and showed that for the same task turkers answered questions quite differently if they were provided different knowledge in advance. Local search relevance is limited to topical relevance and geographical aboutness [85]. Paiement et al. used interannotator agreement as a quality measure for MTurk labels and discussed a simple approach to select only the most reliable labels. Wikipedia improves through the aggregation of many contributors' efforts [86]. Kittur and Kraut showed that adding more editors to an article improved article quality only when they used appropriate coordination techniques and was harmful when they did not. Implicit coordination through concentrating the work was more helpful when many editors contributed, but explicit coordination through communication was not [87].

4.3.5.4 General tasks

Some work focused on the quality management of general tasks. Huang et al. introduced a general approach for automatically designing tasks on MTurk. They constructed models for predicting the rate and quality of work. These models were trained on worker outputs over a set of designs, and were then used to optimize a task's design. They demonstrated that their models can accurately predict the quality of output per unit task and generate different designs depending on the quality metric. Voyer et al. presented a two-phase, hybrid model for generating training data. They used named entity recognition as an example. In the first phase, a trained annotator labels all named entities in a text irrespective of type. In the second phase, naive crowdsourcing workers complete binary judgment tasks to indicate the type(s) of each entity [88]. Decomposing the data generation task in this way results in a

flexible, reusable corpus that accommodates changes to entity type taxonomies. In addition, it makes efficient use of precious trained annotator resources by leveraging highly available and cost effective crowdsourcing worker pools in a way that does not sacrifice quality.

4.3.5.5 Cheating detection

Due to the anonymity of crowdsourcing workers, malicious workers often try to maximise their financial gains by producing generic answers rather than actually working on the task. Currently, cheat-detection techniques are either based on control questions which are evaluated automatically or rely on manual checking by the requester. Eickhoff and de Vries inspected the commonly observed methods of malicious crowdsourcing workers, such as task-dependent evaluation, interface dependent evaluation and audience-dependent evaluation. Based on experimental results, they concluded that malicious workers are less frequently encountered in novel tasks that involve a degree of creativity and abstraction, and prior crowd filtering can greatly reduce the number of malicious workers [89].

For the crowdsourcing systems that control questions are not applicable and manual re-checking is ineffective, Hirth et al. presented two crowd-based approaches to detect cheating workers: a MD (Majority Decision) and an approach using a CG (Control Group) to re-checking the main task. For MD, the same task is given to several different workers and the results are compared. The result which most of the workers submitted is assumed correct [90]. For CG, a single worker works on a main task and a control group consisting of certain other workers rechecks the result, whether it is valid or not. Usually the main task is expensive, while the recheck task is cheaper. A task is considered valid, if the majority of the control group decides the task is correctly done. Experimental results showed that crowd-based cheat-detection mechanisms are cheap, reliable, easy to implement, and their applicability to different types of typical crowdsourcing tasks.

For some suitations, hiring experts for fraud detection is very expensive. Almendra and Schwabe proposed the use of crowdsourcing to improve precision and recall of current fraud detection techniques for online auction sites. They showed that they could distinguish fraudsters from common sellers before negative feedback arrived and looking just at a snapshot of seller profiles [91].

4.3.6 Platform

The crowdsourcing platform is where the actual crowdsourcing task happens. The crowdsourcing platform can be for internet or mobile devices. Without regarding platform type, all platforms have responsibilities for participants of a crowdsourcing application. Security and privacy of participant's data should be an important concern of the platform. Moreover, it is the duty of the platform to provide safe money transactions and should have control mechanism for the payments [78] related to the tasks performed by the crowd users. All payments are performed using online payment services. For more complicated tasks (like software development, projects with more than one step) platform should push Crowdsourcer to make payment on every milestone finished by crowd user. Platform should have a control mechanism for tasks posted by Crowdsourcer, and reject tasks that introduce life risk of crowd users. Identifying illegal tasks and taking action by closing account or preventing illegal work to be done in the crowdsourcing platform are among the responsibilities of the platform. Platform should provide easy usage and good-looking design in order to attract wide category of people [92]. Crowdsourcer (requester) and Crowd user (worker) login once where they act both as a worker and as an employer. However, in some platforms, Crowdsourcer and crowd user have separate login. Crowdsourcing system models are developed by using centralized or decentralized approaches. The centralized approach is employed when the data generated by the crowd is shipped to a server and the server performs processing. Currently, most of the social networking sites (including Twitter, YouTube, and Facebook) are employing centralized approach. The decentralized systems, in contrast, send queries to smartphones, where all computations and communication are performed locally. In such a model, captured objects remain local on their owner's smartphones and the searches take place by using a lookup in order to minimize energy consumption, and search delay. Although this method decreases energy consumption with less uploads in one direction, it performs poorly in energy consumption if it sends complicated tasks to participants. Continuously transferring data from smartphones to the query processor can deplete the smartphone battery and increase user-perceived delays. In addition, it demands that users disclose their personal data to a central authority.

After reviewing the current literature, we identified four distinct features for the crowdsourcing platform. These features, and their definitions, are as follows: ese features, and their definitions, are described below.

4.3.6.1 Crowd-related interactions

Crowd-related interactions are interactions provided by the crowdsourcing platform between the crowd and the platform. These interactions include, but are not limited to:

- Providing an enrolment mechanism for the crowd to enrol in the crowdsourcing platform.
- Providing an authentication mechanism to authenticate the crowd.
- Providing a declaration mechanism for the crowd to declare their skills and abilities.
- Providing an assignment mechanism for assigning crowdsourced tasks to the crowd.
- Providing an assistive mechanism to assist the crowd in different activities that happen in the crowdsourcing platform, e.g. helping the crowd to enrol, and helping the crowd to perform the crowdsourced task.
- Providing a submission mechanism for the crowd to submit their results.
- Providing a coordination mechanism to coordinate the crowd during crowdsourcing activities.
- Providing a supervision mechanism to supervise the crowd during crowdsourcing activities.
- Providing a feedback loop mechanism to give feedback to the crowd about their crowdsourcing activities.

4.3.6.2 Crowdsourcer-related interactions

Crowdsourcer-related interactions are interactions provided by the crowdsourcing platform between the crowdsourcer and the platform. These interactions include, but are not limited to:

- Providing an enrolment mechanism for the crowd to enrol in the crowdsourcing platform.
- Providing an authentication mechanism to authenticate the crowd.

- Providing a broadcast mechanism for the crowdsourcer to broadcast their crowdsourced task.
- Providing an assistive mechanism to assist the crowdsourcer in different activities that happen in the crowdsourcing platform, e.g. helping the crowdsourcer to enrol, and helping the crowdsourcer to broadcast the crowdsourced task.
- Providing a time negotiation mechanism for the crowdsourcer to negotiate the deadline or duration of the crowdsourced task with the crowd. This mechanism also allows the crowdsourcer to determine a deadline or a permitted duration without negotiation.
- Providing a price negotiation mechanism for the crowdsourcer to negotiate the financial incentives or rewards of completing the crowdsourced task with the crowd. This mechanism also allows the crowdsourcer to determine a fixed price or a reward without negotiation.
- Providing a verification mechanism for the crowdsourcer to verify the results, which are obtained from the crowd.
- Providing a feedback loop mechanism to give feedback to the crowdsourcer about their crowdsourcing activities.

4.3.6.3 Task-related facilities

- Task-related facilities are facilities provided by the crowdsourcing platform about the crowdsourced task. These facilities include, but are not limited to:
- Providing an aggregation mechanism to aggregate the results of a crowdsourced task. The outcome of such aggregation will be sent to the crowdsourcer for further verification, and may also be partially sent to the crowd as part of the feedback.
- Hiding results obtained from one participant in the crowd from other participants.
- Storing history of completed tasks, either for every task, for every crowdsourcer, for every participant, or a combination of those. Such history may be useful in deciding for future task assignments, or for preventing one participant from completing one certain crowdsourced task multiple times, etc.

- Providing a threshold mechanism for the quality of the obtained results to ensure a minimum quality is met.
- Providing a threshold mechanism for the quantity of the obtained results to ensure a minimum and/or maximum quantity is met.
- Platform-related facilities: Platform-related facilities are facilities provided by the crowdsourcing platform about the crowdsourcing platform itself. These facilities include, but are not limited to:
- Providing an online environment which is inherent to online crowdsourcing. However, if crowdsourcing is performed in a real (offline or in-person) environment in a particular crowdsourcing activity, there should still be an environment in which crowdsourcing will take place.
- Managing platform misuse, either by the crowd or by the crowdsourcer.
- Providing an easy, feasible interface both for the crowd and the crowdsourcer to work.
- Providing an attractive, appealing interface both for the crowd and the crowdsourcer to work.
- Providing an interactive interface both for the crowd and the crowdsourcer to work.
- Providing a payment mechanism to allow the crowdsourcer to pay a certain amount of money to the participants. If the reward is something other than money, the platform should also provide mechanisms for it. For example, if the participants should get a certain piece of software or mobile application for free in exchange for their task completion, the platform should provide a download mechanism for the participants.

4.4 Placing Crowdsourcing Reference Model in Cloud Architecture

4.4.1 Cloud-centric architecture

Recently, we can see that corporates moved their applications to the Internet of Things architecture (Cloud-centric). Fig. 4.4 basically illustrates cloud-centric architecture. Moving applications on cloud bring benefits to corporates considering aspects of computing power, higher storage, and effective data analysis and visualization. Furthermore, cloud-centric architecture allows to use applications as SaaS, develop platforms as service (PaaS) and infrastructure as service (IaaS) over it.

Cloud-centric architecture interconnects billions of objects uniquely offering services based on service your get, which is known as pay as you go fashion. Our work, inspect ways of how crowdsourcing applications/platforms can be implemented as SaaS on cloud-centric architecture.

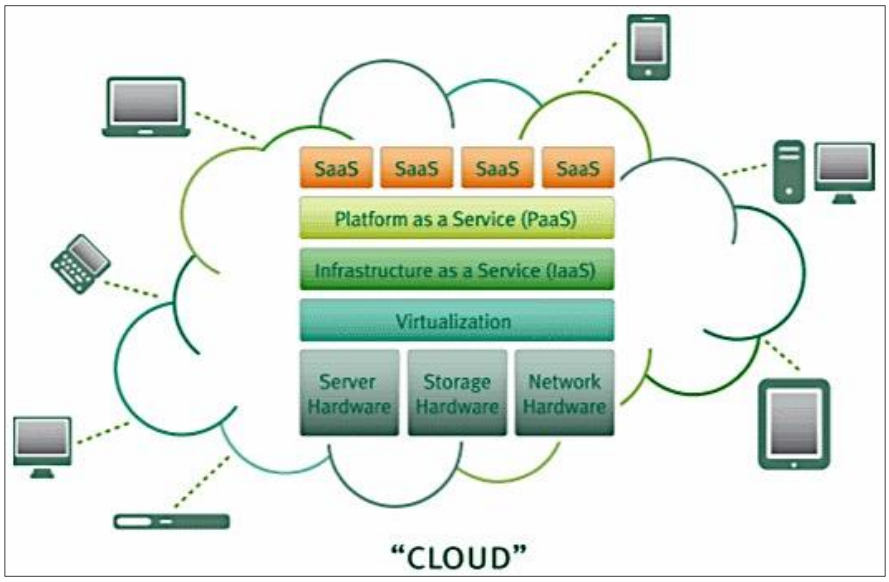

Figure 4.4 : Cloud architecture.

4.4.2 Cloud service models

There are three cloud service models:

- Infrastructure as a Service (IaaS),
- Platform as a Service (PaaS), and
- Software as a Service (SaaS).

With each cloud service model, certain responsibilities are shifted to the cloud service provider allowing consumers of cloud services to focus more on their own business requirements and less on the underlying technologies.

IaaS abstracts the underlying infrastructure and data center capabilities so that consumers no longer have to rack and stack hardware, power and cool data centers, and procure hardware.

PaaS takes us one level higher in the stack and abstracts that operating system, database, application server, and programming language.

Consumers using PaaS can focus on building software on top of the platform and no longer have to worry about installing, managing, and patching LAMP stacks or Windows operating systems. PaaS also takes care of scaling, failover, and many

other technical design considerations so that developers can focus on business applications and less on the underlying IT "plumbing".

SaaS is the ultimate level of abstraction. With SaaS, the entire application or service is delivered over the web through a browser and or via an API. In this service model, the consumer only needs to focus on administering users to the system.

4.4.3 Applicability of the proposed model as service to cloud architecture

Online crowdsourcing replaces the existing business model, which requires massive amount of human intervention with automated, flexible and cloud-inspired service provisioning concept [93]. As cloud computing introduces the concept of Everything as a Service (XaaS) [94], sensing, communication, computing, storage resources of other devices through which recommendations are submitted. Furthermore, concentration, aggregation, storage, analysis and visualization of the crowdsourced data require scalable resources dedicated to the application. Thus, the cloudification of crowdsourcing applications is inevitable.

As cloud computing is transforming the way that the ICT business is run, companies can significantly economize their expenses on crowdsourcing by receiving “crowdsourcing as a service”. There are several crowdsourcing platforms that are tailored for specific applications. Some of these applications need access to the built-in sensors of mobile devices like OpenStreetMap [32] or Gigwalk [30] whereas some are only micro-task-based web platforms like MTurk [50], or InnoCentive [42]. Next generation wireless communication technologies have made the mobile handheld devices participate in sensing, computing and communication tasks. Despite the availability of crowdsourcing platforms and applications, the state of the art does not provide any reference model. To the best of our knowledge, in this paper we fill this gap by defining a generic reference model that uses the components and attributes of "crowdsourcing as a service" concept.

Generally, crowdsourcing systems have limitations on resources, like storage, computation power, platform dependencies and management. Cloud-centric architecture appears as a promising solution for such systems. Implementing our crowdsourcing model in cloud-centric architecture can improve the computation capability, increase participation of the crowd and energy efficiency of mobile

devices by offloading computation tasks onto cloud servers. Such architecture requires a service based work model. Fig 4.5 depicts crowdsourcing as a service architecture implemented over the cloud. Incoming request for services through web server are handled by the task manager which finds the proper task, and publishes it on online social network applications. Crowds connected with Internet can co-operate on each task simultaneously and submit their results.

On other hand, mobile crowdsourcing is a type of electronic commerce service, where mobile users form a mobile cloud to sell cloud resources and services (e.g., data collecting, computing, and processing) for service consumers. Different from the traditional cloud computing that depends on Internet connection, mobile crowdsourcing can provide pervasive cloud services for both online and local terminals. The main difference between the two kinds of mobile crowdsourcing models is that all the Internet-connected mobile users can potentially be a service provider in the Internet-based mobile crowdsourcing, while only the mobile users in the vicinity can provide cloud services in local-based mobile crowdsourcing.

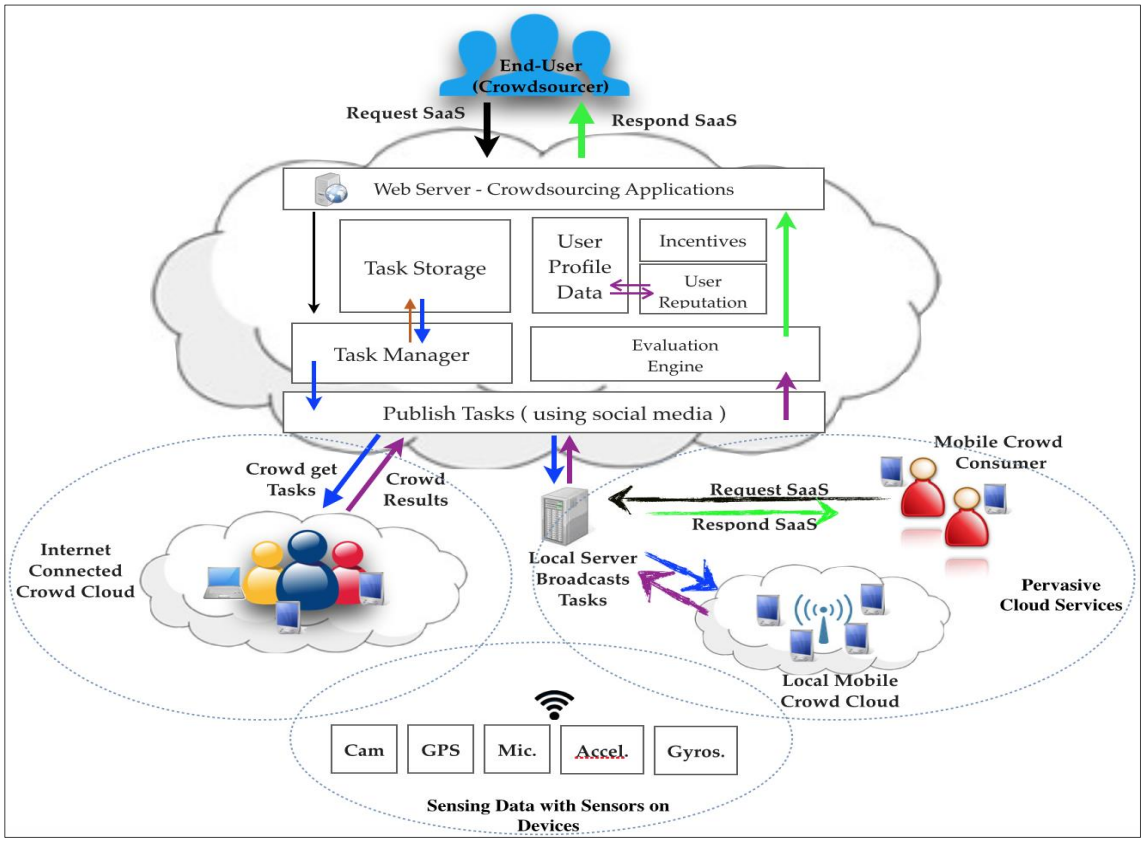

Figure 4.5 : Crowdsourcing as service over a cloud platform

Furthermore, a large number of mobile devices connect with each other via wireless

networks, forming an unprecedentedly powerful mobile cloud to provide pervasive data collection, processing, and computing services. As mobile users become service providers, social relationships and interactions play a significant role in mobile crowdsourcing. It poses a particular challenge on exploiting the underlying social impacts, such as personal social attributes, preference, selfishness, etc.

Therefore, with the advent of mobile cloud model, mobile crowdsourcing has gained momentum as a feasible solution for solving very large-scale problems. By outsourcing tasks to the mobile cloud, cost-effective and pervasive cloud services can be achieved, using a possibly huge number of mobile users and devices to work together in a distributed way.

5. CONCLUSIONS

In this thesis, we study extensively existing literature on crowdsourcing, including platforms, applications, books, papers, articles and blogs. The main research objective of the thesis was to identify the components and activities within crowdsourcing process and design a reference model. While designing the reference model we improved reference models proposed earlier and adapt it to cloud systems. Previous models have concentrated on entities separately without presenting a clear process overview. We analyzed and extract all known components and properties of different type of applications and proposed a reference model. We outlined crowdsourcing process in four phases, present the components of each phase in detail, and give the interaction between these components considering cloud structure. We started with user registrations, task creation and distribution, evaluation techniques, user reputations and participations and finally concluded with payments. We tried to clarify the specificities and issues related to Crowdsourcing in order to contribute to this growing phenomenon. Hopefully that our work will show directions for web and mobile platform designers for better designs employing crowdsourcing as a service.

Future directions

Crowdsourcing is enabled through the technology of the web, and have spread widely with mobile computing in the form of smartphones, tablets, and so on. Today tremendous amount of human computation power is available for accomplishing jobs almost for free. The tremendous data generated by web platforms and smartphone applications (or sensors) will enable new challenging applications and solve harder problems than that is possible with crowds today. The focus of future work in Crowdsourcing field will be concentrated more in development of specialized platforms. The proposed reference model in this thesis aims to contribute to this aspect. Moreover presenting crowdsourcing as a service will lead it to be employed widely.

Endowing crowdsourcing services with customizable, rich, and robust quality-control techniques will contribute to the succsess of the crowdsourcing platforms. Integrating crowdsourcing platforms with such quality control system seamlessly is an important issue to be considered. Building such a complete framework can be a challenging future direction.

Topics such as machine versus human trustworthiness, workflow design for such tasks, and conflict resolution between human and machine judgments will all need to be addressed.

It has been almost a decade when Jeff Howe introduced a new research area named Crowdsourcing. Although various applications have been developed, we believe that crowdsourcing will continue to attract the attention of the developers and researchers. With the development of remote working culture, crowdsourcing will be a part of our daily life as an employer or an employee in the next few years. There is no doubt that new generation platforms and crowdsourcing applications will continue to improve and keep a very important place in our daily life in the future.

REFERENCES

[1] **Tokarchuk, O., Cuel, R. and Zamarian, M.** (2012). *Analyzing Crowd Labor and Designing Incentives for Humans in the Loop***,** Vol. 16, Issue 5, pp. 35 – 51, Sept.-Oct. 2012.

[2] **Howe, J.** (2006). Crowdsourcing: A Definition, Crowdsourcing Blog

[3] **Url-1** <http://www.businessnewsdaily.com/4025-what-is-crowdsourcing.html>, data retrieved 02.01.2015.

[4] **Url-2** < http://essentialsofbusiness.ufexec.ufl.edu/> , data retrieved 02.01.2015.

[5] **Cheung, S.** (2012). *How Companies Can Leverage Crowdsourcing*, *MSc Thesis,* MIT, Cambridge, MA.

[6] **Schenk, E. and Guittard, C.** (2009). Crowdsourcing: What can be Outsourced to the Crowd, and Why?

[7] **Hosseini, M., Phalp, K., Taylor, J. and Ali, R.** (2014). The Four Pillars of Crowdsourcing: a Reference Model, *IEEE 8th International Conference on Research Challenges in Information Science*, Marrakech, Morroco, 28-30 May

[8] **Chatzimilioudis, G., Konstantinidis, A., Laoudias, C. and Yazti, Z. D**. (2012). Crowdsourcing with Smartphones, *IEEE Internet Computing*, Vol. 16, Issue 5, pp. 36 – 44, Sept.-Oct. 2012.

[9] **Url-3** < http://www.forbes.com>, data retrieved 15.01.2015.

[10] **Brabham, D. C.** (2008). Crowdsourcing as a Model for Problem Solving, *The International Journal of Research into New Media Technologies, Vol.* 14, Issue 1, pp. 75 – 90.

[11] **Consolvo, S.** (2008). Activity Sensing in the Wild: A Field Trial of Ubifit Garden, *26th Annual ACM SIGCHI Conf. Human Factors Comp. Sys.*, pp. 1797–1806.

[12] **Miluzzo, E.** (2008). Sensing meets Mobile Social Net- works: The Design, Implementation, and Evaluation of the CenceMe Application, *6th ACM SenSys*, pp. 337–50.

[13] **Mun, M.** (2009). Peir, the Personal Environmental Impact Report, as a Platform for Participatory Sensing Systems Research, *7th ACM MobiSys*, pp. 55–68.

[14] **Thiagarajan, A.** (2009). VTrack: Accurate, Energy-Aware Traffic Delay Estimation Using Mobile Phones, *7th ACM SenSys*, Berkeley, CA,

[15] **Url-4** < http://traffic.berkeley.edu/>, data retrieved 17.02.2015.

[16] **Lu, H**. (2009). Sound-Sense: Scalable Sound Sensing for People-Centric Applications on Mobile Phones, *7th ACM MobiSys*, pp. 165–78.

[17] **Url-5** <http://sensorlab.cs.dartmouth.edu/>, data retrieved 17.02.2015.

[18] **Campbell, A. T., Lane, N. D., Miluzzo, E., Peterson, R. A., Lu, H., Zheng, X., Musolesi, M. and Fodor, K.** (2008). The Rise of People-Centric Sensing. *IEEE Internet Computing*, Vol. 12, Issue 04, pp. 12-21, July –Aug 2008.

[19] **Url-6** <http://www.clickworker.com>, data retrieved 13.02.2015.

[20] **Url-7** <http://www.myhumangrid.com>, data retrieved 11.02.2015.

[21] **Url-8** <http://www.vworker.com>, data retrieved 13.02.2015.

[22] **Url-9** <http://www.crowdflower.com>, data retrieved 12.02.2015.

[23] **Url-10** <http://www.odesk.com>, data retrieved 11.02.2015.

[24] **Url-11** <http://www.microworkers.com>, data retrieved 11.02.2015.

[25] **Url-12** <http://www.shorttask.com>, data retrieved 12.02.2015.

[26] **Url-13** <http://www. atizio.com>, data retrieved 12.02.2015.

[27] **Url-14** <http://www.fotolia.com>, data retrieved 13.02.2015.

[28] **Url-15** <http:// www. threadless.com>, data retrieved 14.02.2015.

[29] **Url-16** <http://www. easyshiftapp.com>, data retrieved 14.02.2015.

[30] **Url-17** <http://www.Gigwalk.com>, data retrieved 15.02.2015.

[31] **Url-18** <http://www.work.mobileworks.com>, data retrieved 11.02.2015.

[32] **Url-19** <http://www.openstreetmap.org>, data retrieved 12.02.2015.

[33] **Url-20** <http:// www. mClerk.com>, data retrieved 14.02.2015.

[34] **Kantarci, B. and Mouftah**, **H. T.** (2015).“Sensing Services in Cloud- Centric Internet of Things: A Survey, taxonomy and challenges,” *IEEE ICC Workshop Proceedings,* pp.9611-9616, June 2015.

[35] **Kantarci, B. and Mouftah**, **H. T.** (2014). “Reputation-based Sensing as as Service for crowd management over the cloud,” in *IEEE Intl. Conf. on Communications (ICC)*, June 2014.

[36] **Kantarci, B. and Mouftah**, **H. T.** (2014). “Trustworthy sensing for public safety in cloud-centric Internet of Things,” *IEEE Internet of Things Journal*, vol. 1/4, pp. 360–368, Aug 2014.

[37] **Yang, D., Xue, G., Fang, X. and Tang**, **J.** (2012). “Crowdsourcing to smartphones: Incentive mechanism design for mobile phone sensing” *Intl Conf. on Mobile Computing and Networking*, Aug. 2012, pp. 173–184.

[38] **Al-Fagih, A.E., Al-Turjman, F.M., Alsalih, W.M. and Hassanein**, **H.S.** (2013). “A priced public sensing framework for heterogeneous IoT architectures,” *IEEE Trans. on Emerging Topics in Computing*, Vol. 1/1, pp. 133–147, June 2013

[39] **Rodríguez, A. F. and Müller, H.** (2012). Ground Truth Generation in Medical Imaging: A Crowdsourcing-Based Iterative Approach, ***CrowdMM '12*** *Proceedings of the ACM multimedia 2012 workshop on Crowdsourcing for multimedia,* pp.9-14. New York, NY.

[40] **Akgu ̈l, C., Rubin, D., Napel, S., Beaulieu, C., Greenspan, H. and Acar, B.** (2011). Content–based image retrieval in radiology: Current status and futuredirections. *Journal of Digital Imaging*, 24(2):208–222,

[41] **Chanal, V., Laurence, M. and Fasan, C.** (2008). How to invent a new business model based on crowdsourcing: the Crowdspirit case, *EURAM 2008 Conference,* Lubjana, Slovenia.

[42] **Url-21** <http://www.innocentive.com>, data retrieved 31.01.2015.

[43] **Url-22** <http://www.oxforddictionaries.com >, data retrieved 31.01.2015.

[44] **Fraternali, P., Castelletti, A., Soncini-Sessa, R., Vaca Ruiz, C. and Rizzoli, A. E.** (2012). Putting humans in the loop: Social computing for Water Resources. *Environmental Modelling & Software*, Vol. 37, November, pp. 68–77

[45] **Matyas, S., Matyas, C., Schlieder, C., Kiefer, P., Mitarai, H. and Ka- mata, M.** (2008). Designing Location-Based Mobile Games With A Purpose: Collecting Geospatial Data with CityExplorer. *In ACM ACE*.

[46] **Mellebeek, B., Benavent, F., Grivolla, J., Codina, J., Costa-Jussa, M.R. and Banchs**, **R.** (2010). Opinion Mining of Spanish Customer Comments With Non- Expert Annotations on Mechanical Turk. *In Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk,* CSLDAMT '10, pp. 114–121, Morristown, NJ, USA.

[47] **Jain, S., Chen, Y. and Parkes**, **D. C.** (2009). Designing Incentives for Online Question and Answer Forums. In Proceedings of the 10th ACM conference on Electronic commerce, EC '09, pp. 129–138, New York,NY, USA.

[48] **Brabham**, **D. C.** (2008). Moving the Crowd at iStockphoto: The Composition of the Crowd and Motivations for Participation in a Crowdsourcing Application.

[49] **Estellés-Arolas, E. and González-Ladrón-de-Guevara**, **F.** (2012). Towards an Integrated Crowdsourcing Definition. *Journal of Information Science,* 38(2), pp. 189-200.

[50] **Url-25** <http://www.mturk.com >, data retrieved 19.02.2015

[51] **Jagadeesan, A. P., Lynn, A., Corney, J. R., Yan, X. T., Wenzel, J., Sherlock, A. and Regli, W.** (2009). Geometric Reasoning Via Internet Crowdsourcing. *In 2009 SIAM/ACM Joint Conference on Geometric and Physical Modeling*, SPM '09, pp. 313–318, New York, NY, USA.

[52] **Heer, J. and Bostock, M.** (2010). Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. *In Proceedings of the 28th international conference on Human factors in computing systems*, CHI '10, pp. 203–212, New York, NY, USA.

[53] **C.Higgins, E.McGrath, and L.Moretto**.Mturkcrowdsourcing: Aviable Method For Rapid Discovery Of Arabic Nicknames? In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language*

Data with Amazon's Mechanical Turk, CSLDAMT '10, pp. 89–92, Morristown, NJ, USA.

[54] **Finin, T., Murnane, W., Karandikar, A., Keller, N., Martineau, J. and Dredze, M.** (2010). Annotating Named Entities In Twitter Data With Crowdsourcing. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, CSLDAMT '10, pp. 80–88, Morristown, NJ, USA.

[55] **Mellebeek, B, Benavent, F., Grivolla, J., Codina, J., Costa-Jussa, M.R. and Banchs**, **R**. (2010). Opinion Mining of Spanish Customer Comments With Non- Expert Annotations on Mechanical Turk. *In Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk,* CSLDAMT '10, pp. 114–121, Morristown, NJ, USA.

[56] **Gordon, J., Benjamin, V. D. And Schubert, L. K.** (2010). Evaluation of Commonsense Knowledge With Mechanical Turk. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, CSLDAMT '10, pp.159–162, Morristown, NJ, USA.

[57] **Wang, R. and Callison, C.** (2010). Burch. Cheap Facts and Counter Facts. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk,* CSLDAMT '10, pp. 163–167, Morristown, NJ, USA

[58] **Alonso, O.,Rose, D. E. and Stewart, B**. (2008). Crowdsourcing for relevance evaluation. *SIGIR Forum*, 42:9–15, November 2008.

[59] **Akkaya, C., Conrad, A., Wiebe, J. and Mihalcea, R.** (2010) Amazon Mechanical Turk for Subjectivity Word Sense Disambiguation. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk,* CSLDAMT' 10,pp. 195–203, Morristown, NJ, USA.

[60] **Callison-Burch, C. and Dredze, M.** (2010). Creating Speech and Language Data with Amazon's Mechanical Turk. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, CSLDAMT '10, pp. 1–12, Morristown, NJ, USA.

[61] **Gao, Q. and Vogel, S.** (2010). Consensus Versus Expertise: A Case Study of Word Alignment with Mechanical Turk. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, CSLDAMT '10, pp. 30–34, Morristown, NJ, USA, 2010.

[62] **Jha, M., Andreas, J., Thadani, K. Rosenthal, S. and McKeown, K.** (2010) Corpus Creation for New Genres: A Crowdsourced Approach to pp Attachment. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, CSLDAMT '10, pp. 13–20, Morristown, NJ, USA.

[63] **Parent, G. and Eskenazi, M.** (2010). Clustering Dictionary Definitions Using Amazon Mechanical Turk. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, CSLDAMT '10, pages 21–29, Morristown, NJ, USA.

[64] **Skory, A. And Eskenazi, M.** (2010). Predicting Cloze Task Quality for Vocabulary Training. In *Proceedings of the NAACL HLT 2010 Fifth Workshop on Innovative Use of NLP for Building Educational Applications*, IUNLPBEA '10, pp. 49–56, Morristown, NJ, USA.

[65] **Maisonneuve, N., Stevens, M., Niessen, M. E., Hanappe, P. and Steels, L. V** (2009). Citizen Noise Pollution Monitoring. In *Proceedings of the 10th Annual International Conference on Digital Government Research: Social Networks: Making Connections between Citizens, Data and Government*, pp. 96–103, North America,

[66] **Choffnes, D. R., Bustamante, F. E. and Ge, Z**. (2010). Crowdsourcing Servicelevel Network Event Monitoring. *SIGCOMM Comput. Commun. Rev.*, 40:387–398

[67] **Golder, S. A. and Huberman, B. A.** (2005). The structure of collaborative tagging systems. *CoRR*, abs/cs/0508082.

[68] **L. von Ahn.** (2006). Games With a Purpose. *IEEE Computer*, 39(6):92–94, June.

[69] **L. von Ahn and Dabbish, L.** (2004). Labeling Images With a Computer Game. In *ACM SIGCHI Conference on Human Factors in Computing Systems*,

[70] **L. von Ahn, Liu, R. and Blum, M**. (2006). Peekaboom: A Game for Locating Objects in Images. In *ACM SIGCHI Conference on Human Factors in Computing Systems.*

[71] **L. von Ahn, Ginosar, S., Kedia, M. and Blum, M**. (2007). Improving Image Search with PHETCH. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2007.

[72] **Url-23** < http://www.gwap.com/gwap/>, data retrieved 19.02.2015.

[73] **Chan, K. T., King, I. and Yuen, M.-C.** (2009). Mathematical Modeling of Social Games. In *CSE '09: Proceedings of the 2009 International Conference on Computational Science and Engineering*, pp.1205–1210. IEEE Computer Society.

[74] **Url-24** < http://www.thesheepmarket.com/>, data retrieved 23.02.2015.

[75] **Leimeister, J., Huber, M., Bretschneider, U. and Krcmar. H.** (2009). Leveraging Crowdsourcing: Activation-Supporting Components for It-Based Ideas Competition. *J. Manage. Inf. Syst.*, 26:197–224, July 2009.

[76] **Kittur, A., Smus, B. and Kraut, R.** (2011). Crowdforge: Crowdsourcing Complex Work. In *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems*, CHI EA' '11, pp. 1801–1806, New York, NY, USA

[77] **Little, G., Chilton, L. B., Goldman, M. and Miller, R. C.** (2010). Turkit: Human Computation Algorithms on Mechanical Turk. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, UIST '10, pp.57–66, New York, NY, USA.

[78] **Ipeirotis, P. G., Provost, F. and Wang, J.** (2010). Quality Managament on Mechanical Turk. In Proceedings of HCOMP '10 Proceedings of the ACM SIGKDD Workshop on Human Computation, pp.64–67, New York, NY, USA.

[79] **Mason, W. and Watts, D. J.,** (2010). Financial Incentives and the "Performance of Crowds". *SIGKDD Explor. Newsl.*, 11:100–108.

[80] **Snow,R., O'Connor, B., Jurafsky, D. and. Ng, A. Y.** (2008). Cheap and Fast – But is it Good?: Evaluating Non-Expert Annotations for Natural Language Tasks. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '08, pp. 254–263, Morristown, NJ, USA.

[81] **Sheng, V. S., Provost, F. and Ipeirotis, P. G.,** (2008). Get Another Label? Improving Data Quality and Data Mining Using Multiple, Noisy Labelers. In *ACM KDD*, 2008.

[82] **Nowak, S. and Rüger, S.** (2010). How Reliable are Annotations via Crowdsourcing: A Study About Inter-annotator Agreement for Multi-label Image Annotation. In *Proceedings of the international conference on Multimedia information retrieval*, MIR '10, pages 557–566, New York, NY, USA. ACM.

[83] **Rashtchian, C., Young, P., Hodosh, M. and Hockenmaier, J.** (2010). Collecting Image Annotations Using Amazon's Mechanical Turk. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, CSLDAMT '10, pp. 139–147, Morristown, NJ, USA.

[84] **Hsueh, P.-Y., Melville, P. and Sindhwani, V.** (2009). Data Quality From Crowdsourcing: A study of Annotation Selection Criteria. In *Proceedings of the NAACL HLT 2009 Workshop on Active Learning for Natural Language Processing*, HLT '09, pp. 27–35, Morristown, NJ, USA,

[85] **Feng, D., Besana, S., Boydston, K. and Christian, G.** (2010). Towards High-Quality Data Extraction Via Crowdsourcing. In *Proceedings of the CrowdConf*.

[86] **Paiement, J.-F., Dr. Shanahan, J. G. and Zajac, R.** (2010). Crowdsourcing Local Search Relevance. In *Proceedings of the CrowdConf 2010*, CrowdConf.

[87] **Kittur, A. and Kraut, R. E**. (2008). Harnessing The Wisdom of Crowds In Wikipedia: Quality Through Coordination. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, CSCW '08, pp. 37–46, New York, NY, USA, ACM.

[88] **Voyer, R., Nygaard, V., Fitzgerald, W. and Copperman, H.** (2010). A Hybrid Model For Annotating Named Entity Training Corpora. In

Proceedings of the Fourth Linguistic Annotation Workshop, LAW IV '10, pp. 243–246, Morristown, NJ, USA.

[89] **Eickhoff, C. and A. P. de Vries.** (2011). How Crowdsourcable Is Your Task? In *Proceedings of the Workshop on Crowdsourcing for Search and Data Mining (CSDM 2011)*, WSDM 2011, pp. 11–14, New York, NY, USA, ACM.

[90] **Hirth, M., Hoßfeld, T. and Tran-Gia, P.** (2010). Cheat-Detection Mechanisms For Crowdsourcing. Technical Report 474, University of Würzburg.

[91] **Almendra, V. and Schwabe, D**. (2009). Fraud Detection by Human Agents: A Pilot Study. In *Proceedings of the 10th International Conference on E-Commerce and Web Technologies*, EC-Web 2009, pp. 300–311, Berlin, Heidelberg, Springer-Verlag.

[92] **Ipeirotis, P. G.** (2010). Analyzing The Amazon Mechanical Turk Marketplace. *XRDS*, 17:16–21.

[93] **Ben Amor, I., Ouziri, M., Sahri, S. and Karam, N.** (2014). "Be a Collaborator and a Competitor in Crowdsourcing System," *IEEE 22nd Intl. Symp. Modelling, Analysis & Simulation of Computer and Telecommunication Systems*, pp. 158—167, Sep. 2014.

[94] **Moreno-Vozmediano, R., Montero, R.S. and Llorent**, **I. M.** (2013). "Key challenges in cloud computing: Enabling the future internet of services," *IEEE Internet Computing*, vol 17/24, pp. 18—25, 2013

CURRICULUM VITAE

Name Surname: Arbër Murturi

Place and Date of Birth: Pristina, Kosovo 12.09.1985

E-Mail: amurturi@gmail.com

EDUCATION:

B.Sc.: Yıldız Technical University – Computer Engineering

PROFESSIONAL EXPERIENCE AND REWARDS:

Senior SAP Technical Application Consultant

Certified Technical Application Consultant (ABAP Programming)

Certified Ethical Hacker, EC-Council

Certified in PLD systems as programmer, University of Leipzig Germany

Istanbul Technical University, Former Teaching assistant on Operating Systems

Work as developer in HP Software Innovation Centre

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

Arber Murturi, Burak Kantarcı and Sema Oktuğ (2015). " A Reference Model for Crowdsourcing as a Service", *IEEE 4th International Conference on Cloud Networking*, Niagara Falls, Canada, October 201

Arber Murturi, Prof. Dr. Ahmet Coşkun Sönmez 2009: "Aidbook End Poverty" Yıldız Technical University, EMO project competition August 2009, Istanbul.