RESEARCH ARTICLE   OPEN ACCESS

Crowdsourcing: A Survey

Rukhsana Thaker
Department of Computer Science, BGSB University, Rajouri, J&K

Abstract-
Crowdsourcing is an online, distributed problem solving and production model that revolutionized the internet and mobile market at present. It turns the customers into designer and marketers. The practice of Crowdsourcing is transforming the web and giving rise to a new field. Today the leading enterprises are embracing the next paradigm shift in the distribution of work by outsourcing to the crowd in the cloud. Everyday millions of people make all kind of voluntary online contribution. With the number of people online approaching 3 billion by 2016 and projected to reach 5 billion by 2020, new workforce has emerged that are now used for different purposes. Available on-demand this workforce has abundant capacity and the expertise knowledge to perform work from simple to complex and solve problems and grand challenges. This paper gives an introduction to Crowdsourcing, its theoretical grounding, model and examples with case study. In this paper we show that Crowdsourcing can be applied to wide variety of problems and that it raises numerous interesting technical and social challenges. Finally this paper proposes an agenda for using Crowdsourcing in NLP.
Index Terms—**Innovation, crowdfunding, crowdwisdom, crowdlabour.**

I. Introduction

The term "crowdsourcing" is a portmanteau of "crowd" and "outsourcing," coined by Jeff Howe and Mark Robinson and published in a June 2006 *Wired* magazine article "The Rise of Crowdsourcing".[1] It has been argued that crowdsourcing can only exist on the Internet and is thus a relatively recent phenomenon.[2] In crowdsourcing, we can outsource the task to not only a small group of person, but also tens of thousands of people. That's the genuine advantage of the crowdsourcing, bringing in mass intelligence to solve problems of all kinds with affordable price. Despite being a somewhat new term, "crowdsourcing" has been existed in our society for years in many forms. For instance, video game's beta invitation is a form of crowdsourcing, where a developer releases a beta version of the game to the public for testing. This has been going on in the software world for years, where certain softwares are tested by crowdsourcing companies, while others are tested publicly. YouTube receives hundreds of thousands of videos daily. Every minute, 24 hours of videos are uploaded to youtube. Crowdsourcing is often used for translation service**.** For instance, Facebook has been using crowdsourcing to translate its website into different languages since 2008. Wikipedia can be categorized as crowdsourcing as well: an encyclopaedia written by people from all over the world, professionals and amateurs alike. Crowdsourcing has been with us for a while and now, it just got a catchy name. Companies nowadays not only employ Research and Development (R&D) teams, but list the problems they're working on to the public as an attempt to get

help from anyone**.** Certain companies even offer huge prizes for solving a problem on crowdsourcing site like InnoCentive, and the cost is fairly cheap compared to employing a huge R&D team. Crowdsourcing has captured the attention of the world recently. Numerous tasks or designs conventionally carried out by professionals are now being crowdsourced to the general public to perform in a collaborative manner [5]. This new Web-based business model called *crowdsourcing* attempts to harnesses the creative solutions of a distributed network of individuals established with the goal to *outsource* tasks to these workers. This network of humans is typically an open Internet-based platform that follows the open world assumption and tries to attract members with different knowledge and interests [9]. Large IT companies such as Amazon, Google, or Yahoo! have recognized the opportunities behind such mass collaboration systems for both improving their own services and as business case. While the aforementioned specifications have been created to model human interactions in BPEL, it remains an open issue how to apply them to crowdsourcing. Specifically, the WS-HT specification does not define any particular mechanisms to find or select people in open and dynamic environments. Instead, a *Logical People Group* is used to query an organizational directory. We believe that human interactions in SOA need to be supported in a flexible manner; in particular, it should be possible to use crowdsourcing for process execution.

As previously mentioned, crowdsourcing is best suited for simple tasks**.** For instance, crowdsourcing is a great option for web designers and designers in general to get some usability feedback on their work before making it public. If you're looking for a new logo, hiring a professional designer is an option, but if you are a start up and you cannot afford such professional, you can consider placing a competition on the crowdsourcing site to get the product with an affordable price. Anyone building a huge library of photos, products, etc, can easily take advantage of crowdsourcing features. You may be building a new online shopping site, where you could use crowdsourcing and pay a few cents for people to describe, categorize or tag the inventory**.** Or perhaps you're working on a new website design, you could ask people to fill out a questionnaire regarding the new design**,** and from there you can get some inspiration**.**

II. Steps involved in Crowdsourcing

Crowdsourcing is a method of open innovation that calls on a large group of people or a community to generate new ideas, solve complex problems, or complete a task. It pulls on the idea that there is "wisdom in the crowds," and that through Internet resources we can harness this wisdom. This kind of open innovation has proven to be an important research and discovery tool.

Three main steps are:

Open Innovation – challenges that encourage public participation in research and lower barriers to entry.

Data Collection – collect targeted datasets at low cost to supplement and add depth to traditional survey research data.

Analysis – use humans to do analysis of social media data.

A. Open Innovation

The goal of open innovation is to Lower or eliminate barriers to entry in research. Use open innovation in the research design process. Challenge all members of research communities to contribute to idea generation and research solutions. Experts are smart, but they do not always have the answers. You may discover an idea you never would have thought of.

Case Study:

In 2012 RTI started a research challenge. It planned an omnibus survey in Chicago. They asked researchers from a variety of subject areas to contribute to the survey. Two pages of synopsis with 10 survey questions were to be submitted within 23 days. They received total 76 entries. Research topics were emerging tobacco products, testing astrological theories, Perceptions of law enforcement. Scored were based upon evaluation criteria i.e. Quality of research idea, Quality of survey questions, Relevance to Chicago. Eventually they selected nine winners.

Percentage of participant and winners from different area is show below:

Table 1: RTI Research challenge result

Employment	Percentage of Participants	Percentage of Award Winners
Academia	66.1%	22.2%
Nonprofits	1.7%	-
Government	5.1%	11.1%
Private Sector	13.6%	-
Student	25.4%	66.7%
Other	5.1%	-

B. Data Collection

The abundance of structured and social data on the Web coupled with ability to solicit feedback from crowds has the potential of changing the way we search for information and enabling new classes of applications on the Web. Structured data appear on the Web in several forms, including hidden Web sources exposed through HTML form interfaces, tables, lists, and pages with repeating semi structured cards. Current research efforts for leveraging this data include approaches for extracting and combining results from multiple sources, for surfacing the deep web, and for exposing data through RDF repositories with rich linking, possibly exploiting existing knowledge bases for data annotation and integration. Crowdsourcing and social data are increasingly popular methods for improving search results and enhancing the quality of data on the Web. Social data have huge potential to re-rank and enrich pages and content based on what the user's friends have visited or recommended previously. Crowdsourcing can be used to answer questions that are inherently hard for machines but can be handled relatively easily with human input.

Case Study:

One of the 2012 Research Challenge winners focused on perceptions of snus tobacco. He used Amazon Mechanical Turk for his research. Let us see how this Mechanical Turk Works.

Figure 1. Amazon Mechanical Turk

Mechanical Turk is used to offer HITs (Human Intelligence tasks) to workers. First of all requester designs HIT template with payment detail. After that it launches HIT. Interested workers work on those

jobs and after completion requester analyse the result. If the result is satisfactory then requester pays worker otherwise reject work that is unsatisfactory.

Now we will see how this mechanical Turk helps in finding the tobacco retailers of Chicago who sold snus tobacco. First step is to break the complex task into small task. Then asks workers to do each task one by one.

Task 1: Collect name and address of tobacco retailer.

Task 2: Get the phone number of these retailers.

Task 3: Call retailer and ask if he sells snus tobacco product.

All these tasks are performed by Mechanical Turk workers.

C. Sentiment Analysis

It is about determining the subjectivity, polarity (positive or negative) and polarity strength (weakly positive, mildly positive, strongly positive, etc.) of a piece of text. Text analysis software is getting better but it still has some problem. It cannot always detect sarcasm, cannot understand slang. Again we can use Mechanical trunk for such analysis.

III. Applications of crowdsourcing

A. Oxford Dictionary

In the late-nineteenth century Professor James Murray was leading a literary project that draws from the knowledge, expertise and time of tens of thousands of volunteers. Operating in Oxford, England, he received hundreds of thousands of slips of paper over the course of several decades, each containing the definition of a particular English word. To Murray, these contributors were strangers -- unpaid, but worked together as one to collate the definitions and origins of every word in the English language. It took 70 years to complete, but the *Oxford English Dictionary* was arguably history's first massively-crowdsourced collation of English knowledge.

Initially published steadily in numerous volumes, Volunteers requested books from Murray and were tasked with recording each instance of a given word, along with its date of first use, its meaning within context, and other data. Who these volunteers were was of trivial importance. In fact, little did Murray know that for 20 years, one of the most prominent contributors was a mentally unstable ex-US military surgeon named William Chester Minor. Minor contributed from the Broadmoor asylum after killing a man during a paranoid episode, but being found not guilty on account of his insanity. But it didn't matter. Minor was remarkably well-read and highly educated, so his extremely detailed contributions of thousands of definitions was just as valuable - even though the personal circumstance behind them was unknown for two decades, at least to Murray. The "metadata" recorded by Minor and other volunteers was recorded on special slips of paper, and then sent to Oxford. Multiple citations and definitions of identical words were manually considered, before being entered into the dictionary.

B. Crowd Sourcing in Business

What happens when a company lets consumers design and vote on their own product? The business model is based on so-called crowd sourcing, or community-based design. One of those companies is RYZ, a tiny, high-end sneaker company in Portland, Ore. Like other companies relying on community design, RYZ doesn't need a large marketing or design staff. It uses potential customers for that. Would-be designers use a template from the company's Web site to create a pair of high-rise sneakers. The sneaker designs are posted online, and viewers vote on which ones they like. Winning designs are produced, and the designer gets $1,000 plus 1 percent royalties.

There's practically no overhead involved. Marketing costs? Practically nothing. The business model relies largely on the Internet, hoping that online voting and buzz on sites such as MySpace and Facebook will create demand for specific products.

C. Web Security Words Help Digitize Old Books- crowdsourcing

People who use the Internet to talk to friends, set up free e-mail accounts or buy concert tickets are often unknowingly helping to digitize vast libraries of old books and newspapers. That's because more than 40,000 Web sites — including popular ones such as Ticketmaster, Facebook and Craigslist — are using a new kind of security program called reCAPTCHA. It's the brainchild of Luis von Ahn, a computer scientist at Carnegie Mellon University in Pittsburgh, who helped develop another commonly used Web security system. That one, called CAPTCHA, will allow people to access a Web site only if they prove they are human — and not a spammer's computer — by typing in a sequence of letters or numbers that appear on the screen in a distorted or garbled image. Each time we type one of these, our brain is doing something amazing. Our brain is performing a task that, despite 50 years of research in computer science, we cannot yet get computers to do. The trouble is, each time we type in one of these garbled words, and we are also wasting time. Approximately 200 million of these are typed every day by people around the world. Each time we type one of these, essentially we waste about 10 seconds of our time. If we multiply that by 200 million, we get that humanity as a whole is wasting around 500,000 hours every day, typing these annoying squiggly characters. But with reCAPTCHA, von Ahn has come up with an idea for harnessing all that human brain power. He knew that lots of libraries have huge efforts under

way to digitize their collections. These projects first scan books or newspapers by basically taking a picture of each page. Then a computer takes the image of each word and converts it into text, using optical character-recognition software. But computers often come across printed words they just can't recognize. Especially for older documents, things that were written before 1900, where the ink has faded and the pages have yellowed out, the computer makes a lot of mistakes. A human being has to look at those words and decipher them. It occurred to von Ahn that he could link this kind of activity to security devices used on the Internet. Instead of asking people to prove they're human by copying random sequences of distorted letters and numbers, he could ask them to decipher mystery words from scanned books and newspapers. So he got together with *The New York Times,* which is digitizing newspapers going back to 1851, and nonprofits called the Internet Archive, which is digitizing thousands of books. In the journal *Science*, he and his colleagues report that over the last year Web users have transcribed enough text to fill up more than 17,600 books, with better than 99 percent accuracy.

D. Web T-shirt Company using crowdsourcing

The online company Threadless makes funny and satirical T-shirts. But they are a little different from those made by other T-shirt companies. All Threadless shirts are designed and approved by the site's members.

Each week, the Web site stages a design contest to determine the next T-shirts it will produce. The members vote, the winners get printed, and the T-shirts sell out, sometimes in just days. The site was started by two young college-dropouts with about $1,000 and a desire to build a Web-based community that would also be a business. Threadless emphasizes that everyone is welcome to submit to the T-shirt design contest, whether they're a trained artist or a novice doodler. There's a MySpace/Friendster element to the Web site. Each registered user has a profile, and can post pictures and start discussions with other members. The site gets about 150 design submissions a day. Entries have seven days to get scored by the community. At the end of the week, 10 of the highest scoring designs are selected and about 1,000 of each are printed. Anyone can submit a design idea that gets ranked by members of the crowd. If the design gets enough votes, it makes it into the store and the designer gets a big payout – $2000 in fact. In addition to the cash, the winning designer gets a $500 gift certificate for Threadless merchandise. For each subsequent printing, rewards the designer with another $500 in cash. Today, the company is introducing a crowdsourcing platform called Threadless Atrium,

which is initially targeted at causes and cause-based marketing campaigns.

E. Crowdsourcing in academic humanities domain

Four factors were identified that characterise, very broadly speaking, crowd-sourcing for the humanities. These characteristics are not exhaustive, but they usefully highlight commonalities between some of the activities we have observed:
a) The existence of a clearly-defined humanities direction or research question. The question could be designed by an academic team, or by an individual with particular knowledge or interests. This seems to preclude some categories identified by some authors [11] and will preclude elements of others [12]. It is suggested here that this characteristic is especially significant, since the *academic* component of academic humanities crowd-sourcing implies some form of professional rigour. However, we do not assume that the source of that rigour must necessarily originate from a Higher Education Institution.
b) The potential for a group with open membership to transform or add value to primary material or the interpretation of this material. However, a distinction has been made elsewhere between 'community sourcing' and crowd-sourcing, with the latter typically dealing in open calls for participation and the former being more closed. [10]
c) There needs to be a definable task, or some meaningful and replicable way of breaking the workflow down into sets of definable tasks.
d) The activity should be scalable, both to different volumes of data and different levels of participation.
 In the light of these characteristics, crowd-sourcing is considered to be distinct from the production of general user-generated content (UGC) on platforms such as Google Earth, as there is no clearly-defined direction or question, although such platforms could be components of crowd-sourcing projects if such a direction were present. Equally, the harvesting and analysis of so-called transactional data, that is information about people's (usually online) activities, is not considered here to constitute crowd-sourcing, as whatever additional value is added to the data does not result from public participation. There is an obvious distinction to be made between motivations for crowd-sourcing and motivations driven by market economics, which suggest that people will only contribute effort or submit to regulation in return for some benefit, usually material. Academic participants in projects in universities have relatively clear motivations, including, but not limited to, the fact they are materially rewarded by salaries and grants, professional recognition in their field, career advancement, and publication. Most crowd-sourcing projects however do not reward their contributors in material or professional ways, and members of

crowds who contribute to crowd-sourcing projects are not subject to discipline or sanction in the way that members of conventionally configured research organizations are. That said, contributors may receive "social" rewards, for example through rankings, increased standing in the crowd-sourcing community, or being credited and named in publications. Similarly, contributors may be subjected to social sanctions, such as banning, which can adversely affect their reputation and enjoyment, and may even in rare cases reflect on their professional standing. However, it is clear that the motivations of academic crowd-sourcing participants are more intrinsic to the activity.

F. Software crowdsourcing

The way of organizing software development in all these practices of software crowdsourcing is changing from traditional software factory or open source teams to decentralized, peer-production based ecosystems of software developers[8].Different from open source software, the openness of software crowdsourcing does not refer to free access to source code. Instead, it denotes an OPEN call for participation in any tasks of software development, including documentation, design, coding and testing. These tasks are normally conducted by either internal members within a software enterprise or people from contracting firms. But in software crowdsourcing, all the tasks can be assigned to anyone from the general public. Moreover, software crowdsourcing introduces more explicit incentives such as financial rewards to motivate community developers than open source software. Thus, it greatly extends the concept of open source community into the notion of market-driven software ecosystem.

Software crowdsourcing platforms including Apple's App Store, TopCoder [6], and uTest [7] demonstrate the advantage of crowdsourcing in terms of software ecosystem expansion and product quality improvement. Apple's App Store is an online IOS application market, where developers can directly deliver their creative designs and products to Smartphone customers. These developers are motivated to contribute innovative designs for both reputation and payment by the micro-payment mechanism of the App Store. Within less than four years, Apple's App Store has become a huge mobile application ecosystem with 150 000 active publishers, and generated over 700 000 IOS applications. Around the App Store, there are many community-based, collaborative platforms for the smart-phone applications incubators. For example, AppStori introduces a crowd funding approach to build an online community for developing promising ideas about new iPhone applications. Another crowdsourcing example–TopCoder, creates a software contest model where programming tasks are posted as contests and the developer of the best solution wins the top prize. Following this model, TopCoder has established an online platform to support its ecosystem and gathered a virtual global workforce with more than 250 000 registered members and nearly 50 000 active participants. All these TopCoder members compete against each other in software development tasks such as requirement analysis, algorithm design, coding, and testing.

G. Crowdsourcing in remote sensing

Global land cover is one of the essential terrestrial baseline datasets available for ecosystem modeling, however uncertainty remains an issue. We use tool like google earth for land cover validation. It gives fine spatial resolution image upto 50cm X 50cm. Every internet user from any region can involve in global validation exercise. The Geo-Wiki Project is a global network of volunteers' who wish to help improve the quality of global land cover maps. Since large differences occur between existing global land cover maps, current ecosystem and land-use science lacks crucial accurate data e.g., to determine the potential of additional agricultural land available to grow crops, volunteers are asked to review hotspot maps of global land cover disagreement and determine, based on what they actually see in Google Earth and their local knowledge, if the land cover maps are correct or incorrect. Their input is recorded in a database, along with uploaded photos, to be used in the future for the creation of a new and improved hybrid global land cover map.

H. Crowdsourcing and Hollywood

Besides being one of the earliest companies to introduce a crowdsourcing service in the form of the Mechanical Turk platform, Amazon is now harnessing the valuable data generated by its enormous consumer base to crowdsource the creation of original content. Amazon.com is producing its own movies and TV programming using the consumer tracking and data crunching skills it developed while becoming the world's largest Internet retailer. Since late 2010, the company's Hollywood studio, Amazon Studios, has let aspiring screenwriters and film makers upload thousands of scripts to its website. Amazon has been uploading scripts for films and TV series for crowd-sourced reviews, then producing original movie and TV content based on the consumers' feedback. For instance, Amazon took its nine best test movies from 2011 and posted them on Amazon Instant Video, the company's streaming video service. Customers viewed the projects hundreds of thousands of times, according to the company. It is using reviews and

feedback to re-write scripts .Amazon also collected data on how long customers watched the test videos and how many watched all the way through. So far, the company has produced movies such as "Zombies versus Gladiators" and the children's TV series "Magic Monkey Billionaire." In April of 2013, Amazon has announced the launch 14 new TV pilots available for streaming.

IV. Case Studies of Crowdsourcing

A. Google Map maker

Perhaps one of the first companies to realize the potential of crowd generated data, Google uses crowdsourcing for everything from optimizing search results, to translation. Earlier this year,Google announced that it plans to also implement a crowdsourcing element into its Google Maps platform. The crowdsourced version of the popular map application will examine every possible route from the chosen destination to the target point, then cross-reference these routes against the driving routes picked by people who have taken the trip in the past to come up with the best route. The map will show the driver landmarks to help them navigate, however the rest of the map will the be excluded from view to avoid clutter.

Walmart

You may think that this offline retail giant would have a hard time tapping into the mainly online trend of crowdsourcing, however Walmart has its own ideas on this subject. The company plans to introduce a crowdsourced delivery service to its stores to help grow its online shopping division. The service will rely on offline Walmart customers volunteering to disclose their home addresses and signing up to deliver packages to online shoppers who live along their route home. The success of this model could truly revolutionize the way companies think about delivery logistics, so keep an eye on future developments.

B. ChipIn

Unlike its hipper and edgier competitors, Microsoft is a latecomer to the crowdsourcing arena. However, in June of this year, the company has launched its first crowdsourced service, named ChipIn. The idea behind ChipIn is to find a crowdsourced way for high school and college students to pay for their first computer. After choosing one of 15 Windows devices, including tablets, hybrids, and laptops, ChipIn prompts the user to sign in to Facebook, and fill out a full profile with their personal details. The student's Facebook friends will then receive an alert asking them to ChipIn for his/her computer, and the

progress of the funding campaign can be monitored using the ChipIn dashboard.

C. I-Prize program

Like several other companies, Cisco has implemented a crowdsourcing element into its research and development process, in the form of the I-Prize program. I-Prize is a global competition that encourages entrepreneurs from all over the world to submit innovative business or technological ideas for Cisco to fund and help develop. The I-Prize program is currently in its 5th year, with previous winners coming from Germany, Russia, and Mexico among others. The latest Cisco I-Prize has been awarded to a Russian foundation for encouraging entrepreneurship.

D. Netflix

Netflix has experience with crowdsourcing, and this collaborative problem-solving model has impacted the company's success. From its first crowdsourced problem in 1996 in the form of a contest with a $1 million prize, Netflix has employed several crowdsourcing solutions consistently. It is also currently experimenting with using the crowd to caption its video library.

The successful crowdsourcing efforts at Netflix highlight the company's strengths that contribute to the effectiveness of the methods it uses. Among its strengths are rigorous testing and innovation. Perhaps Netflix's biggest strength, however, is that it understands what users want and understands that using the collaborative power of the human mind is the best way to provide it.

Netflix's first attempt at crowdsourcing was in 2006 when it asked the general crowd to find a better way to predict user viewing recommendations by improving on its prediction algorithm. This initiative shows how Netflix gets to the heart of the matter to determine what users want in order to provide it. In fact, Netflix also employs more than 40 people in-house as what it calls "taggers" to tag videos with at least 100 tags to work in conjunction with its recommendation function – showing how important the human component is in a computerized function.

Testing with crowds

Netflix takes its use of crowdsourcing and uses it to test, exhaustively, two different scenarios using 10,000 or more of its users. If something is successful during the crowdsourced testing, it is made available to all Netflix subscribers.

Captioning with the crowd

As its most recent crowdsourcing initiative, Netflix is utilizing the crowd for captioning. With more than 50,000 videos in its online streaming collection, Netflix is first starting a test phase of the crowd captioned content. Initial indications are that the program is successful. These online streaming videos

are likely to be watched and captioned by crowd laborers around the globe. While the participants in this effort are volunteer, they demonstrate just how powerful a crowd of collaborators can be for certain projects.

E. SellaBand

German startup Sellaband.com is hoping to leverage the wisdom, and cash, of the crowd to produce high quality independent music for free download on their site. It's a fascinating prospect even if it seems unlikely to succeed.

The way it works is this: bands upload sample music to Sellaband.com, promote the heck out of their profile page and ask fans to chip in $10 per share of a recording that will be produced when the band raises $50,000. The fans can take their money back out at any time before the goal is met. Once recordings are made, they are offered for free on the Sellaband site, where ad revenue will be split between the bands (60%), Sellaband (30%) and the hired producer and manager. Fans each get a copy of the recorded CD and bands are free to offer them any other benefits, like concert tickets, that they wish. Sellaband retains rights on the music for 12 months. The company seems confident that bands will be able to find 5,000 supporters (called "Believers") willing to put up $10 apiece.

One week since signing on, most of the 130 bands on the site have raised between $200 and $500. One Goth band from the Netherlands has raised $4500. Sellaband says that the $50,000 goal is what's needed to provide the kind of quality equipment and support that a major record label could provide. The company takes none of this money, only 30% of the site's ad revenue.

F.   The Scottish Words and Place-names (SWAP) Project

The Scottish Words and Place-names project was a JISC-funded activity to gather words in Scots and, ultimately, to offer selected words for inclusion in the word collection of the Scottish Language Dictionaries resource. Scots is not a distinct language in the way that, for example, Gaelic is, but it has a community of people who believe strongly in its cultural and linguistic importance. The aim of the SWAP project was to see which Scots words are in current use, and where/how they are used. The interface was relatively straightforward: a user could enter a word into a box on the website, and this was then harvested into a backend database. There was no moderation or checking done at this point, but words that the project felt were suitable were passed to the DSL (Dictionary of the Scots Language), where they were scrutinised by expert lexicographers. Words

were also gathered via Facebook and Twitter. The Facebook page was an important venue for developing conversations around the forms and meanings of individual words. Contributions via Facebook were less formal and structured, taking the form of discussions with project staff in comment threads. This required more of an overhead in the time spent processing and extracting them. However, it generated valuable material, and allowed the project to target specific areas and questions that were of interest, such as local dialects and rhyming slang. For example: "Thanks for all your place-name, alcohol and miscellaneous Scots info so far! This week we'd like to know about your local dialect rhyming slang. Some examples we already know about are: mammy mine - wine (Glasgow), Mick Jagger - lager (Gl), Lee Van (Cleef) - deif (Edinburgh), broon breid - deid (Ed & Gl), Mars bar - scar (Gl), Oscar (Slater) - later (Ed). Do you use these? Do you know any others? Please let us know!"[13].This generated responses such as: "I remember Caroline MacAfee's book on Glasgow: Language Varieties around the World collected some of these: askits (= shoulders) askit pooders, shooders); benny (lynch, cinch), etc. I never heard them used, though my godmother in Ayur [*sic*] still addresses me as 'china'."This highlights the importance of active engagement with interest communities for gathering contemporary linguistic content. In general, the project found that they received more useful information when they asked about specific words rather than framing open questions. The project was extremely successful in building up a community of followers and contributors among schoolchildren and teachers. This is largely because it was able to utilise and capitalise on the GLOW intranet network which connects Scottish schools. There was a slight potential ethical tension between the success of using Facebook to generate discussions, and the success of generating networks using GLOW: one is not allowed to have a Facebook account if under the age of 13. In June 2012 the project ran a competition for schoolchildren, with judges including Louise Welsh. This was again a highly successful way of engaging the community, and furthering the project's aims of both capturing and encouraging the use of Scots.

G. My Starbucks ideas

Using crowdsourcing for new idea generation. One the most challenging tasks for companies is to innovative with the goal of offering products /services to consumers.via   crowdsourcing platform,My starbucks Ideas, the company is able to have a hub where consumer can share their ideas regarding anything that is linked to the brand, specifically : products, in-store experience ,

involvement (i.e social responsibility, building community, etc). to sustain level of engagement via the platform, starbucks rewards the community members for their idea with points that they are accumulated and listed on the leader board.The most impressive section of the site is "Idea in Action". This section provide a listing of ideas that community members submitted and their status in regards of being implemented by the coffee house chain.

H. InnoCentive

InnoCentive is the open innovation and crowdsourcing pioneer that enables organizations to solve their key problems by connecting them to diverse sources of innovation including employees, customers, partners, and the world's largest problem solving marketplace. InnoCentive's proven Challenge Driven Innovation methodology, community of millions of problem Solvers, and cloud-based technology platform combine to fundamentally transform the economics of innovation and R&D through rapid solution delivery and the development of sustainable open innovation programs. Leading commercial, government, and nonprofit organizations such as Eli Lilly, Life Technologies, NASA, nature.com, Popular Science, Procter & Gamble, Roche, Rockefeller Foundation, and The Economist partner with InnoCentive to solve problems and innovate faster and more cost effectively than ever before. Since its inception in 2001, InnoCentive has established a formidable track record.
Total Registered Solvers: More than 250,000 from nearly 200 countries
Total Solver Reach: 12+ million through our strategic partners
Total Challenges Posted to InnoCentive.com: 1,300+
Project Rooms Opened to Date: 381,000+
Total Solution Submissions: 27,000+
Total Awards Given: 1,000+
Total Award Dollars Posted: $34+ million
Range of awards: $5,000 to $1 million based on the complexity of the problem
Average Success Rate: 50%+

I. P&G Connect & Develop

Procter & Gamble crowdsourced several well known brands through P&G's program "Connect + Develop". Launched in 2001, as a program in which at least half of the innovation comes from crowdsourcing external sources, Connect + Develop, is tracking to triple the impact of this already successful program through increased product innovation, development, and existing product reinvention by investing a 30% increase in marketing expenses. Connect + develop open innovation strategy has established more than one thousand active agreement with innovation partners. Connect + develop enables P&G to share R&D, commercialization and brand strength with partners worldwide, bringing great ideas to market and into the lives of consumers faster. Connect + develop are aggressively looking for solutions for their needs, they will consider any innovation — packaging, design, marketing models, research methods, engineering, technology, etc. that would improve their products and services and the lives of the world's consumers. Connect + develop has been at the forefront of open innovation with it's partners, suppliers and customers in the belief that "together, more value will be created than P&G could ever create alone".

J. InTrade

Intrade is an exchange that facilitates the matching of trade orders from its customers. Intrade ensures that trading profits and losses are transferred between customers in a timely manner and allows customers to close out positions by trading with any other customer. Intrade does not enter into trades on the exchange. Intrade members trade directly with each other. Their trading service allows members to transact in the most innovative, transparent and exciting way on political, financial, current and similar event futures. The Intrade markets offer a way to track the market-based consensus on political and current events. Participants stake real money in forecasting predictions, arguably driving more reliable results and more accurate that methods involving snap judgments in opinion polls or soliciting views of pundits where no-stakes are at risk.

VI. Future work
Crowdsourcing can be used for natural language processing in which we can collect data for processing from the crowd. We can also check the validity of data. We can see which words of a language are in current use, and where/how they are used.

VII. Conclusion
Crowdsourcing is a process of obtaining needed services, ideas or contents from large group of people. It is a way by which a information is obtained from people especially online community. In this paper we have shown how crowdsourcing can be employed in different fields. It shows how different organisations are engaging people for better output. In this paper we have given different applications of crowdsourcing.In this paper we have shown different case studies. We also have shown that how crowdsourcing can be used for natural language processing which will be our future work.

References

[1] J. Howe "The Rise of Crowdsourcing",Wired, 2006.

[2] Brabham,Daren, "Crowdsourcing as a Model for Problem Solving: An Introduction and Cases", *Convergence: The International Journal of Research into New Media Technologies* **14** (1): 75–90,2008.

[3] http://www.wired.co.uk/news/archive/2011-01/13/the-oxford-english-wiktionary.

[4] Special issue on structured and crowd-sourced data on theWeb Marco Brambilla · Stefano Ceri · Alon Halevy, The VLDB Journal (2013) 22:587–588

[5] Doan A, Ramakrishnan R, Halevy A Y. Crowdsourcing systems on theWorld-WideWeb. Communications of the ACM, 2011, 54(4): 86–96

[6] Lakhani K, Garvin D, Lonstein E. Topcoder (a): developing software through crowdsourcing. Harvard Business School General Management Unit Case, 2010. Available at SSRN: http://ssrn.com/abstract=2002884

[7] uTest. https://www.utest.com/

[8] WenjunWU 1, Wei-Tek TSAI2,3, Wei LI1,"An evaluation framework for software crowdsourcing" , springer 2013

[9] D. Schall ,B. Satzger , H. Psaier, "Crowdsourcing tasks to social networks in BPEL4People" , springer , 2011

[10] http://millennialmarketing.com/2010/09/crowd sourcing-vs-community-sourcing

[11] Wiggins and Crowston ,"*Action*" category of 2011.

[12] Bonney et. al. ,"Co-created" projects of 2009.

[13] http://www.facebook.com/scotswap/posts/230 076623689593