Mechanical Turk

WILLIAM LUNDIN FORSSÉN
and ADAM RENBERG

KTH Computer Science
and Communication

Bachelor of Science Thesis
Stockholm, Sweden 2010

Mechanical Turk

W I L L I A M L U N D I N F O R S S É N
a n d A D A M R E N B E R G

Bachelor's Thesis in Computer Science (15 ECTS credits)
at the School of Computer Science and Engineering
Royal Institute of Technology year 2010
Supervisor at CSC was Mads Dam
Examiner was Mads Dam

URL: www.csc.kth.se/utbildning/kandidatexjobb/datateknik/2010/
lundin_forssen_william_OCH_renberg_adam_K10066.pdf

Royal Institute of Technology
School of Computer Science and Communication

KTH CSC
100 44 Stockholm

URL: www.kth.se/csc

Abstract

Amazon Mechanical Turk (mTurk) is an infrastructure where tasks can be published for a crowd to solve. Those publishing the tasks reward the crowd for solving them. The system has increased in popularity since its release in 2005 but is still in a beta stage. This paper surveys the infrastructure, suggests improvements and debates the social and ethical questions that has arisen, such as "is this a virtual sweatshop?". We investigate the similarities and differences between this presumed virtual sweatshop and real world ones, and find that they closely resemble each other in wages and the menial nature of the tasks performed, but that they differ in working conditions and legal complexities.

Referat

Mechanical Turk

Amazon Mechanical Turk (mTurk) är en infrastruktur där uppgifter kan publiceras för en folkmassa att lösa. De som publicerar uppgifterna belönar folkmassan för att lösa dem. Systemet har ökat i popularitet sedan dess uppstart 2005, men är fortfarande i ett beta-stadium. Denna uppsats undersöker dess infrastruktur, föreslår förbättringar och debatterar de sociala och etiska frågor som uppstått, såsom "Är detta en virtuell sweatshop?". Vi undersöker de likheter och skillnader mellan denna påstådda virtuella sweatshop och fysiska sådana, och upptäcker att de liknar varandra i löner och den simpla strukturen av uppgifterna som utförs, men att de skiljer sig i fråga om arbetsvillkor och rättsliga komplikationer.

Contents

Chapter 1

Introduction

The purpose of this paper is to give a general overview of Amazon Mechanical Turk's website and work-flow, and the social/ethical questions that this type of workplace raises. When the number of available workers gets high and the workload gets low, what are the consequences? Will this generate a workplace with similar conditions to a sweatshop? Or does it provide the necessary framework for people to be able to make ends meet? Our main task is to put ourselves in a workers position and experience the life of a 'turker'.

As our experience builds up we will suggest improvements and investigate who the actual turkers are. Perhaps surprisingly, a large majority of them have an education equivalent to a college degree, or higher [1]. This, combined with the fact that most of them are Americans, might give those believing mTurk to be a sweatshop pause. But with the latest shift towards more Indian workers [10], and the significant 18% who are somewhat or fully depending on their pay from mTurk to subsist [1] gives reason to investigate further.

Amazon Mechanical Turk is an example of a greater movement in community work combined with Internet called crowdsourcing, and we will discuss some other examples of this.

Chapter 2

Background

2.1 Crowdsourcing

Crowdsourcing is the act of outsourcing a job to a group of people or community by publicly publishing the job, asking for contributions. The community submits solutions, and vets them for accuracy. The outsourcer/employer then chooses a desired solution, and hands out rewards accordingly [10].

Crowdsourcing can be synthesized into the following steps [10]:

1. A company has a problem.
2. The company broadcasts/publishes the problem (often online, but also through other media).
3. The target group of people/community ("crowd") is asked for solutions.
4. The crowd submits solutions.
5. The crowd verifies the accuracy of the solutions.
6. The company rewards the "winning" solutions.
7. The company now owns the solutions.
8. The company profits.

Crowdsourcing commonly takes the shape of competitions: Design a new slogan for a product and win a prize. For instance, Pepsi launched a campaign in early 2007 asking consumers to design a new look for the Pepsi can [13]. There was a monetary reward of $10'000, and the winners got their artwork featured on 500 million Pepsi cans. This way Pepsi can market their product, and creates a "crowd community" that does work for them and generates buzz around their products and the company, leading to a kind of viral marketing.

Another example of crowdsourcing, that follows the eight steps above very closely, is Wikipedia. The problem is: Write encyclopedic articles about everything. The crowd is the community of editors (anyone with a computer and an Internet connection), and they submit articles. These are vetted by their peers (and superiors in the form of administrators etc., but that is irrelevant), and often hotly debated. Wikipedia now owns the articles, and thrives as the largest encyclopedia online [11]. (It doesn't make a profit in a monetary sense, as it is a non-profit organization, but succeeds in other ways, such as providing a quality free service).

Wikipedia gives no extrinsic rewards but the accreditation of having done it. The concept of asking people to work for free has been given an interesting twist in the form of the ESP Game [12][19]. On the surface, the ESP Game is an anonymous two-player online game where both players are given an image to describe. The game marks certain words as "taboo" which the players cannot use to describe the image. If the players manage to describe the image using the same word (also called tag or label), they get points. This is a fun game, with many people playing several hours a week. The genius is that the data generated this way - the tags to describe the images - is actually very accurate and useful, e.g. in-image search (Google has as of 2006 bought a license to make their own version of the ESP Game to improve their search results [20]).

The ESP Game is also interesting in the way it uses human computation to solve problems that computers cannot easily solve. Amazon Mechanical Turk is a marketplace for such non-computer-computable tasks.

In the last decade or so, several crowdsourcing marketplaces have sprung up online [13], where people can publish jobs for anyone to do in exchange of money. One of them is Rent a Coder [21], which does exactly what the name describes. It is an auction house for programming tasks, where the lowest bidder with sufficient qualifications wins.

2.2 Amazon Mechanical Turk

Amazon Mechanical Turk takes its name from the 18th century chess-playing automaton. The machine toured Europe, claiming to be the world's first fully automated chess player. It was later revealed that there was a chess master hiding inside the machine, controlling its operations. Similarly, Amazon's Mechanical Turk is a machine that harnesses the work of hidden humans. It attempts to match people with small tasks that are currently unsuitable for machines.

Some example of tasks on mTurk are:

- Writing a review of a blog, website or product.
- Translating a text from one language to another in its natural spoken form.
- Labeling images or music clips.
- Writing a short plot description of a movie.

These tasks are called HITs (Human Intelligence Tasks). The ones who broadcast them online are referred to as 'requesters' and those who solve these HITs are 'workers' or 'turkers'. When a HIT has been completed it needs to be verified by whoever broadcast it and, if accepted, the worker receives payment (usually around a few cents). Sometimes a HIT comes with extra bonuses for those who do especially excellent work, or for the top turker of the week.

Amazon created this infrastructure to meet the demands of companies or people who had problems that they could not solve by themselves because of different reasons, time and money being the most common. An example is a man who needs to verify addresses for companies that his own company has done business with. The records he has are over 15 years old so he needs to get them up to date. At first he would attempt to update the records himself with the newly updated addresses of the companies. But after a while he finds out that it would take him several weeks for him to do it on his own. So he spends $300 and publishes the job on mTurk. Three days later the work was done and he even received e-mails from workers who were alerting him that some companies on his list now had two different addresses or didn't exist anymore.

mTurk was launched on November 2, 2005. It quickly attracted many users, partly due to the Slashdot effect. The current amount of registered users is more than 100,000 with about 55,000 HITs available at any time [9].

Chapter 3

Working as a turker

3.1 Hitting it off with mTurk

We tried several different HITs, such as "Tag and rate an audio clip", "Movie sentiment analysis" and "Search for bugs in Java code".

In the first, we listened to several 10 seconds long music clips with many different music styles, and then wrote down some words describing the music. The tricky part of this HIT was the list of prohibited tags, and the long list of instructions telling us what *not* to do. The second HIT gave us around fifteen twitter posts somewhat related to a specific movie (we got High School Musical and Avatar, for instance). The task was then to select whether these tweets spoke positively or negatively (or both) about the movie. The total work time was no more than two minutes, and we got $0.01 for our work.

The last HIT was larger and less menial in nature, which was represented in the reward of $4.50. The work consisted of looking at about 40 snippets of Java code, and tell which, if any, line had a bug in it. The code snippets was 100-200 lines of different algorithm and data structure implementations, such as quick sort and linked list. There were also some examples of Java API specific code, such as Swing apps. The HIT was part of a research project from an American University that studied how easy it is to spot bugs. It took about one hour to complete, and was actually quite fun and instructive. At the end there was one major problem though, one which we will dissect further in the next section: We did the HIT before we had formally accepted it, so when we were done someone else had gotten it accepted, and we couldn't get paid.

3.2 Improvements

Amazon has built its community on the basis of user reviews. However, this has not been carried over to mTurk since you are currently unable to review/rate a requester. This leads to fake HITs by non-serious requesters which in turn has lead to a separate part in the community forum Turker Nation, called Hall of Shame

[7]. The workers post which requesters are good and bad, this in turn creates a lot of overhead for serious turkers. The so called "Speed Turkers" who specialize in doing HITs quickly in order to maximize their profits are especially affected. Now everyone is required to check the seriousness of the requester before they start working which takes considerable time, time that could be better spent doing HITs. One solution to this problem is to make requesters possible to rate and have their rating show up whenever one of their HITs is shown.

A peculiarity we encountered is the capability to start doing work on a HIT before you've accepted it. This seems to be a beginner's problem, although it is not something that needs to be there. This could be solved by making the "Accept HIT" button (that appears with every HIT description) more eye catching, like a friendly reminder to push the button before you start working. Or perhaps you should be prohibited to start any actual work until you've formally accepted the HIT.

Another problem concerning HITs is the requesters option to timeout the HIT. If you get a timeout then the HIT is put back into the pile and you loose any money that would have been bestowed upon you should you have completed the HIT. This could be considered aggravating for workers when they have spent time making sure they do the HIT correctly. Some HITs have a lot of rules and CAPTCHA-like control questions which takes time to read and understand. This is a lose-lose situation for both the requester and the worker. The requester receives more HITs with questionable quality [8] and the worker looses the time spent doing the HIT and the money that came with it. This problem could also be blamed on the requester which has put the time for the HIT to short which in turn relates back to the fact that beginners may deliberately choose to start working on the HIT before accepting it just so they won't get a timeout.

There is also the case when the time limit is a necessity. Maybe the requester wants the HIT to be completed very quickly or it could be that the goal is to make users not think to much before they complete the HIT. An example would be a question like "How do you feel on a scale from 1 to 10?", if the question isn't answered intuitively and the user is given time to think it could ruin the point of the question and thus the HIT itself.

Since the launch of mTurk it has become clear that some tasks are better suited in this environment than others. The most common tasks could be put into a generalized structure template. For instance, all review based HITs could be put into a "review-template" while tagging pictures or music could be put into a "tag-template" and so on. This would make it easier to understand the HIT and also lead to quicker completion.

Below is the most common page seen as a turker (Figure 3.1). The website seems a bit old fashioned in regards to graphics and layout. It kind of reminds you of a non-professional HTML attempt, not something you would expect from a company as web-centric as Amazon.com. It is a constant reminder that the service is still in a beta stage [9].

Figure 3.1. HITs - page for displaying available HITs on Amazon Mechanical Turks website.

Chapter 4

Statistics and demographics

4.1 Facts

The spread of turkers around the world is actually not such a wide spread. In the middle 2009, approximately 57% of the users hailed from the US and 3% from Canada [1] suggests that the majority of the users are living in western 'modern' society. About 32% are from India which is considered a 'poor' country [3]. However, the reason that the huge part of the user base (89%) are American or Indian can be related to the fact that Amazon can only do bank transfers or give checks to users in those countries [4]. If you are from any other part of the world you would have to convert your turker credits (money made from mTurk) into gift certificates which can only be used to shop from Amazon.com.

Research done in early 2008 gives another picture. Then just over 76% of the turkers were American, and only 8% where from India [10]. This shift in demographics might mean that the general turker population is shifting towards the poorer India, but it might also be a sign of the inherent bias of the surveys used to determine turker origins. Using a survey to establish user origins and status has the unfortunate design flaw that it attracts "people-who-like-surveys" to answer it [1].

In general when speaking about computers and the Internet, it is common to associate this business with males. In mTurk there are 55% female users [1]. The explanation for this could be that females make up the majority of Internet consumers [5].

The education level among turkers is very high. 87% of the users have some kind of college experience and 42% even have a bachelor degree[1]. Even though the user base is quite educated, unemployment is high. A whole 31% of the users are unemployed [1].

The average turker spends about 8 hours per week performing HITs and earns around $10 during that time. Only 3% earns more than $50 per week [1].

It turns out that 50% of the users don't really care for the money they make on mTurk [1]. 18% sometimes or always need the money to make ends meet.

40% of the users are 18-24 years old which comes as no surprise as the payments are very small. Young people tend to accept working for less since they are inexperienced. 54% of the users make less than $30'000 a year [1].

4.2 Discussion - is this a sweatshop?

Is Amazon Mechanical Turk a 'virtual sweatshop'? To answer that question we need to define the word 'sweatshop'. Wiktionary defines sweatshop as "a factory or other place of work where pay is low and conditions are poor or even illegal" [16].

First, lets take a look at the first part of the definition: "where pay is low". The turkers make small amounts of money, so small that most of them don't even care if they get it or not. Keeping that in mind there are the 18% of the users who depend on this money to make ends meet which should be considered alarming. Those who depend on this money are most likely also unemployed and/or they might live in India, however, there are no statistics to cross-reference this assumption. If we were to judge only on the workers wage it is certainly comparable to real sweatshop wages [25]. The reasons for working in these conditions differ considerably. Where mTurk workers don't care about their pay, a third world country sweatshop worker certainly do, because they are dependent on the income for their survival. In some cases the alternatives to sweatshop work is subsistence farming, prostitution, trash picking, or starvation by unemployment [23] [24].

The second part of the definition states: "conditions are poor". The working conditions of a turker can not be considered to be poor when compared to a third world country sweatshop. For example, on mTurk users can work in the comfort of their own home or outside (with the help of a laptop), take as many breaks as needed, take time off whenever they please and everyone is their own boss. However that is not to say there aren't any physical conditions that can be related to working in front of the computer several hours every day. It can cause eyestrain, back and neck pain, headaches and so on [17]. But if we compare these inconveniences to the physically exhausting and often brutal real world sweatshop conditions, their severeness diminishes quite a bit [18].

The last part of the definition is: "the conditions are illegal". There are three parties to take into account when discussing the legality of this service, Amazon, the requester and the worker. Amazon does not commit a crime when providing the infrastructure the requesters and workers use when conducting business. The requester doesn't commit a crime but has quietly side-stepped the laws of minimum wage, overtime and workers compensation due to the fact that workers are payed as contractors and not as employees. The worker, however, is left with the burden of reporting their income as self-employment income. So conditions are not illegal by the rule of law but could be considered rather immoral and/or unethical [9].

There is also a controversy of whether sweatshops are good or bad [14]. Some neo-liberal economists, Paul Krugman of Princeton University among them, claim that they are a natural step in the maturing process of developing countries, and

that all currently developed countries used them in their youth [22]. People would not choose to work there if they had a better option, and in many cases sweatshops offer higher salaries than the national average [6]. However, earning more than the average does not tell you everything. The average person may still be very poor and the income might still not be sufficient to accommodate a decent standard of living. Averages should be treated with a bit of skepticism since they may not reflect reality because the 'average' person may not exist. For example, in an economy with no middle class, there is only the rich and the poor, but the average person would have had the appearance of a 'decent' lifestyle.

Sweatshop critics point out that while companies may pay more than what the average worker in that country earns, it is not the whole truth. That statement refers to the fact that the foreign companies sometimes use domestic subcontractors, who actually produce the goods. They in turn have their own profit interest, and thus the surplus is taken by that subcontractor's managers and the workers may never see a penny of it [25]. There have also been reports of abusive behavior by some of the domestic subcontractors but since these abuses are against local law it is not something that the foreign company can be held responsible for - after all, that is the responsibility of the local police [18].

Proponents of sweatshops claim that they are not bad for a country or its people per definition. To generalize them all would be to neglect the fact that most poor countries are better off with sweatshops than without because of the substantial positive wage difference it has when compared to the country's general income per capita [25].

Chapter 5

Conclusion

Sweatshops tend to exist in places where there is the possibility of cheap labour. Like the real world kind, mTurk has also become a place for cheap labour but not necessarily a sweatshop, using the definition of [14]. Since the average turker don't even care about the money they make it is difficult to say that they are working in a sweatshop - they are not really working in poor or illegal conditions. The issue of the workers pay is a question of supply and demand. There are currently to many people willing to work and too little work to be done to satisfy any demand for higher pay. This has given the requesters a very advantageous position.

Those workers who are using mTurk as a necessity to make ends meet, are in a precarious position. However, to even be able to work, you need a computer with an Internet connection, costing an average of $46 per month in the US [15], not to mention the cost of the computer itself, which implies a "decent" standard of living.

All this talk about working conditions and sweatshops paints a rather bleak picture. But there are stories of turkers who manage to do quite well on mTurk, who even earn decent monthly wages [2]. Specializing on certain types of HITs or trying for those desirable extra bonuses seems to be viable strategies for these crafty few.

If handled correctly, Amazon Mechanical Turk could become an important job market on an international scale. Combining a very free working environment and market economy wage negotiations with a seemingly unlimited supply of workers, it is in theory a win-win situation. Crowdsourcing at its finest, most capitalistic, moment.

Bibliography

[1] Joel Ross, Andrew Zaldivar, Lilly Irani, Bill Tomlinson, *Who are the Turkers? Worker Demographics in Amazon Mechanical Turk*. 2009.
http://www.ics.uci.edu/~jwross/pubs/SocialCode-2009-01.pdf

[2] Katharine Mieszkowski, *"I make $1.45 a week and I love it"*. Salon.com, 2006.
http://www.salon.com/tech/feature/2006/07/24/turks/index.html

[3] The World Factbook, CIA, 2004.
https://www.cia.gov/library/publications/the-world-factbook/rankorder/2004rank.html

[4] Worker Web Site FAQs, Amazon, 2010.
https://www.mturk.com/mturk/help?helpPage=worker

[5] Pia Sarkar, *GENDER GAP / Computer industry skewed toward male perspective despite users demographics*. San Francisco Chronicle, 2002
http://articles.sfgate.com/2002-01-30/business/17527568_1_computer-science-jane-margolis-science-students

[6] Johan Norberg, *The Noble Feat of Nike*. 2003.
http://www.johannorberg.net/?page=articles&articleid=53

[7] *Hall of Shame / Fame*. Turker Nation, 2010.
http://turkers.proboards.com/index.cgi?board=shame

[8] Aniket Kittur, Ed H. Chi, Bongwon Suh, *Crowdsourcing for Usability: Using Micro-Task Markets for Rapid, Remote, and Low-Cost User Measurements*.
http://www.parc.com/content/attachments/crowdsourcing_user_studies_6280_parc.pdf3

[9] *Amazon Mechanical Turk*. Wikipedia, 2010.
http://en.wikipedia.org/wiki/Amazon_Mechanical_Turk

[10] Panos Ipeirotis, *Mechanical Turk: The Demographics*. 2008.
http://behind-the-enemy-lines.blogspot.com/2008/03/mechanical-turk-demographics.html

[11] *Traffic stats for wikipedia.org*. Alexa.com, 2010.
http://www.alexa.com/siteinfo/wikipedia.org

[12] Luis von Ahn, *The ESP Game*. 2006.
http://video.google.com/videoplay?docid=-8246463980976635143

[13] *List of Crowdsourcing projects*. Wikipedia, 2010.
http://en.wikipedia.org/wiki/List_of_crowdsourcing_projects

[14] *Sweatshop*. Wikipedia, 2010.
http://en.wikipedia.org/wiki/Sweatshop

[15] *Average broadband monthly subscription price, by country, USD PPP*. OECD, Oct 2008.
http://www.oecd.org/dataoecd/22/44/39575002.xls

[16] *Wiktionary - sweatshop*. Wiktionary, 2010.
http://en.wiktionary.org/wiki/sweatshop

[17] *Computer-related injuries*. Better Health Channel, January 2009.
http://www.betterhealth.vic.gov.au/bhcv2/bhcarticles.nsf/pages/Computer_related_injuries

[18] Dexter Roberts, Pete Engardio, Aaron Bernstein, Stanley Holmes, Xiang Ji, *Secrets, Lies, And Sweatshops*. Nov 27 2006.
http://www.businessweek.com/magazine/content/06_48/b4011001.htm

[19] *The ESP Game*. 2010.
http://www.gwap.com/gwap/

[20] *Google Image Labeler*. 2010.
http://images.google.com/imagelabeler/

[21] *Rent a Coder*. 2010.
http://www.rentacoder.com/

[22] Paul Krugman, *In Praise of Cheap Labor*. MIT.
http://web.mit.edu/krugman/www/smokey.html

[23] Allen R. Myerson, *In Principle, a Case For More 'Sweatshops'*. The New York Times, June 22 1997.
http://www.nytimes.com/1997/06/22/weekinreview/in-principle-a-case-for-more-sweatshops.html

[24] Nicholas D. Kristof, *Inviting All Democrats*. The New York Times, January 14 2004.
http://www.nytimes.com/2004/01/14/opinion/inviting-all-democrats.html

[25] Benjamin Powell, David B. Skarbek, *Sweatshops and Third World Living Standards: Are the Jobs Worth the Sweat?*. The Indepentent Institute, September 27 2004.
http://www.independent.org/publications/working_papers/article.asp?id=1369