

# **An Innovative Crowdsourcing Approach for Amazon Mechanical Turk**

Hanieh Javadi Khasraghi  
MSc Student in Information Technology  
K. N. Toosi University of Technology

Shahriar Mohammadi  
Assistant Professor  
K. N. Toosi University of Technology

## **ABSTRACT**

Web2 and the evolving vision of Web3 have a great effect on facilitation of information sharing, information aggregation, interoperability, user-centered design, collaboration on the World Wide Web, and crowd-centered services. New concept of Web is the intuition that drives crowdsourcing, crowd servicing, and crowd computing. With crowdsourcing emergence people get motivated to work through internet without being limited by time or geographical location. On the other hand employers could have their jobs done faster and cheaper.

This paper is going to introduce an innovative approach for Amazon Mechanical Turk (AMT) crowdsourcing marketplace. In current AMT marketplace, workers especially new ones need to qualify themselves for each requester that has submitted Human Intelligence Tasks (HITs) in AMT, and there is lack of shared reputation system; some workers may cheat on tasks in order to maximize their income, as a result requesters are uncertain of the quality of results, so they offer lower rewards and consequently qualified workers leave the marketplace.

Because of the above shortcomings, we introduce a new approach for AMT crowdsourcing marketplace. In our proposed approach we offer to distribute HITs among Amazon's customers and ask them to work on tasks in exchange for discount. The distribution of HITs is based on customers' interests and skills that Amazon has this information in its database. Through our proposed approach the HITs will be done by more qualified people, and spammers will be decreased to the minimum. This innovative approach is very efficient, time saving, and user friendly (because workers don't need to search for HITs of their interests).

## **General Terms**

Human Computer Interaction, Crowdsourcing.

## **Keywords**

Crowdsourcing, Amazon Mechanical Turk (AMT), Human Intelligence Tasks (HITs), Classifying, Distributing.

## **1. INTRODUCTION**

Crowdsourcing has provided a chance for organizations to have their tasks done by large groups of individuals with different talents and skills. Employers rely on intelligence of crowd of people instead of having an employee or a machine to do the tasks. Some tasks such as "image labeling", "transcribing old books and newspapers", and "describing some patterns in music" need human intelligence to be done with quality, and machines are unable to do these kinds of tasks with acceptable quality. Crowdsourcing has a wide range of application and enables many enterprises to access scalable networks of people on-line.

Wikipedia uses crowdsourcing for information aggregation, Amazon Mechanical Turk uses for doing human intelligence tasks, Google uses for image labeling, ReCAPTCHA uses for digitizing old books and newspapers, Crowd Spring uses for Logo and marketing design, Children's Digital Library uses for translation and so on. It's obvious that crowdsourcing is going to be the future of businesses; and organizations will build their business model based on crowdsourcing.

In every project or job the time is of great importance. If a project or job doesn't finish in the predefined deadline, it could be considered failure. In crowdsourcing process, if a requester doesn't be sure of task's completion time, she/he will not crowdsource it. To be sure of the completion time we have to motivate more people to do tasks, and have a right approach for distributing tasks among qualified and skilled people. Having jobs done by qualified workers is of great importance for employers too. There are many prior researches for ensuring quality of crowdsourcing results. Most of the works uses majority voting to improve overall quality of results [1], Le et al. applied training data approach to improve quality of workers' results [2], Kittur et al. offers to design tasks in such a way that completing them in good faith take the same or less effort than malicious completion [3]. There are various approaches in prior works for assuring quality of results. In most of the crowdsourcing efforts requesters should apply quality assurance approaches which necessitate constantly working on better approaches to ensure quality of results. Our work aims to improve quality by routing the tasks to qualified and interested customers of Amazon. Through our proposed approach we could have qualified workers to do more complex tasks with high quality and in less time. We'll focus on Amazon Mechanical Turk to introduce our new approach for crowdsourcing tasks.

In summary, in this paper we're interested in the question of how to reduce the cost of some potential workers finding tasks that match their interests and skills and how to help requesters to have their tasks be done with high quality and in less time.

This paper is divided into four parts. In section 2 we review the literature of crowdsourcing and Amazon Mechanical Turk marketplace. In section 3 we will introduce our innovative crowdsourcing approach. In section 4 of this paper we represent the results of the survey that we conducted to see if Amazon customers like to participate in crowdsourcing efforts or not. Section 5 describes our new approach's advantages in Amazon Mechanical Turk marketplace. In the last section we'll conclude from our research and we'll define some insights for future works.

## **2. RELATED BACKGROUND**

### **2.1. Crowdsourcing Literature**

The term crowdsourcing, first coined by Jeff Howe in an article in Wired magazine [4], his definition of crowdsourcing is:

*“Crowdsourcing is the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call.”*[4]

There are other definitions for crowdsourcing and all seems to be compatible with Howe’s:

*“Crowdsourcing as a concept as well as a practice refers to the idea that the web can facilitate the aggregation or selection of useful information from a potentially large number of people connected to the internet.”*[5]

*“Crowdsourcing is emerging as the new on-line distributed problem solving and production model in which networked people collaborate to complete a task.”*[6]

All of the above definitions refer to crowdsourcing as a new problem solving model that harness human intelligence to solve problems that computers or machines are unable to do.

There are three roles in crowdsourcing process: requester (also known as employer), crowdworkers (also known as workers or providers), and crowdsourcing platform. Crowdsourcing requester submits a task (or couple of tasks) request in crowdsourcing platform, and provides the tasks description, the required skills that each participant should have to be able to do tasks, due date for task completion, the time each worker could spend on the tasks, and useful information for workers to do the tasks with high quality.

Crowdworkers according to their talents and skills and the time they want to spend for each task, choose appropriate task to work on. The results of tasks are submitted by workers to crowdsourcing platform; if submitted results meet requesters’ criteria, workers will be rewarded. Crowdsourcing platform acts like an interface between requesters and workers ensuring the successful completion of tasks and payment process [6].

Currently, many organizations are turning to the crowd to complete various tasks of organizations. Amazon Mechanical Turk (AMT) is a widely used crowdsourcing marketplace that many organizations rely on it to crowdsource their tasks. In the following section AMT and its’ drawbacks will be discussed.

## 2.2. Amazon Mechanical Turk

Amazon Mechanical Turk (AMT) is a crowdsourcing marketplace for small tasks that employers (are known as “requesters”) post their tasks (which are called “HITs”) to be done by workers.

AMT uses monetary payment; it aims to facilitate access to scalable human intelligence. Ipeiritis conducted a survey in 2010, collecting demographics of 1000 Mechanical Turk users. The results indicated that 46.80% of users are from USA, 34% are from India, and 19.20% are Miscellaneous. Ipeiritis presented the demographic analysis separately for American and

Indian workers; the results show that the distribution of workers depends on gender, age, educational level, income level, marital status, and etc. In general, workers are generally young and have low income and small families [7]. Researchers discussed some drawbacks of current micro-task markets. Ambati et al. represent that the workflow in AMT is sub-optimal for both requesters and the workers, they discuss that requesters should constantly apply some quality assurance approaches to distinguish high quality results from low quality ones and also to identify workers’ accuracies; this approach is time-consuming for requesters . On the other hand, workers are unable to find the tasks that match their interests and finding appropriate tasks that align with their interests is time-consuming too[8].

By scrutinizing most of the papers about AMT we extracted the following as drawbacks of Amazon Mechanical Turk marketplace:

1. Workers of AMT are limited to some special countries, there are not enough qualified workers, and finding workers with certain capabilities are difficult.
2. AMT is a price-effective task completion marketplace [7]; workers are limited by the current user interface. Most of the workers choose HITs that offer more reward; so there is a high degree of unpredictability in completion time of some tasks.
3. Most of the tasks are done by low quality workers before high quality ones find the tasks that match their skills and interests. So high quality workers may choose to work on tasks that they are not interested in or skilled at, but just attractive for them due to the reward [8].
4. Requesters offer very low wages (because there is lack of shared reputation system and increasing wages will attract spammers); as a result qualified workers will leave the marketplace.
5. Workers can obtain a Qualification by browsing or searching through the available Qualifications and requesting ones that appeal to them, but it’s time consuming.
6. As there is not acceptable shared reputation system, for each HIT or HIT group, requesters may conduct a qualification test. Because new workers needs to qualify themselves to each requester, redundancy will increase in AMT, and probably new workers become tired of these qualification tests and leave the marketplace.

Figure 1 illustrates current AMT crowdsourcing process from qualification test and task distributing perspective, as you can see there is redundancy in this marketplace. We’re going to introduce a solution to the problems that mentioned about AMT, in the following section.

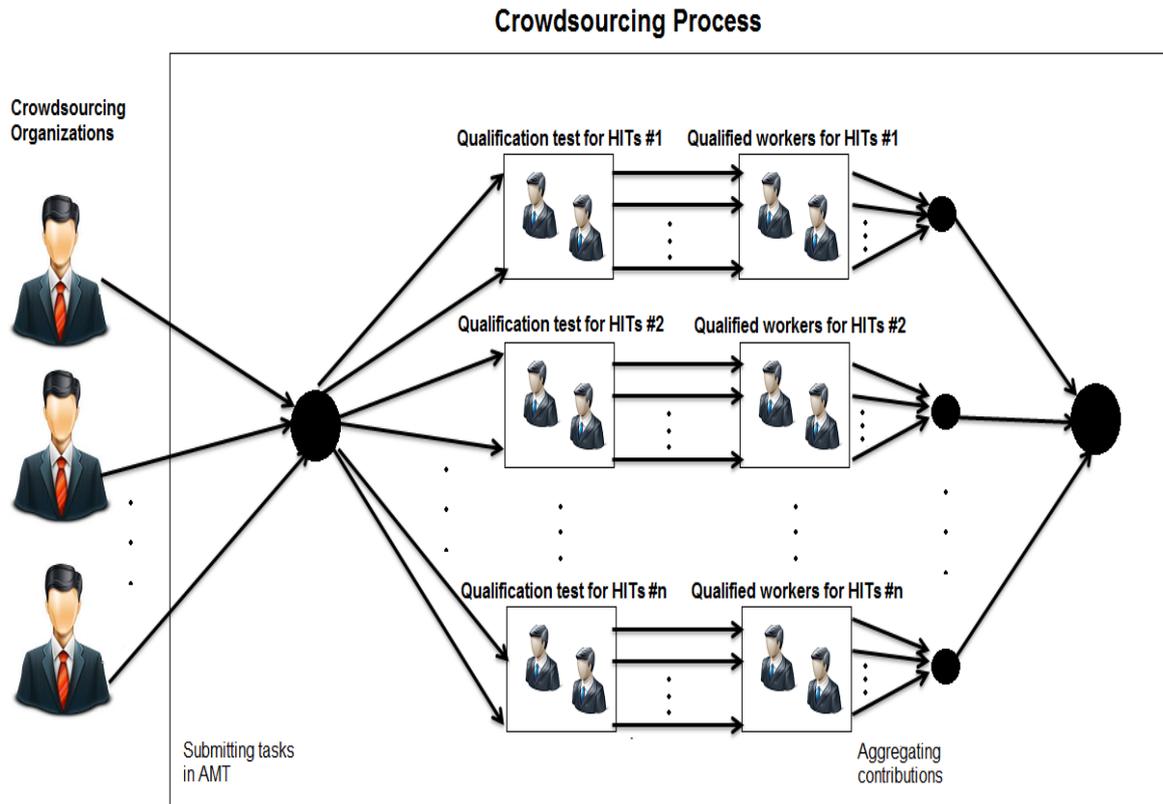


Fig 1: Crowdsourcing process in Amazon Mechanical Turk

### 3. AN INNOVATIVE APPROACH FOR AMT CROWDSOURCING MARKETPLACE

Amazon according to users' previous searches and their purchases from Amazon site collects some information about customer's interests. By data mining techniques, companies could recognize significant facts about customers' trends, characteristics, and purchase behaviors. We could use this information not only for personalizing and continually improving customers' shopping experience at Amazon, but also for AMT's crowdsourcing objectives, such as distributing AMT's HITs among Amazon's customers based on the type of HITs and customers' interests. For making the distribution process easier, Amazon could classify customers into groups according to their interests and skills, and distribute tasks based on type of tasks and characteristics of each group (figure 2). For leveraging this process companies could apply data mining techniques. Data mining could be used to classify customers into groups according to their interests, skills, trends, purchase behaviors, and categories of products that they have purchased; therefore we could predict which class of customers are the most appropriate one for each task that has been crowdsourced.

Through this approach tasks will be recommended to the right customers who may be interested in participating in crowdsourcing effort, and the outcome will be of high quality.

Suppose a requester has submitted HIT that is to describe patterns in music; people who have good taste of music would definitely do these kinds of tasks better than ordinary people; and most of the people who belongs to these groups usually purchase music tools, music books and so on. Via Amazon we could know which customers have good taste of music, so we could distribute HITs related to music among these customers. About other categories of tasks we apply the same approach. To achieve the desired outcome we should apply data mining techniques to classify workers and tasks into right categories with consideration of more various features of customers. After categorization of tasks and customers, we route tasks to the appropriate categories and recommend them to customers. If recommended tasks align with customers' interests, they may work on them. The extent to which customers may participate in crowdsourcing efforts will be evaluated by a sample task in the next section.

Figure 2 illustrates our proposed AMT crowdsourcing process from task distribution perspective. With comparing figure 1 and figure 2 it's obvious that in our proposed scheme there is no redundancy in selecting qualified workers. The process order in this approach is RCW (Requester, Computer, and Worker), because the computer plays an active role in doing tasks. When a requester submits HITs, computer searches the database of customers to find matches between tasks and customers' interests to offer right matches of HITs to customers.

## Crowdsourcing Process

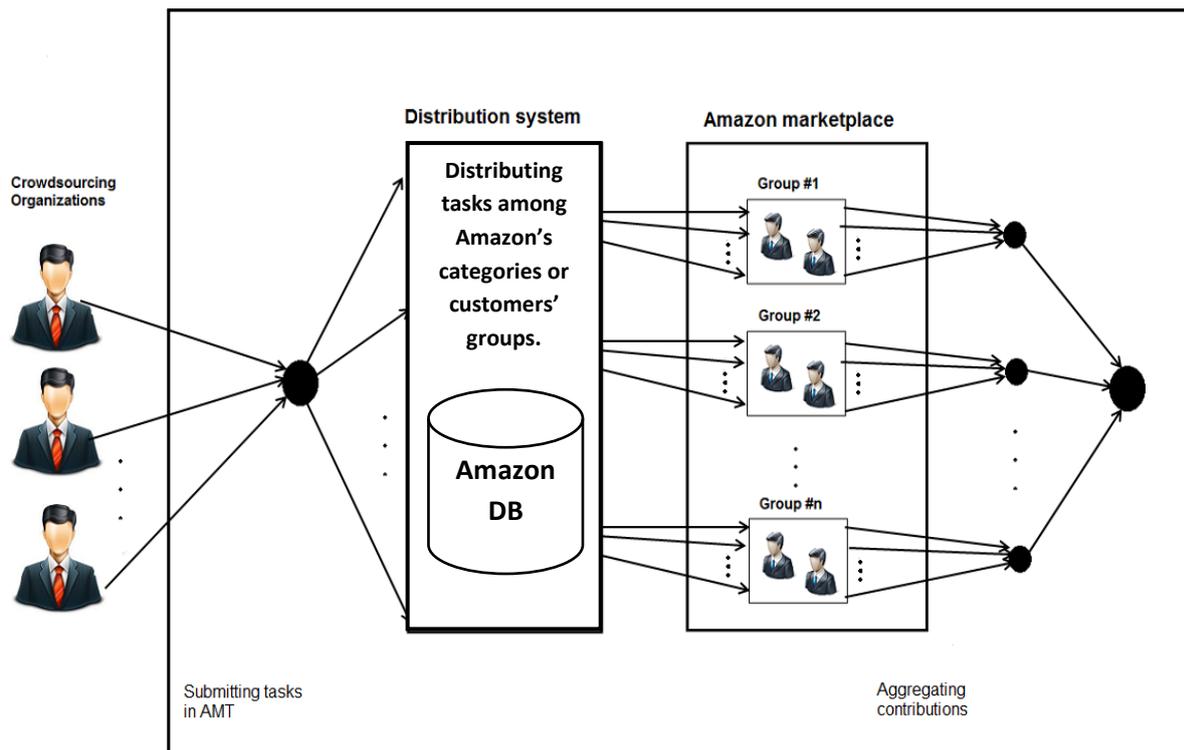


Fig 2: Crowdsourcing process in our proposed AMT marketplace

The question that may come to our mind is what motivates Amazon's customers to do HITs?

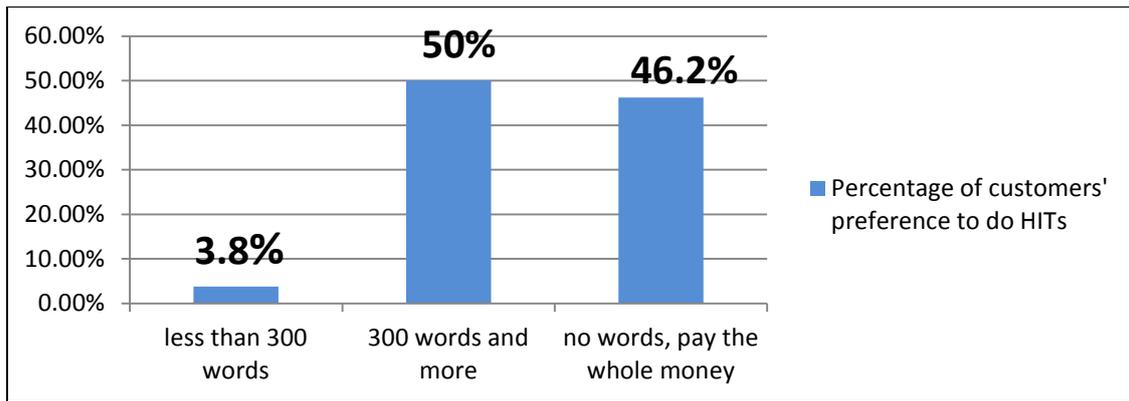
We could bring HITs into purchase transaction in Amazon market. A customer who wants to buy something (e.g. book) has two choices: the first one is the ordinary one in which the customer pay for book (one click shopping); the second one is crowdsourcing approach in which customer are asked to do some recommended tasks of her/his interests. The more tasks a customer does the better discount she/he will get. According to complexity of tasks we define appropriate discount. Through this approach requesters can reward the workers appropriately, and they don't have to deal with spammers.

This kind of crowdsourcing process has many advantages that we'll represent in the following sections. To be sure of the quality and correctness of results we could apply *output agreement* (also known as majority voting in some papers), which was introduced by Quinn and Bederson (2011) as a way for quality control, so Amazon distributes one task between couples of customers simultaneously and customers work independently and simultaneously in different locations and if there was a match between answers, the customers could get discount at the moment and there is no need to wait for confirmation. Another solution is *multilevel review*, which one set of workers does the work, and immediately a second set reviews the work and rate its quality [9]. By integrating other quality assurance approaches we could increase the overall quality of crowdsourcing results and leverage our approach's reliability.

To prove that routing tasks to Amazon's customers is a good approach and Amazon customers are gravitating toward this approach, we conducted a survey. Participants of the survey are from different countries with various skills and interests. The following section represents the results of the survey in detail.

#### 4. EVALUATING AMAZON'S CUSTOMERS TENDENCY TO PARTICIPATE IN DOING HITs

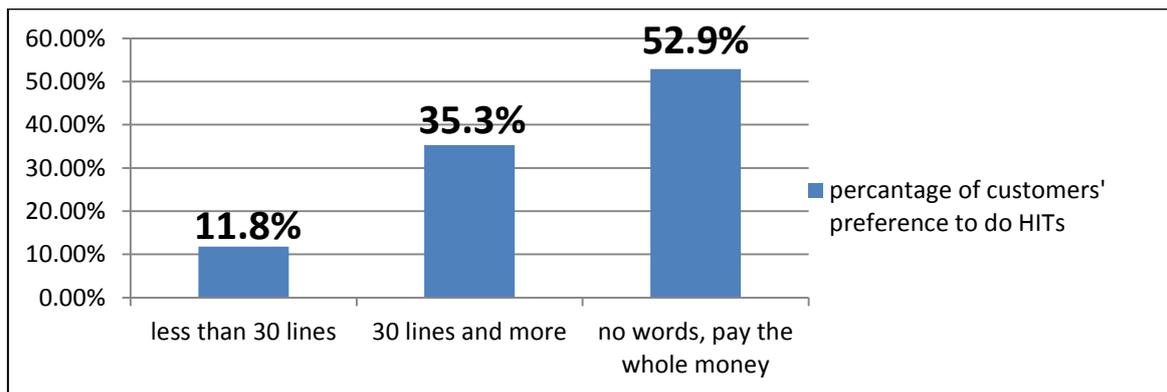
This survey's results shows that how many people around the world prefer to do tasks to get discount on their purchase instead of paying whole money of purchased goods, and more specifically how many of Amazon customers are interested in getting discount by doing tasks in purchase interaction. In total we got 1200 Amazon customers from various countries. We're going to analyze Amazon customers' answers to the survey. The results indicate that majority of people prefer to do tasks to get discount from Amazon. We asked two questions, one was "what's your maximum threshold to type security image (CAPTCHA) to get 3\$ discount for each 300 words? CAPTCHA here is words of scanned books and it is easily readable by human eye. The result of this survey is shown in figure 3. It shows that just 46.2% of Amazon's customers don't like to do HITs to get discount, but 53.8% prefer to work on HITs to get discount, from this percentage 50% prefer to do more and more HITs to get more and more discount. We could conclude that Amazon customers prefer to do tasks to get discount.



**Fig 3. Customers' preference to fill security image textbox to get 3\$ discount for each 300 words that they fill**

If 53.8% of Amazon customers do tasks, the qualified and skilled workers of AMT will increase. Studies show that the number of active Amazon Customer accounts is almost 144,000,000. If 53.8% of these customers participate in doing tasks, the amount of AMT workers will increase to 77,472,000. This is much more than the amount of 15,059 and 42,912 AMT's workers that Fort et al. hypothesized [10]. So through this approach we bring much more qualified workers to do tasks besides previous AMT workers.

The other question that we asked from our participants was, "what's your maximum threshold to translate words from one language to other language that you know well, to get 3\$ discount for each 300 words?" Figure 4 shows that 47.1% of people prefers to do translation HITs (complex HITs) to get discount on their purchases. In comparison with the first question the percentage is lower; this shows that people inclined to do simple tasks.



**Fig 4. Customers' preference to translate words from one language to other language that they know, to get 3\$ discount for each 300 words that they translate.**

## 5. THE ADVANTAGES OF OUR PROPOSED AMT CROWDSOURCING APPROACH

We mentioned some of the AMT's drawbacks in section 2.2; in this section we're going to represent our new approach's advantages that solve the previous method's drawbacks.

We brought HITs into Amazon's purchase transaction. As we represented in previous section, about 50% of Amazon's customers prefer to do HITs to get discount on their purchase, as a result more HITs will be done each day, the completion time of tasks will decrease, and more requesters will be motivated to post their tasks in AMT.

Because the distribution of tasks is according to type of tasks and customer's interests and skills, and people who want to purchase something could do tasks, the amount of spammers will decrease to the minimum. Also in our proposed approach more requesters will post tasks and they specify rewards appropriately, consequently more qualified workers will do the tasks, so requesters could submit complex problems to be done by qualified workers.

Amazon collects customers' information according to their purchases, searches, and other features; so distribution of tasks

could be done according to the information that Amazon have from its customers. With this approach skilled, qualified, and interested customers will do the tasks, so the results will be of high quality. On the other hand if customers cheat on doing HITs for several times, they could be banned from doing other tasks or be fined.

In our approach, tasks that don't need special talents or skills are distributed randomly between customers with general knowledge, and tasks which need special talents are distributed between people who have talent and interest in those kinds of tasks, so no HIT remains unsolved.

## 6. CONCLUSION AND FUTURE WORKS

The purpose of this paper is to introduce a new approach to bring more qualified and skilled workers to participate in crowdsourcing efforts in AMT. This could be done by distributing tasks among different purchase transactions that customers have in Amazon site. The distribution process is according to the type of tasks and the characteristics of each group of customers. The creation of groups is according to customers' previous purchases, searches, trends, purchase

behaviors, and information that customers have provided to Amazon. Through this way the distribution of tasks will be easier and more efficient, because groups and their characteristics have been specified before.

We conclude that our proposed approach will increase the quality of results, it's time saving, efficient, and has less spammers, on the other hand more qualified workers will participate to do the tasks. To prove that our innovative model is practical and is satisfactory for Amazon customers, we conducted a survey. Based on the results of the survey and current literature we conclude that our approach is applicable and Amazon customers have tendency to do tasks to get discount on their purchases.

In summary our main contribution is to represent a new reputation system for Amazon Mechanical Turk marketplace. By our proposed reputation system, many people will participate in crowdsourcing efforts, and the efficiency of current AMT marketplace will increase.

One direction for future researches is to provide other incentives to motivate more Amazon customers to participate in doing tasks. Other directions for future work are to define better ways to distribute tasks among qualified customers, and finding ways to get more information about customers' characteristics. In future researches we will consider new features such as the amount of time each customer prefers to allot for each task, to classify customers into more appropriate groups. If we could define and improve these aspects, Micro-Tasks marketplaces specially AMT will definitely be the future of businesses. We conclude that many large tasks can be effectively designed and carried out via this method at a fraction of the usual expense and time.

## 7. REFERENCES

- [1] Sheng, V. S., Provost, F., and Ipeirotis, P. G. 2008. Get another label? improving data quality and data mining using multiple, noisy labelers. In *KDD '08: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, 614–622. New York, NY, USA: ACM.
- [2] Le, J., Edmonds, A., Hester, V. and Biewald, L. 2010. Ensuring quality in crowdsourced search relevance evaluation: The effects of training question distribution. In *Proceedings of the ACM SIGIR 2010 Workshop on Crowdsourcing for Search Evaluation*.
- [3] Kittur, A., Chi, Ed H., and Suh, B. 2008. Crowdsourcing user studies with Mechanical Turk. In *Proceeding of CHI*. April 5-10. Florence, Italy. ACM.
- [4] Howe, J. 2006. The Rise of Crowdsourcing. *Wired*.
- [5] Davis, J. 2011. From Crowdsourcing to Crowdservicing. *IEEE Internet Computing*.
- [6] Vukovic, M. 2009. Crowdsourcing for Enterprises, in *Proceeding of the 2009 Congress on Services – I*. IEEE Computer Society, Washington,DC, USA, 686-692.
- [7] Ipeirotis, P. 2010. Analyzing the Amazon Mechanical Turk Marketplace.
- [8] Ambati, V., Vogel, S., Carbonel, J. 2011. Towards Task Recommendation in Micro-Task Markets. *AAAI Workshops*, North America.
- [9] Quinn, A.J., and Bederson, B.B. 2011. Human Computation: A Survey and Taxonomy of a Growing Field. In *proceeding of CHI*.
- [10] Fort, K., Adda, G. and Cohen, K. B. 2011. Amazon Mechanical Turk: Gold Mine or Coal Mine? In *Association for Computational Linguistics*, Volume 37, Number2.
- [11] Geiger, D., Seedorf, S., Schulze, T., Nickerson, R. 2011. Managing the Crowds: Towards a Taxonomy of Crowdsourcing Processes, In *Processing of the seventeenth Americas Conference on Information Systems*.
- [12] Pan, Y., Blevis, E. 2011. A Survey of Crowdsourcing as means of Collaboration and Implications of Crowdsourcing for Interaction Design, *IEEE*.
- [13] Chen, J., Menezes, N., Bradley, A. 2011. Opportunities for Crowdsourcing Research on Amazon Mechanical Turk, Amazon.com, Inc.
- [14] Paolacci, G., Chandler, J., Ipeirotis, P. 2010. Running experiments on Amazon Mechanical Turk, *Judgment and Decision Making*, Vol. 5, No. 5.
- [15] Shaw, A., Horton, J., Chen, D. 2011. Designing Incentives for Inexpert Human Raters, *CSCW 2011*, March 19-23, 2011, Hangzhou, China.
- [16] Khanna, Sh., Ratan, A., Davis, J., Thies, W. 2010. Evaluating and Improving the Usability of Mechanical Turk for Low-Income Workers in India, *ACM DEV'10*, December 17–18, 2010, London, United Kingdom.
- [17] Chen, J., Menezes, N., and Bradley, A. 2011. Opportunities for Crowdsourcing Research on Amazon Mechanical Turk. In *proceeding of CHI 2011 Workshop on Crowdsourcing and Human Computation*
- [18] Eickhoff, C., De Vries, A. 2011. How crowdsourcable is your task? In *Proceedings of the Workshop on Crowdsourcing for Search and Data Mining (CSDM) at the Fourth ACM International Conference on Web Search and Data Mining (WSDM)*, pp. 11–14.
- [19] Dow, S., Kulkarni, A., Bunge, B., Nguyen, T., Klemmer, S. and Hartmann, B. 2011. Shepherding the crowd: managing and providing feedback to crowd workers. In *Proceedings of the CHI Extended Abstracts on Human factors in computing Systems*. New York, NY, USA. ACM. 1669–1674.
- [20] Brabham, D. 2008. Crowdsourcing as a model for problem solving. *The International Journal of Research into New Media Technologies*. Vol 14(1). pp. 75-90.
- [21] Ipeirotis, P.G., Provost, F., & Wang, J. 2010. Quality management on Amazon Mechanical Turk. *HCOMP'10*.
- [22] Kittur, A., Kraut, R.E. 2008. Harnessing the wisdom of crowds in wikipedia: quality through coordination, in *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, San Diego, California, USA.
- [23] Zaidan, O. F. Callison-Burch, Ch. 2011. Crowdsourcing Translation: Professional Quality from Non-Professionals. In *Proceeding of 49th Annual Meeting of the Association for Computational Linguistics*. Portland, Oregon. June 19-24. Association for Computational Linguistics. pp. 1220-1229.