● Andrea Carignani y Lorenzo Negri
Milan (Italia)

"Do we need education?" Required abilities in online labor markets: an empirical research

¿Necesitamos formación? Habilidades necesarias en los mercados de trabajo online: un estudio empírico

ABSTRACT
The paper presents a study conducted through a cross sectional research design and a quantitative content analysis method to categorize and provide a numerically based summary of the different abilities requested in online routine labor markets like Mechanical Turk. These markets are growing nowadays for the outsourcing of routine information processing tasks. This is mainly due to the existence of tasks that computerization is not capable to substitute. This research is a first attempt to study abilities required by employers from low skilled information processing workers in virtual marketplaces. It also points out the fact that a new generation of temporary workers is appearing on the labor market raising questions about their characteristics, behaviors and differences compared to high skilled knowledge workers. The unit of analysis is a sample of tasks published by requesters on Mechanical Turk.

KEYWORDS
Distributed work, Information worker, Work practices, Human information processing.

Introduction

Every human being that has almost a basic literacy is able to accomplish simple routine information-processing tasks like categorizing objects, writing simple sentences, reading and summarize brief texts or describing, labeling and comparing images. In the nineteenth century, technological innovations fostered the demand for these kind of tasks as testified by the sharply increase of clerking occupations. More recently, computerization substitutes for workers in executing routine tasks given rise to an increase in the demand for more educated workers and non routine cognitive tasks for which computer capital complements workers rather than substitute them (Autor et al., 2003; Spitz-Oener, 2006).

Nevertheless, there is a boundary to the substitution carried on by computer capital towards routine work. This boundary is given by the existence of tasks that are trivial and simple to describe for humans but challenging even for the most sophisticated computer program like image recognition, natural language processing or scenes understanding in presence of varying degrees of background noise, incompleteness and distortion (Ahn, 2008; Kenaw, 2008). These activities are informally known in the field of artificial intelligence as AI-complete problems because computers should be as intelligent as people to solve them (Mueller, 1987). Therefore, for all this tasks human capital rather than computer capital seems to be more suitable and productive (Carignani and Negri, 2010).

Nowadays, in response to this need, we are experiencing the growth of new labor market intermediaries. Over the Web (Autor, 2001) businesses have the chance to outsource routine work from remote locations to an indistinct mass of workers and people could lend their basic information processing capabilities to employers (Zittrain, 2008) becoming "processors in a distributed system, each performing a small part of a massive computation (Ahn, 2005). Furthermore, these new intermediaries are structured around the concept of distributed human computation, that is: "the strategy of combining the strengths of computers and humans by delegating parts of the problem to large numbers of humans connected via the internet - usually spread out geographically" (Quinn and Bederson, 2009).

In this work we will focus our research on online labor markets in which routine tasks are exchanged, although, over the Web, are also appearing labor market intermediaries which regulate the exchange of non routine work activities. We motivate this choice by our research problem: what are the skills required by employers in online labor markets in which routine tasks are exchanged? That is, what are nowadays the characteristics that low skilled information workers must have to participate in online routine labor markets? In particular, we will focus on a new electronic market intermediary called Mechanical Turk (MT) and on requirements expressed by employers (known in MT as requesters). Over the last two decades we have seen a growing body of research regarding how the complexity of jobs is increasing causing an upgrading of the skill level required in most occupational fields. A proof of this phenomenon is the fact that the labor market is becoming increasingly polarized with a rapid growth in high skill professionals with high skills wage and low skill workers belonging to the service sector. Computerization has been indicated as a primary contributor of the present skill biased technological change because it substitutes routine work based on information, leaving available only high abstract tasks and manual tasks (i.e., professionals and service work) and decreasing de facto clerical jobs (Hilton, 2008; Maxwell, 2008). What is not deepened in this literature is that new online labor markets are growing nowadays for the exchange of routine information processing tasks that are trivial to describe for humans but difficult to be codified for computers. In this way, not only low skilled information processing workers are not disappearing but they are a numerous and multi nationality force (online labor markets are accessible from all over the world) eligible of research.

To the aim of studying which kind of abilities (skills) are required by requesters in Mechanical Turk we will address four strands of research: skill biased technological change, to convey a general point of view about which skills are more requested today both in high skilled that in low skilled occupations; literature about human computation, useful in deepening which kind of tasks are not reusable by computers; literature about internet skills, indeed, since the output of work in online labor markets is information we expect that some com-

puter related skills will be required and assignment literature, that suggests to consider both education and experience together with job characteristics when approaching a study about labor markets. This study is significant in that it provides a first attempt to study abilities required by employers from low skilled information processing workers in virtual marketplaces. It also points out the fact that a new generation of temporary workers is appearing on the labor market raising questions about their characteristics, behaviors and differences compared to high skilled knowledge workers.

Therefore, the purpose of this study is an attempt to categorize and provide a numerically based summary of the different abilities requested in online routine labor markets. The unit of analysis will be a sample of tasks published by requesters on Mechanical Turk starting from 24th to 31th January 2010. Because, to the extent known by the authors, there is no previous literature about what abilities are required in Mechanical Turk we decided to use a descriptive content analysis guided by two research questions. The codebook used to code observed tasks descriptions into categories of required abilities was developed starting from the theoretical framework. The intent, in this work, is to provide an initial description of what abilities are required in Mechanical Turk in order to be able to depict more precise hypotheses in the near future. Finally, an attempt was made to construct a dictionary of terms for future content analysis in this domain.

Given the descriptive purpose of this research we will use a cross sectional research design and a quantitative content analysis method to provide a numerically based summary of abilities.

Related Literature

Even if humans could be used like processors in a distributed system, they would not actually do it without an incentive. Examples of incentive are altruism, fun or the classic monetary incentive. Mechanical Turk acts as an intermediary between employers and people. In this online labor market employers are charged for their use of the web platform and people are motivated to join in by a monetary incentive: this mechanism has led some authors to speak about the creation of limited access markets (Kleeman and Voß, 2008). Additionally, an exhaustive definition of Mechanical Turk is given by its creators: "Mechanical Turk is a service that provides a virtual marketplace where requesters advertise micro tasks and human workers complete them. Amazon Mechanical Turk is based on the idea that there are still many things that human beings do more efficiently than computers […] Amazon Mechanical Turk gives businesses immediate access to a diverse, geographically dispersed, on-demand, scalable workforce and gives workers a selection of thousands of tasks to complete whenever and wherever it's convenient." This definition clearly states the scope of Mechanical Turk and its main characteristic: adjectives like on-demand and immediate emphasize the absence of a stable relationship between workers and employers in favor of the freedom of the choice of whenever or wherever to work.

While such new intermediaries in electronic labor markets are not receiving much attention from literature, two recent proposals could help to conceptualize their properties and characteristics. The work from Quinn and Bederson (2009) offers taxonomy to compare key characteristics of distributed human computation systems. The taxonomy has been built around six dimensions, namely: motivation, quality, aggregation, human skill, participation time and cognitive load. Mechanical Turk is cited in this proposal as an example of mechanized labor, that is a form of crowd sourcing distinguished by the use of pay as a motivation to work; the use of expert review (review of work done by employers) to support quality; the presence of a high number of small independent and not aggregable tasks; and variable human skills, participation time and cognitive load required. Malone et al. (2009) propose a similar approach to help classifying collective intelligent systems by providing a set of building blocks that address who is performing the task, why are they doing it, what is being accomplished and how is it being done. Even if the authors do not cite directly Mechanical Turk we could argue from their framework that who undertakes the activity is the crowd that do it for a financial gain, creating something new (doing tasks which are mainly independent), giving origin to a collection of well done tasks for an employer.

© ISSN: 2255-1514

As stated above, to the aim of studying which kinds of abilities (skills) are required by requesters in MT we will refer to four different strand of research. During the exploratory phase of our work, we included each strand of research in order to create a coding scheme as complete as possible. The process we used was a combination of deductive reasoning (supported from a specific literature strand) and inductive reasoning on data analyzed. This approach, that challenges the a priori design of a coding scheme, is necessary due to the fact that there is no previous literature that specifically addresses our research questions (Neuendorf, 2002).

Skill Biased Technological Change

In recent literature about skill biased technological change (SBTC) two general skills emerge as relevant from different studies applied to various working contexts and datasets related to knowledge workers, namely complex problem solving skills (using specific knowledge) and communication skills (de Grip, 2005). What is notable here is that knowledge workers need, nowadays, to develop social, interactive, interpersonal skills together with high specialized technical skills.

Nevertheless, less well studied in literature is the shift in abilities observed within low skilled workers. According to Hilton (2008) also in low wage/skilled service occupations is growing the importance of social and aesthetic skills (especially for those that work in contact with customers). Moreover, Maxwell (2008) finds that in low skill jobs office clerical skills, mechanical skills, and the new basic skills of reading, math and communication paid average higher wages than did other low skill jobs. To encompass these findings in our work we decided to include a set of variables that refers to generic clerical abilities. These abilities derived from the Occupational Information Network database1 (O*NET) and belong to the profile known as General Office Clerks[2] are: active listening, reading comprehension, speaking, writing, social perceptiveness, oral comprehension, oral expression, speech clarity, speech recognition, information ordering, number facility and mathematical reasoning. Nevertheless, since our focus is on an online labor market we suppose that there are no tasks requiring mechanical abilities, indeed all published tasks must end with an information output and not with a physical one. For this reason mechanical skills will not be included in our codebook.

Mechanical Turk is based on the idea that there are still many tasks that human beings do more efficiently than computers. Starting from this argument in a next step we reviewed literature about Artificial Intelligence and Human Distributed Computation to see if traditional clerical abilities, like those quoted before could be better specified to bring evidence of characteristics for which humans are better than machines.

Human Intelligent Tasks

In Artificial Intelligence literature the view that a machine could have the intelligence to accomplish any intellectual task that a human being can has been quite criticized. Basically, the characteristic feature of human beings is to have a holistic perception of events whereas in a machine this recognition is based on the perception of isolated stimuli. The holistic perception enhances the possibility to perceive not only the object of interest but also the context where that object is situated. Doing so human beings reduce the ambiguity of meanings about sentences, gestures, etc. (Kenaw, 2008). Starting from the work of Stork (1999) through the literature about Human Computation (Ahn, 2005) we collected examples of tasks that are trivial and simple to describe for humans but challenging for computer programs, that is: tasks that require computer programs as intelligent as people to solve them. From these tasks we identified four basic abilities. The first two will include traditional clerical abilities giving them a more specified meaning: natural language understanding that includes and specifies both speech recognition and oral comprehension; recognizing and creating categories of objects and actions from images or video clips that includes and specifies the ability of category flexibility. The last two, namely Providing personal opinions, comments, vision of the world, personal attitudes and experiences (we will use the abbreviated form providing personal opinions and comments in the following discussion) and Turing test refer to the ability of people to reason about the world generating visions of the reality and to the capacity for people to prove in every moment to be human and not a machine. This last point is more and more

necessary today when navigating the Web: indeed, a lot of tests have been developed in order to prevent spamming and distinguish between internet users and automated tools (BOT) like the completely automated public Turing test to tell computers and humans apart (CAPTCHA).

Internet Digital Skills

The Web is the natural element through which Mechanical Turk tasks are displaced, completed and submitted to requesters. Moreover, the use of Internet as a source of information into daily working life it's a well know topic in literature. Therefore we needed to include variables which address requested abilities to surf, to collect or to manage information over the Web. For this purpose we addressed literature about computer literacy, in particular that strand of research studying Internet skills. We refer here to the work of van Deursen (2009) which summarizes four different research directions providing an integrated view about Internet skills. Following the work of the author we included operational skills (operating an internet browser, operating internet based forms); information internet skills (Being able to locate required information on the Internet) and formal internet skills (navigating on the internet). Strategic internet skills (taking advantage of the internet) which comprises considerations about the objective of the search will be excluded. We will also consider the work of Hargittai (2007), which proposes a framework for studying differences in people's digital media uses, including a category called vote, rank or share content on the web that encompass the author's categories of knowledge to how to contribute to group discussion and knowledge to how to share content.

Assignment Models and Outsourcing of Jobs

While developing our research problem we referred to assignment models as a generic framework that guided us in our exploration of the Mechanical Turk market. Even if a full revision of the assignment models is not consistent with the objective of this work, we would like to state here that we adopt the view that when studying a labor market both human capital investment and job characteristics (that is, required levels of education and skills) are relevant (Sattinger, 1993). It was this view that leads us to consider what abilities are required in MT and is for the same reason that we will include requested education and requested experience in our analysis. Moreover, when the output of work is information then there is the possibility to divide it into parts and to move it through the web to low wages area. Through Mechanical Turk the outsourcing of jobs becomes an inexpensive reality. Indeed, a website like MT is reachable from anywhere there is an internet connection. This reduces the cost to move people and machinery undertook by employers in traditional forms of outsourcing. Also, employers don't have the problem to export specific knowledge: if works posted on MT don't require any particular knowledge all they have to do is to explain how the tasks must be accomplished. For this reasons we will also include in our analysis a category called localization that refers to those tasks for which an employer has specified the geographic location of requested workers.

Methodology

In this research we identified the following research problem and questions. The research problem could be stated as: what are the skills required by employers in online labor markets in which routine tasks are exchanged? That is, what are nowadays the characteristics that low skilled information workers must have to participate in online routine labor markets?

- RQ1: What are the most requested abilities in Mechanical Turk?
- RQ2: Which of the most requested abilities require more education, experience or localization?

Research Design, Methods

We adopted a cross sectional research design and a quantitative content analysis method to provide a numerically based summary of abilities requested by employers in Mechanical Turk. We developed, tested and

revised a codebook that was subsequently applied to a sample of tasks published by requesters on Mechanical Turk starting from 24th to 31th January 2010. The codebook comprises 17 nominal level variables and the process of coding was carried on by two coders. The unit of data collection and analysis is every single task published by a requester. Specifically, for every single task we analyzed the title, the description and keywords used by employers to facilitate the online research of that task inside Mechanical Turk. A reliability test was conducted through Cohen's Kappa: all the variables (categories) resulted reliable with coefficients ranging from 0.50 to 0,75. Descriptive statistics and word counting procedures where used to represent the percentage of different abilities and the word frequency of tasks titles and descriptions.

The approach of studying the manifest content of messages published online by employers in order to infer what characteristics should have employees in specific labor market is not new in literature (Kuhn and Kailing, 2009). Additionally, as stated in Kennan (2008) a fundamental assumption of these kind of approach is that the content of job requests or ads is a valid picture of the labor demands of employers. Given that this approach is consistent with our research problem and questions we assume that studying the manifest content of published tasks is a valid method to understand what are the most requested abilities and between those abilities for which of them requesters require different education, experience and localization.

Population and Sample Characteristics

We first derive a sample size using formulas for categorical data and correcting it by the Design Effect (the intra cluster correlation was calculated with a previous pilot study) in order to have an effective sample size. Secondly, we identified as the population of interest all the tasks published at 12.00 p.m. starting from 24th to 31th January 2010. Then we applied a cluster sampling with probability proportional to size (PPS) with the aim of extracting 564 tasks. We opted for cluster sampling with probability proportional to size because the amount of tasks published in every moment in Mechanical Turk is highly variable so we need a procedure to keep the sampling fraction constant (Shmueli 2005). Using this method four cluster (days) where selected, namely 24th, 27th, 29th and 31th each composed by 141 tasks.

For an operational definition of variables see the codebook in Appendices A.

Results

Table 1 presents the fourteen categories considered in this work together with their observed frequency, percentage and cumulative percentage in the sample. The analysis of this table is the answer to the first research question: what are the most requested abilities in Mechanical Turk?

Table 1. Abilities required in the sample and their frequency

	Frequency	Percentage	Cumulative Percentage
Providing personal opinion and comments	152	27,0	27
Locate information on the web	109	19,2	46,2
Writing expression	57	10,1	56,3
Natural language understanding	55	9,8	66,1
Recognizing and creating categories of objects and actions from images or video clips	38	6,7	72,8
Operating internet based forms	37	6,6	79,4
Operating Internet Browser	31	5,5	84,9
Navigating on the Internet	29	5,1	90
Vote or ranking content	21	3,7	93,7
Turing Test	17	3,0	96,7
Reading comprehension	11	2,0	98,7
Communication	4	0,7	99,4
Mathematical reasoning	2	0,4	99,8
Information ordering	1	0,2	100
Total	564	100	

Table 1. Abilities required in the sample and their frequency

According to empirical results the first five categories represents the 70% of the composition of requested abilities. The most requested ability is providing personal opinion and comments: a simple word counting applied to title, descriptions and keywords inside this category (Table 2) reflects the fact that what is most requested here is the opinion of people and their response to specific short questions and survey. Examples of tasks under this category are "Answer a Short Survey about your lunch habits", "Write a short answer to a question about parenting", "Give detailed feedback about how you experienced a website" or "What do you consider when buying a car?"

Table 2. Word counting of Providing personal opinion, writing expression							
Providing personal opinion				**Locating information on the Web**			
Content		Keywords		Content		Keywords	
Word	Count	Word	count	Word	Count	Word	count
question	124	answer	102	Term	127	research	67
answer	118	question	62	describe	67	question	18
Short	64	write	61	research	66	choice	9
Review	59	survey	52	explain	63	answer	9
Write	47	opinion	44	Find	40	multiple	9
survey	32	short	42	website	35	data	7
provide	11	advice	42	information	20	video	7
Tell	9	review	27	question	18	write	6

Table 2. Word counting of Providing personal opinion, writing expression.

The second most requested ability is the capacity to locate information on the web: action verbs like research, find, explain or nouns like term and website clearly states what is asked to do within these tasks (Table 2). Examples are "given a website, find the phone number for this business" or "Find an article on Trees from this website". Third, we find the writing expression ability where words like write, word, article and sentence are often recurring (Table 3). Examples of tasks which required this ability are "Write a 200-300 word original article", "Rewrite a 250-word article" or "Write a science fiction themed short story".

Table 3. Word counting of writing expression and Natural language understanding							
Writing Expression				**Natural Language Understanding**			
Content		Keywords		Content		Keywords	
Word	Count	Word	count	Word	Count	Word	count
write	58	write	40	audio	90	transcribe	49
word	32	sentence	12	transcription	43	cw	47
article	25	article	12	difficult	43	podcast	45
short	23	rewrite	11	premium	22	castingword	45
question	17	easy	9	express	14	English	42
title	14	article	8	test	12	mp3	39
review	13	fast	8	transcript	12	transcription	7
description	13	title	7	New	12	edit	4

Table 3. Word counting of writing expression and Natural language understanding.

Natural language understanding accounts for 9,8% of the total abilities required and it requires mainly to transcribe audio files. This kind of tasks have often all the same description starting with "Difficult Audio Transcription: …". In this sample the word difficult emerges as strictly connected to this kind of ability (Table 3).

Recognizing objects from images and videos requires classifying, label and categorize images and videos as emerge from the word counting of keywords and descriptions. Operating internet based forms, Operating Internet Browser and Navigating on the Internet increase the cumulative percentage by approximately 17%.

Operating forms requires testing data using forms as emerges from the word counting. Nevertheless, what is notable here is the word test, indeed more than simple operating internet based forms it seems that this ability encompass the need to test online websites or software as some of the published tasks clearly depict: "Simply cut and paste data from an email into a form" or "Test data collection, input data, test email submit-Fast & Amp; Simple".

Table 4. Word counting of Recognizing objects from images and videos and Operating internet based forms

Recognizing objects from images and videos				Operating internet based forms			
Content		Keywords		Content		Keywords	
Word	Count	Word	count	Word	Count	Word	count
image	65	image	43	test	31	survey	31
object	20	categorization	21	data	15	fast	30
adult	18	photo	20	survey	14	quick	27
classify	17	label	20	form	13	easy	21
product	16	photos	19	download	12	money	14
depict	16	object	18	simple	12	web	14
concept	16	classify	16	page	11	data	13
pick	16	product	8	new	11	test	12

Table 4. Word counting of Recognizing objects from images and videos and Operating internet based forms

Operating Internet Browser, consistently with the description of this category in the codebook requires mainly to visit a website or to bookmark it. Whereas, in order to complete navigating on the internet tasks the only ability required is to click on a link (Table 5).

Table 5. Word counting of Operating internet browser and Navigating on the internet

Operating internet browser				Navigating on the internet			
Content		Keywords		Content		Keywords	
Word	Count	Word	Count	Word	Count	Word	Count
site	20	bookmark	14	link	38	click	20
bookmark	16	social	9	click	36	money	10
website	15	home	8	visit	16	survey	10
go	11	easy	7	easy	14	simple	9
page	10	fast	6	website	13	work	9
find	9	work	5	test	12	job	9
copy	8	data	5	hit	12	easy	7
url	8	quick	5	fast	11	fast	5
twitter	6	business	4	payout	9		

Table 5. Word counting of Operating internet browser and Navigating on the internet.

The last six categories (vote or ranking content, Turing test, reading comprehension, communication, mathematical reasoning and information ordering) account only for a 10% of the entire percentage. Examples of tasks descriptions belonging to last six categories are, in order: "Help me Rank my Brand New Article on How to Burn Fat and Gain Muscle" for the vote or ranking content category; "Do 20 image captchas for us and answer few questions about yourself. Should take you between 1 and 3 mins". For the Turing test category; "Given a topic and an extraction from a document, is the highlighted sentence relevant to the topic? Or, given an event, is the highlighted sentence relevant to the event?" for the reading comprehension category; "Be available for a 5 minute phone call about Honeymoon Registries." For the communication category and finally "Solve a geometry problem" and "What Order Do These Pictures Go In?" respectively for the mathematical

reasoning and the information order category (Table 6).

Table 6. Word counting of the last six categories in order of importance							
Vote and ranking content				**Turing Test**			
Content		Keywords		Content		Keywords	
Word	Count	Word	Count	Word	Count	Word	Count
rate	13	bookmark	5	create	10	money	7
vote	7	social	4	email	7	sign	6
Reading Comprehension				**Communication**			
Content		Keywords		Content		Keywords	
Word	Count	Word	Count	Word	Count	Word	Count
sentence	13	search	4	dialog	2	phone	4
given	11	relevance	4	utterance	2	conversation	2
Mathematical reasoning				**Information ordering**			
Content		Keywords		Content		Keywords	
Word	Count	Word	Count	Word	Count	Word	Count
diagram	2	problem	1	order	3	memory	1
solve	2	draw	1	given	1	predictable	1

Table 6. Word counting of the last six categories in order of importance.

In Table 7 there is the answer to the second research question, that is: within the most requested abilities, for which of them is required more education, experience or localization?

Table 7. Education, Experience and Localization requirements				
Category id	**Category**	**Education**	**Experience**	**localization**
A	Providing personal opinion and comments	4	63	50
F	Operating Internet based Form	5	14	23
B	Writing expression	6	32	9
I	Recognizing and creating categories of objects and actions from images or video clips	4	8	7
C	Locating information on the Web	8	30	5
H	Turing	0	10	5
E	Navigating on the internet	3	20	4
D	Operating Internet Browser	2	24	3
G	Vote and ranking content	5	13	2
L	Communication	0	2	2
K	Natural Language Understanding	52	3	1
J	reading comprehension	2	6	1
M	mathematical reasoning	0	2	0
N	Info ordering	0	1	0

Table 7. Education, Experience and Localization requirements.

Before examining in detail the answer to the second research question a clarification is needed about the meaning of the categories education and experience in Mechanical Turk. Indeed, education in Mechanical Turk is not like formal education which is normally requested in the labor market, but is a sort of test of the ability possessed by workers. For this reason these kind of tests are called qualifications in MT. For example, if we look a table 7 we could see that Natural language understanding has the biggest request of education. The most frequent education request in this category is to have a score in CastingWords PPT. the higher the score the higher the possibility to work with difficult audio transcription and consequently to earn more money. In the example Catingwords is a requester that has determined a minimum level of education that differs from task to task. Therefore in this work, education is qualifications created by requesters in order to match a desired workforce with a set of tasks. Otherwise, Experience is a measure of worker's history and reputation. In

Mechanical Turk experience is calculated for every worker regardless of the specific requester. Examples of experience are "HIT[3] approval rate is not less of 95" that is, not less of the 95 percent of the tasks done have been approved by a requester or also "HIT abandonment rate (%) is less than 50".

Looking at the distribution of education, experience and localization requirements two more comments could be made (Figure 1). First, the majority of categories require no localization. Tasks that need a localized workforce belong mainly to the providing personal opinion category: the reason is that all survey and questions (probably used for marketing research) are shaped around a specific kind of consumer. Information gathered through this category is valuable only in certain economic environments. The second category which requires localization is operating forms. This result was a surprise and it strengthen the idea that the operating forms category needs to be detailed and better specified by gathering and codifying more tasks through a study conducted on a more long range of time. With data sampled in this analysis we could argue that in this category a localized workforce is needed to test websites or to make data entry or that localization is used by requesters for tax purposes. Second, the majority of categories are situated in the area of no education, no localization and little or no experience.

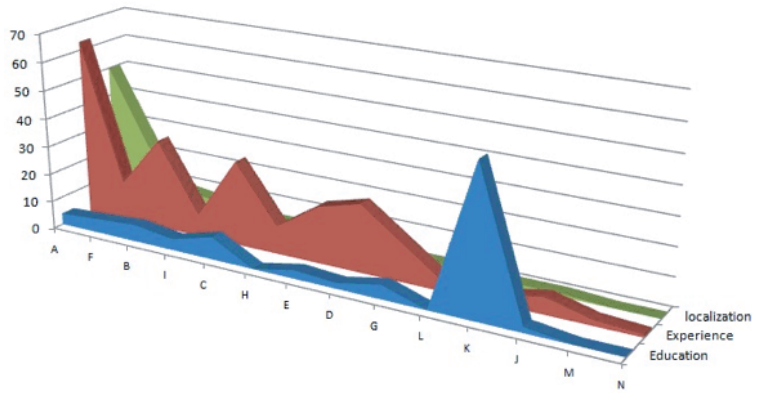

Figure 1. Education, Experience and Localization for Categories.

Discussion

We began this work by depicting the possibility that a mass of low skilled information workers are lending, nowadays, their basic information processing capabilities in online labor markets such as Mechanical Turk.

The evidence of our results supports these hypotheses. Indeed, Mechanical Turk is used primarily to collect information about consumer habits. This is achieved paying workers to answer surveys. Secondly, some tasks that are simple to describe for humans but challenging even for the most sophisticated computer program are outsourced to the mass of workers like locating information on the Web (in order to verify data already in posses of the employer), transcribing audio files or recognizing and categorizing objects from images and videos with the purpose of putting in order catalogue of images of products or similar. Thirdly, a lot of abilities regarding the capacity to create or to share content are requested: like writing articles, ranking them or testing existing websites. These findings state that there are some categories of tasks for which human work is still more productive than the use of software or of information systems. Also, what is relevant here is that there is a marketplace for these tasks where employers are willing to pay for work and people is desirous to make use of their abilities to earn moneys. We believe this has to be taken into consideration in the information system literature when taking decisions about what kind of tasks could enclosed in systems rather than being outsourced to humans. In this way people is becoming like processors in a distributed system made of people and machines. That is, in some situations humans substitute for computers in information processing tasks.

Another issue emerging is that all these tasks require only basic education (like writing or natural language

understanding) and a basic knowledge of how to operate internet browsers and how to use search engines to find information on the Web (like navigating on the internet and operating internet browsers). For these reasons, provocatively, it seems like everyone could work in Mechanical Turk without the need to have a formal education. However, some recent studies dispute and shed light on the essence of the Mechanical Turk workforce and therefore help us to understand the requirements of abilities, education, localization and experience expressed by employers. Ross (2009) reports how the majority of turkers are from the United States (57%) and from India (32%) and the remaining are from countries ranging from Australia to Ukraine. Surprisingly, more than half of the workers have a college or advanced degree. One third of respondents to the authors' survey are employed full time, one third is full or part time students and the other third are currently unemployed. From these findings it doesn't seems like turkers are a low skilled workers population, but even this they do everyday routine tasks. The reason for this behavior is the need to transform spare time into productive time (Carignani and Negri 2010). This need arises for different reasons like a primary work that is boring, a period of unemployment or physical disabilities and brings people to do repetitive tasks that they perceive like funny, brain stimulating or tasks which allow to learn something new (like searching information on the Web) or practice in something useful in real-world works like creative writing or fast typing while hearing audio files.

Therefore, it seems like turkers are a middle skilled workforce using Mechanical Turk for doing brain stimulating tasks or for keeping productive while experiencing an increased amount of spare time in their life. Additionally, the requirements of localization expressed by employers must be interpreted like the desire to gather information only from specific geographical locations. Localization is no more a matter of costs (to outsource a tasks where the work cost less) but a matter of information (to gather information where it is relevant). In this way the ability of providing personal opinion and comments is highly localized. Education and Experience are used by employers to assess the goodness of a worker. An education level is a test that workers must pass to start working for an employer or to access more difficult and well paid tasks. Education differs from experience because this last one is based only on the number of tasks approved by all employers. From our results it appears that the ability to understand natural language is the only one which requires a specific test before accessing to tasks whereas experience is a common requirement along all the categories.

Finally, we would like to suppose in this paper that a new kind of worker (a new kind of work) is appearing on the scene. Certainly turkers are not teleworkers. Indeed, they experience the total absence of a direct relationship with an employer. They use telecommunications not as a way to keep in contact with the office and coordinate their work with that of colleagues but as a mean to access a generic and everyday different labor market. From this perspective turkers are more similar to freelance knowledge workers. Nevertheless, the findings of this paper show how the tasks that they do are routinary and they did not require an high formal education to be completed. An interesting perspective comes from the concept of immaterial labor. Immaterial labor is the conflation between production and consumption where leisure time and working time are strongly linked making life inseparable from work (Coté and Pybus, 2007). Furthermore, according to Lazzarato (1996) the concept of immaterial labor refers to two different characteristics of labor: it refers to the informational content of work and to the increasing need for skills involving cybernetics and computer control; and it involves activities that are not normally recognized as work (like fixing cultural standards, tastes, consumer norms, etc). In the words of the author immaterial workers are "primarily producers of subjectivity". We state here that the concept of immaterial labor suites very well with some of the abilities required working in Mechanical Turk such as providing personal opinions and comments or vote and ranking content. Indeed, the output of these abilities is the creation of subjectivity. Other abilities like locating information on the Web or recognizing and creating categories of object and actions from images and video clips are all activities not normally recognized as work because they are too easy to do to be recognized as such. Indeed, turkers do it, as noted previously, for fun or for increasing their abilities: they approach Mechanical Turk as it were a video-game.

Through this work we have described what kinds of abilities are requested from employers in one of the most used online labor markets of this kind: Mechanical Turk. The majority of the tasks published are, according to the formula of distributed human computation, accessible to all workers despite their geographic dispersion. This is true with the exception of tasks which want to collect information about consumer habits and

tasks that search workers to test websites or to make data entry where localization is used by requesters, probably, for tax purposes. Education, that in Mechanical Turk appears in the form of qualifications created by requesters in order to match a desired workforce with a set of tasks, is used extensively only when a transcribing ability is required: in all the other categories education requirements are low or totally absent.

The worker's profile that emerges from this work is extremely interesting. It seems like no formal education or just minimal education is required: writing, reading, categorizing, understanding of speech are all abilities early learned in most industrialized countries. Moreover, an ability to use the Web only in its basic functioning is required: navigating, searching, and using browsers and comments. Nevertheless, this profile conflicts with the real composition of the Mechanical Turk workforce: more than half of workers have a college or advanced degree. The resolution of this conflict lies in the comparison between the required abilities identified in this work and some observations about the motivation of people working in Mechanical Turk.

Finally, we state here the possibility that working in Mechanical Turk has some similarities with the concept of immaterial labor because it involves activities that are not normally recognized as work as well as a strong focus on the production of subjectivity. For these reasons we advance the hypotheses that a new kind of work, different from telework or knowledge work, is developing nowadays in electronic labor markets.

To further investigate the characteristics of workers in Mechanical Turk it could be interesting to overcome some limitations of this paper: first, it should be useful to expand the range of time under analysis in order to be able to estimate a percentage of a category having as a population a longer time than a week. Second, it could be useful to test the codebook developed here using more than two coders. Third, an attempt to use dictionary emerged from data to develop a computer coding should be made.

Moreover, there are many opportunities for future research. In fact, as stated above, there is little or no literature about Mechanical Turk as a new kind of electronic market. Indeed, while some works have been written regarding the use of Mechanical Turk as tool for research little attention has been paid in literature to understand Mechanical Turk as an electronic market per se. Starting from this descriptive analysis different questions arise: does the kind of ability requested influence the posted reward of a task? Or is this relation better explained by education, experience and localization requirements? In which way Mechanical Turk supports the distribution of work over the Web? Could we assert that a new kind of work (i.e., a new kind of worker) is appearing on the scene? What are its characteristics in relation to other work's typologies like telework or knowledge work?

All of these are possible research questions and, despite limitations, our findings serve as a starting point and we hope they stimulate further research.

Notes

(1) The O*NET program is the United States primary source of occupational information. Central to the project is the O*NET database, containing information on hundreds of standardized and occupation-specific descriptors.

(2) O*NET code: 43-9061.00

(3) Tasks in Mechanical Turk are defined as Human Intelligent Task (HIT)

References

Ahn, L. V. (2005). Human Computation, Unpublished PhD thesis. University of Carnegie Mellon.

Ahn, L. V. (2008). Designing Games With a Purpose. Communications of the ACM, August, (51:8), pp. 58-67.

Autor, D. H. (2001). Wiring the labor market. Journal of Economic Perspectives, winter, (15:1), pp. 25-40.

Autor, D. H., Levy, F., Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. Quarterly

Journal of Economics, November, (118:4), pp. 1279-1333.

Carignani A., Negri L., (2010). Transforming spare time into productive time: investigating turkers motivation to work in Proceedings of the IADIS International Conference, 26-31 July, Freiburg, Germany

Cote, M., Pybus, J., (2007). Learning to Immaterial Labour 2.0: MySpace and Social Networks ephemera, 7 (1), 88-106.
De Grip, A., and Zwick, T. (2005). The employability of low-skilled workers in the knowledge economy. London School of Economics, LoWER Final Paper, available at http://rlab.lse.ac.uk/lower/final_papers/grip.pdf (accessed 13 November 2008).

Hargittai, E. (2007). A framework for studying differences in people's digital media uses. In N. K. a. H.-U. Otto (Ed.), Cyberworld Unlimited, pp. 121-137. Wiesbaden: VS Verlag für Sozialwissenschaften/GWV Fachverlage GmbH.

Hilton, M. (2008). Skills for Work in the 21st Century: What Does the Research Tell Us. The Academy of Management, (22:4), pp. 63-78.

Kenaw, S. (2008). Hubert L. Dreyfus's Critique of Classical AI and its Rationalist Assumptions. Minds & Machines, June, (18), pp. 227–238.

Kennan, M. A., Willard, P. , Cecez-Kecmanovic, D. and Wilson, C. S. (2008). IS Knowledge and Skills Sought by Employers: A Content Analysis of Australian IS Early Career Online Job Advertisements.
Australasian Journal of Information Systems, (15: 2). Available at SSRN: http://ssrn.com/abstract=1168044

Kleemann F., Voß G., Rieder K. (2008). Underpaid Innovators: The Commercial Utilization of Consumer Work through Crowdsourcing" Science, Technology & Innovation Studies, (4:1), pp. 5-26.

Kuhn, P. and Kailing, S. (2009). Employers' Preferences for Gender, Age, Height and Beauty: Direct Evidence. No 15564, NBER Working Papers, National Bureau of Economic Research, available at http://econpapers.repec.org/RePEc:nbr:nberwo:15564.

Lazzarato Maurizio, (1996). Immaterial Labour, in Paolo Virno & Michael Hardt (eds.) Radical Thought in Italy: A Potential Politics, Minneapolis: University of Minnesota Press, 133-147.

Malone, T. W., Laubacher, R., Dellarocas, C. (2009). Harnessing Crowds: Mapping the Genome of Collective Intelligence. Working Paper 2009-001, available at http://cci.mit.edu/publications/workingpapers.htm (accessed 02 October 2009)

Maxwell, N. L. (2008). Wage Differentials, Skills, and Institutions in Low-Skill Jobs. Industrial & Labor Relations Review, (61:3), pp. 394-409.

Mueller, E. T. (1987). Daydreaming and Computation, Ph.D thesis, University of California.

Neuendorf, K. A., (2002). The Content Analysis Guidebook, Thousand Oaks, CA, Sage Publications.

Quinn, A. and Bederson, B. (2009). A Taxonomy of Distributed Human Computation. available at http://www.alexquinn.org/ (accessed 20 February 2010)

Ross J., Zaldivar A., Irani L., Tomlinson B. (2009). Who are the Turkers? Worker demographics in Amazon Mechanical Turk. Technical Report, SocialCode-2009-01, University of California, Irvine.

Sattinger, M. (1993). Assignment Models of the Distribution of Earnings. Journal of Economic Literature, June, (31:2), pp. 831-880.

Shmueli, G., Jank W, and Bapna, R. (2005). Sampling eCommerce Data from the Web: Methodological and Practical Issues. In Proceedings of the American Statistical Association, Statistical Computing Section, Alexandria, VA: American Statistical Association.
Spitz-Oener, A. (2006). Technical Change, Job Tasks, and Rising Educational Demands: Looking outside the Wage Structure. Journal of Labor Economics, (24:2), pp. 235-270.

Stork, D., G. (1999). The Open Mind Initiative. IEEE Expert Systems and Their Applications, May-June, (14:3), pp. 16-20.

Van Deursen, A..J.A.M. and Van Dijk, J.A.G.M. (2009). Using the Internet: Skill Related Problems in Users' Online Behavior. Interacting with Computers, 21, pp. 393-402.

Zittrain, J. (2008). Ubiquitous Human Computing. Oxford Legal Studies Research Paper, No. 32/2008, available at http://ssrn.com/abstract=1140445 (accessed 15 November 2009)