Crowdsourcing for Industrial Problems

Victor Muntés-Mulero[1], Patricia Paladini[1], Jawad Manzoor[2], Andrea Gritti[2], Josep-Lluís Larriba-Pey[2], and Frederik Mijnhardt[3]

[1] CA Technologies, Spain
{Victor.Muntes,Patricia.PaladiniAdell}@ca.com
[2] Universitat Politècnica de Catalunya, Spain
{jawad,larri}@ac.upc.edu, andrea.gritti@est.fib.upc.edu
[3] University of Utrecht, The Netherlands
a.f.mijnhardt@uu.nl

Abstract. The generalized use of the Internet and social network platforms has changed the way human beings establish relations, collaborate and share resources. In this context, crowdsourcing (or crowd computing) is becoming a common solution to provide answers to complex problems by automatically coordinating the potential of machines and human beings working together. Several challenges still separate crowdsourcing from its generalized acceptance by industry. For instance, the quality delivered by the workers in the crowd is crucial and depends on different aspects such as their skills, experience, commitment, etc. Trusting the individuals in a social network and their capacity to carry out the different tasks assigned to them becomes essential in speeding up the adoption of this new technology in industrial environments. Capacity to deliver on time, cost or confidentiality are just some other possible obstacles to be removed. In this paper, we discuss some of these issues, provide solutions to improve the quality in systems based on the use of crowdsourcing and present a real industrial problem where we use the crowd to leverage the work capacity of geographically distributed human beings.

1 Introduction

For many years, we have seen a huge increase in the use of sophisticated systems that have enabled world-wide collaboration through our home PCs and, more recently, through ubiquitous hand-held devices. The purposes are as numerous as they are varied: content sharing, whether through blogs or many well-known peer-to-peer (P2P) applications; collaborative computation, starting from the early SETI@home project (setiathome.berkeley.edu) that was one of the first large-scale grid computing instances, and other examples.

People are gaining awareness of the power of collaborating through the network. We have recently seen national revolutions, like in Egypt, where people organized themselves using digital platforms. The *crowd* is becoming aware of its power, and the next natural step is to enhance the tools and modalities for collaborative computing. Powerful devices, like smartphones and tablets, are able to carry out an impressive amount and array of computation. P2P computing has shown to be feasible and efficient. We have some examples, such as Skype, that show the model is valid and can challenge serious cloud-based competitors, such as Google Voice.

J. Nin and D. Villatoro (Eds.): CitiSens 2012, LNAI 7685, pp. 6–18, 2013.

Not only machines but also real people are connected to the network combining their computing and thinking capacity. Trends seem to be pointing to this model as gaining the position to complement (or perhaps substitute) cloud computing: connecting people and machines in a single network. Nowadays, millions of people are asynchronously analyzing, synthesizing, providing opinion and labelling and transcribing data that can be automatically mined, indexed and even learned. Therefore, there is not much difference between this and classical computing: the crowd is working online, taking digital data as input and yielding digital data as output. The main difference is that human brain-guided computation is able to perform tasks that computers can hardly do, at overwhelming speeds. Tagging a picture or a video based on their content or answering questions in natural language, are just a couple of examples.

In this paper, we discuss about the typology of problems that can be solved through crowdsourcing at an industrial level and the main challenges to be solved to make this technology generally adopted. We present two elements that we consider essential for quality. First, we propose the Action-Verification Unit (AV-Unit), a quality control mechanism that helps organizing the crowd to not only perform actions, but also evaluate the quality of the results during the process. Second, we propose to use rewarding systems based on the quality delivered by each single worker. Also, we present a real industrial example: a crowdsourcing platform for translation that it is being developed by CA Technologies[1], using AV-Units. The system we present is novel in the sense that it proposes a model to perform a large number of tasks in parallel, and uses multiple levels of verification to ensure industrial quality standards. To our knowledge, it is also the first proposal that adds a quality-aware rewarding system that rewards workers based on a ranking that measures the quality of their work. Finally, our system allows sizing tasks at our convenience in such a way that we avoid the problem of the lack of context in translations based on isolated sentences.

This paper is organized as follows. Section 2 describes the state-of-the-are in crowdsourcing and presents some industrial applications that use crowdsourcing in order to solve non-trivial problems. In Section 3, we present a method to strengthen the capacity of the system to provide quality in the results. We also discuss about rewarding methods and scalability. Section 4 presents an example of a crowdsourcing platform developed at CA Technologies. Finally, Section 5 concludes and draws some future research lines.

2 Crowdsourcing

The term crowdsourcing was used for the first time by Jeff Howe in 2006 [8], referring to the increasing practice of outsourcing task to internet as an open call over a variety of users. Since then, crowdsourcing has evolved to exploit the work potential of a large crowd of people remotely connected through the Internet. For instance, recent work studies different typologies and uses of crowdsourcing and proposes a possible taxonomyin [6]. They categorized crowdsourcing depending on different methodologies and processes divided according to several dimensions that are shown to impact

[1] CA Technologies is a worldwide software and solutions provider that helps customers to make ICT management more agile, secure and flexible.

the behaviour of workers within the crowd, and the tasks that can be outsourced to the crowd.

As this idea increases in terms of popularity, several general purpose crowdsourcing platforms have appeared in the last years. For instance, Amazon Mechanical Turk (mturk.com) is a crowdsourcing marketplace that enables companies or individuals to utilize the human intelligence to perform tasks that are difficult for computers. The requesters post tasks known as Human Intelligence Tasks (HITs) that can be viewed by workers. Other examples like CrowdFlower (crowdflower.com) or ClickWorker (clickworker.com), extend Mechanical Turk capabilities offering a variety of crowdsourcing services. They improve quality by using gold standard units, redundant reviews of each data unit, etc. Their workflow management system divides complex tasks into smaller units and distributes them among the crowd based on the profile of individuals.

Quality control is a key point in crowdsourcing and it changes depending on the nature of the task which is crowdsourced. For instance, on the one hand, Lease and Yilmaz [10] show how the results obtained from the crowd are more inaccurate compared to laboratory participants. On the other hand, Yan et al. [14] present *CrowdSearch*, a system to search images on mobile phones using the crowd. In their work they show that workers are able to achieve over 95% precision. Other lines of research study the effect that different rewarding systems have on quality. For instance, Harris [7] shows that financial incentives actually encourage quality.

Crowdsourcing markets are traditionally used for simple and independent tasks. For example, labeling an image or finding relevance between search results. In [9], the authors present a framework that enables solving complex and interdependent tasks using crowdsourcing markets. The authors follow an approach similar to MapReduce for breaking down a complex problem into a sequence of simpler subtasks. The subtasks are solved in parallel by the crowd and the results are combined to form the final solution.

In the use case presented in this paper, we focus on the use of crowdsourcing for translation and, especifically, on software localization. Several research works are based on the use of Amazon Mechanical Turk for translation [4,5,11,16]. Zaidan et al. [16] propose some factors to select a good translation among a set of different versions. These factors include the workers country of residence, native language, etc. and each factor has a weight. In this way, the total score is calculated for a specific translated text. Experimental results show that some translations turn out to be very close in quality to the ones made by professional translators. Two other studies [2,4] investigate the use of crowdsourcing to evaluate the output of machine translated natural language. Gao and Vogel [5] present a case study of word alignment tasks performed by crowd on Amazon Mechanical Turk and Matteo et al. [11] use crowdsourcing to create corpora to feed and enrich Statistical machine translation. Other studies focus on using crowdsourcing for post-editing tasks. Bernstein et al. [3] propose the Find-Fix-Verify pattern to perform tasks like text shortening and proofreading. They split the tasks into a series of stages that utilize independent agreement and voting to produce reliable results.

However, these systems might not be suitable for implementing an industrial software localization process for two main reasons. Firstly, these systems are based on the resolution of very small tasks, mostly at sentence level, and their entire quality assurance methodology is based on this reduced amount of information. Secondly, many of them use automatic evaluation methods like BLEU [12] and METEOR [1] to evaluate the quality of translation. These methods however rely on the existence of pre-computed golden translations. Golden translations are rarely available for unpopular languages and cannot be used for new texts. Thus requiring human intervention to decide which translations are to be accepted and which not, adding an additional managerial layer to the process.

2.1 Is the Crowd a Universal Solution for Industry?

Not any problem in industry is suitable to be crowdsourced. First of all, many processes are and can be automated by machines instrumented with intelligent AI-guided algorithms. Secondly, many processes will always require an onsite workforce to manage and monitor the operations. However, between these two a wide variety of applications will be suitable to be crowdsourced. To determine whether a process can be crowdsourced we detect a set of proprietary characteristics that must be taken into account:

- ***Activities must not be easy to automate:*** crowd computing activities usually require to perform actions that are better solved by the human brain than by currently available algorithms. Usual examples are tasks where creativity is essential, such as the proposal of new designs; tasks where the geographic distribution of the individuals provides a higher quality access to information; or tasks where the complexity of the problems posed are so high that there are not predefined mechanisms to solve them. A more complete survey on the type of tasks suitable for crowdsourcing may be found in [15].
- ***Information involved in the process must not be confidential:*** since data must be sent to the crowd, processes that involve sensitive information are not suitable in general for this type of solution, unless data is preprocessed first, for instance, applying anonymization techniques. On the positive side, crowdsourcing also mitigates concerns about loss of privacy, since a single provider does not have a global view of anyone's data. Later on in this paper, we will see a real example that benefits from crowd distributing data so that nobody has a complete picture of all the documents.
- ***Training must be simple or highly automatic:*** complex training processes are not suitable for crowdsourcing since this would imply training thousands or even millions of people, which would be unaffordable.

If these conditions hold, the use of crowdsourcing has multiple advantages. First, crowdsourcing delivers elasticity. Analogously to cloud computing, by working with the crowd, we have a virtually infinite number of resources that may be allocated and deallocated depending on the workload. Therefore, crowdsourcing offers flexibility in processes that include human beings. Secondly, it eliminates middlemen costs. By building a platform to manage the crowd automatically, we gain direct access to the final workers, eliminating intermediate vendors that increase the cost of services.

Real Industrial Applications of Crowd Computing. Crowd computing has been successfully used for industrial applications. Some tasks which are difficult to automate, such as those based on innovation and design, are good candidates for crowdsourcing. NamingForce (namingforce.com) or Threadless (www.threadless.com) are just two examples. Other examples are focused on technical areas. For instance, InnoCentive (innocentive.com) is a problem solving marketplace that brings together solution seeking companies and problem solvers dispersed all over the world. TopCoder (topcoder.com) uses the crowd so solve complex programming challenges. There exist many other examples such as testing, like in the case of uTest (www.utest.com), or journalism, like OpenFile (openfile.ca).

3 Quality, Motivation and Scalability

Three main challenges when developing a crowdbased system can be found in quality, motivation and scalability. This section proposes an approach to deal with the questions: (i) how to deliver quality when working with the crowd; (ii) how to motivate the crowd to participate in industrial processes; and (iii) how to make systems scalable.

3.1 Quality Based on Organized Verification

Crowdsourcing is sometimes associated to low quality. The participation of a massive amount of people who are geographically distributed is intuitively assumed to be a barrier for quality. In general, it is difficult to establish automatic mechanisms in order to monitor quality in crowd-based systems, since the lack of automatic methods to solve problems usually imply a poor definition of quality and a high complexity in order to establish a universally accepted quality measure. How do we measure the degree of innovation of an idea, the beauty of a proposal or the best style for a text? These are measurements that are highly dependant on subjectivity. Nevertheless, many previous proposals are still based on methods to monitor quality based on automatic measures or golden solutions, which might not be realistic in many different scenarios. In this section, we propose a general mechanism to crowdsource the evaluation of quality in the crowd. In order to build a trustworthy crowdsourcing system effectively, we need to work on two essential aspects: mechanisms to coordinate workers to guarantee the proper evaluation of quality, and a reliable mechanism to monitor the skills of workers.

Crowd-Based Quality Evaluation. In this section, we propose a general mechanism in order to guarantee quality when working with the crowd. The basic idea of our proposal is that human beings are the best quality evaluation method in many situations. Therefore, we propose to subdivide any complex task in a series of subtasks that we call *Action-Verification Units* (AV-Unit). An AV-Unit establishes a relationship pattern between the workers of the crowd to help them work collaboratively to provide a higher degree of quality.

Figure 1 depicts an AV-Unit. An AV-Unit is divided into two phases: Action and Verification. In the Action phase a single worker performs a specific action. In the Verification phase a set of workers verify the quality of the output generated in the previous

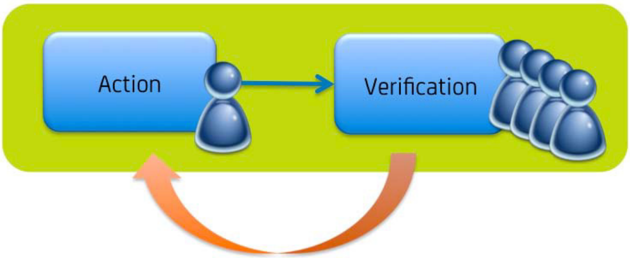

Fig. 1. Action-Verification Unit (AV-Unit)

Action phase. If the workers in the Verification phase consider that the quality provided is below a certain threshold, they might ask the first worker to repeat or improve the action. This process may be repeated iteratively until the output has reached a certain level of quality or the workers in the Verification phase decide to substitute the initial worker (or the worker is not available anymore). In practice, the Verification phase in the AV-Unit acts as a quality filter barrier, that does not allow to proceed with the process until the quality of each step in this process is approved by a set of human evaluators working collaboratively.

Figure 2 presents an example where a complex process is defined as the composition of AV-Units. It is important to remark that AV-Units do not necessarily have to be organized sequentially, and it is possible to create a complex network of interconnected AV-Units to solve complex processes.

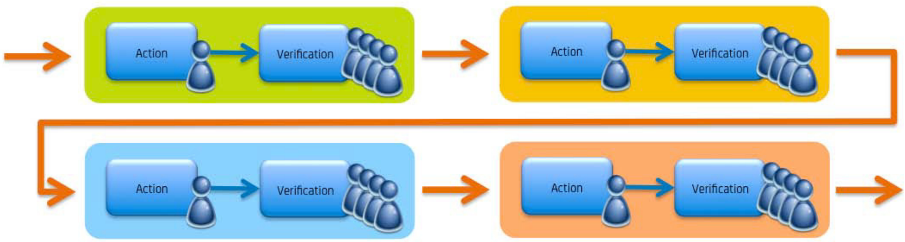

Fig. 2. Example of a use of AV-Units for general and complex problems

Measuring Trustworthiness with Worker Ranking. The success of AV-Units may depend highly on the profile of the workers. Involving many workers with poor skills in an AV-Unit, might have a negative impact on quality. In many industrial processes, quality standards are high and trusting the individuals in the crowd and their capacity to carry out the different tasks assigned to them becomes essential. Because of this, a primary concern of the system is to monitor workers in order to evaluate their skills.

Beyond AV-Units, we propose to develop crowdsourced industrial applications that take into account the quality produced by the actions done by the workers in the system and, their behaviour in general. For this, we propose to use ranking systems that are dynamically modified as the worker yields results and interacts with other workers in the system. Specifically, there are several aspects that might influence a ranking system:

– The quality measured from the output of the work produced by the workers in the crowd: it is necessary to establish rewarding and penalty measures that modify the ranking of the workers in the crowd. In general, the actions with a higher impact for workers are those performed in the Action phase of an AV-Unit. However, it would be also possible to modify the ranking of workers based on their activity when they are acting in a Verification phase.
– The behaviour of the workers in the crowd: other aspects might influence the ranking. For instance, a worker might click very fast in order to get solutions quickly and get an economical reward. Although, improper behaviours will lead with high probability to bad quality, taking into account behaviour patterns may help to multiply the positive or negative impact of action in the corresponding worker's ranking and speeding up the detection of incorrect behaviours.

The main idea behind using ranking systems is that a worker with a higher rank will be more trustworthy than other workers with lower ranks. As we mention at the beginning of this paper, establishing methods to automatically measure quality is complex, specially if we take into account that problems suitable for crowdsourcing are those which are not easy to automate.

3.2 Pay-per-Quality Rewarding System

There might be many different motivations for people to participate in a crowd-based process, ranging from their willingness to participate in a collaborative process to build something new, their will to help the community or their interest to be rewarded economically. Most industrial applications, if not all, pursue lucrative objectives. Because of this, industrial applications based on the crowd are quite more constraint in terms of motivating the crowd and tend to reward workers economically. The work presented in [13], for instance, confirms the importance of money compared to other motivations. We may still classifiy the different crowd systems depending on the rewarding model they use:

– ***Best-gets-paid systems:*** usually, in this type of systems, only the best workers get rewarded. In general, the system provides the tools to present ideas or solutions to a specific problem and a voting system for the crowd to decide the best proposals. This philosophy usually allows to reduce costs drastically and obtain very good quality, although it is in general unfair for workers, given that most of them work and are not rewarded, potentially becoming a source for lack of motivation.
– ***Pay-per-Work systems:*** in this case, workers are in general rewarded by the amount of work done. This is for instance the philosophy of Amazon's Mechanical Turk, where workers execute Human Intelligence Tasks (HITs) and get a predefined amount of money for it.

However, these two systems do not take quality into account. In this section, we propose to couple quality and the rewarding system. With this purpose, we propose to build crowd-computing systems based on a variant of a Pay-per-Work system. We call them Pay-per-Quality systems. The fundamental idea is that workers get paid for their work, but the amount that each worker receives depends on the profile of the worker. In other words, the rewarding system depends on the rankings of the workers. In this way, a trustworthy worker will be better rewarded than an unexperienced worker, or a worker with lower skills in general. In the following section, we present an example.

3.3 Scalability

Since one of the key aspects of crowd computing is elasticity, we must provide crowd-based system with mechanism in order to cope with large problems as fast as possible. Because of this, parallelization becomes essential. There exist several paradigms in order to parallelize the execution of a task in a set of distributed computers. Among those MapReduce has become one of the most popular because of its simplicity and its capacity to scale. In [9] authors discuss MapReduce as an interesting alternative to parallelize tasks in a crowd-based system.

4 A Practical Industrial Example

For over ten years, CA Technologies has been developing and using machine translation technologies and tools to support software localization activities. Like most large software vendors, CA Technologies continuously improves its processes to reduce the cost of translation and increase the number of languages it supports. The most expensive and time-consuming phase of the localization process is the post-editing performed by human translators of the output produced by automatic machine translation, especially when this step has to be outsourced to external translation service providers.

In this section, we present a new semi-automatic management platform that allows the integration of crowd computing for reducing the cost of post-editing phase in software localization. Following the ideas presented in previous sections, in our system quality will be monitored and stored in individual records which will depend on the quality of the texts translated by each worker. Note that this problem fits the three prerequisites described at the beginning of this document: (i) machine translation does not deliver sufficient quality, (ii) translating a user guide or a user interface does not require complex training process and these training processes can be automated, and (iii) the information handled is only partially confidential.

4.1 Limitations of the Current Localization Process

Before describing the platform, we summarize the main motivations that lead to the proposal of a crowd-based system. The main limitations in the current localization process are:

- ***Long time-to-market periods for non-English versions of our products:*** products translated to languages different from English are usually released several months after the English version release because the localization process is time-consuming.
- ***Changing workload management:*** the translation workload is heterogeneous and there are some peaks during the year when a large number of products are released together. Hence, the localization teams cannot cope with the situation forcing the outsourcing of part of the work to external translation service providers.

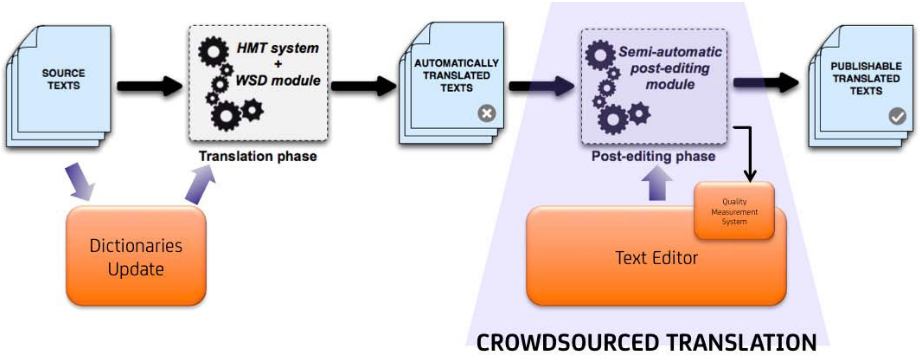

Fig. 3. Use of crowdsourcing for software localization

– ***High cost of extending our market to countries speaking languages that are currently not translated:*** in order to translate to a large number of languages, and open the possibilities for the company to explore new markets, the current approach does not work. Firstly, it is not easy to find translators for all languages. Secondly, it is economically expensive and thus unfeasible in general to hire a team of translators for every language, especially for emerging markets in some countries where the number of products sold is not expected to be huge.
– ***High cost of software localization for languages we currently translate:*** companies invest several million dollars in localization per year both in internal and outsourced localization.

4.2 Building a Crowdsourcing Platform for Software Localization

The objective of our platform is to overcome the above-mentioned issues. In Figure 3, we describe the usual process followed to localize a user guide. The source texts go through the machine translation engine and a first automatic translation is produced. Usually, the original and the machine-translated texts are sent to human translators that post-edit the text written in the target language. After this, the text is ready to be published. When the amount of work exceeds the capacity of the translators, the localization has to be outsourced. With our proposal, our goal is to crowdsource work instead of outsource. Our trustworthy crowdsourcing platform consists of:

– A model to divide a task into subtasks to be distributed among the crowd.
– A ranking function to evaluate the quality of the translations generated by a user.
– A model to organize tasks for its parallel execution based on the MapReduce philosophy.
– A quality-aware rewarding system to remunerate workers based on the quality of their work and other aspects that may range from their training to their location and native language.
– A quality-aware task sequence organization system that guarantees a minimum level of quality independently of the quality of translators, based on AV-Units.

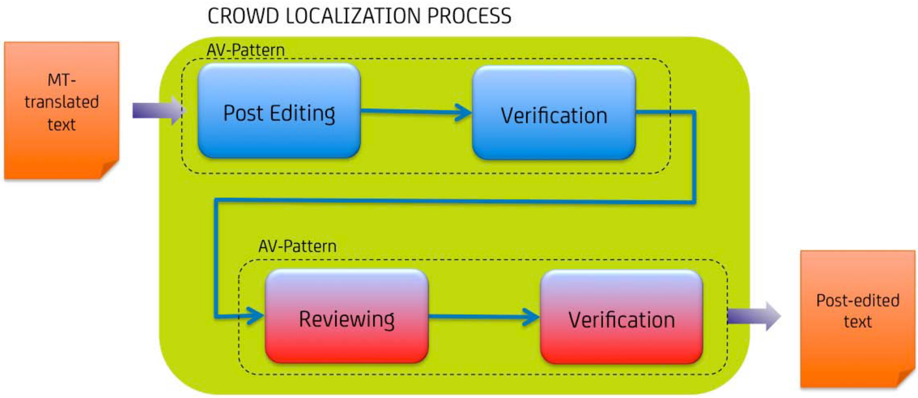

Fig. 4. Example of the use of AV-Units for the localization problems. Blue: bilingual. Red: monolingual.

We divide a task into smaller subtasks that can easily be solved by an individual worker. E-g a user guide containing roughly 100,000 words is divided into subtasks of 2000 words each. This is different from other existing approaches that divide tasks into sentence level or paragraph level subtasks thus losing the context of the text. Our selected subtask size is both convenient for a worker to solve in reasonable time and also preserves the context. Workers work on different jobs like post-editing, fixing errors, verifying, etc. We have developed a quality control system, which comprises two AV-Units to provide quality in the final translation. An important aspect of this system is that, since the quality of their work directly affects the reward obtained after finishing a task, this motivates workers to emphasize on quality. Also, the quality control algorithm adapts to the availability of workers. If high ranked workers are available, a small number of them are assigned to a task. However, if only low ranked workers are available, this number is increased.

The platform manages workers automatically. The software localization process is carried out by professional and non-professional from around the world, while the quality of translations is being maintained by the systems itself.

Quality and Scalability. In Figure 4, we show a possible division of the post-edition process into two AV-Units. In the first AV-Unit, the action consist in post-editing the text given the original text in English and the machine translation in the target language. The second AV-Unit is designed in order to review fluency and naturality in the target language of the translated text. The workers participating in the first unit are bilingual, and those participating in the second can be monolingual speakers of the target language.

Also, in order to achieve a large degree of scalability, we use a MapReduce approach. In the mapping phase, texts are run through the different phases described before. In the reduce phase, post-edited text are merged back to a single document. Figure 5, describes this process visually.

Advantages of Using Crowdsourcing in This Example. Besides the obvious benefits of crowdsourcing, namely scalability, elasticity, etc., the use of this type of crowdsourced platform reduces the cost of the software localization process significantly. It

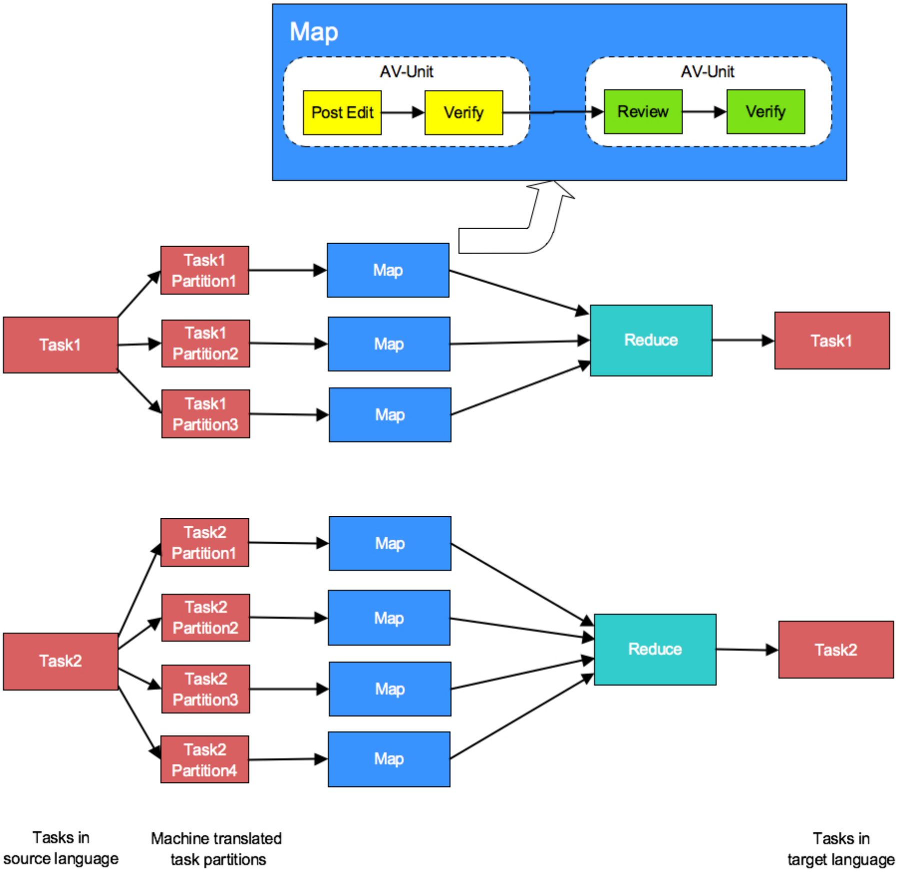

Fig. 5. Crowdsourced tasks are organized following a MapReduce strategy

solves the problem of finding translators for less popular languages because it allows for the participation of any remote translator around the world, and it improves quality. Specifically:

- ***We leverage the capacity provided by the crowd to interact one-to-one with translators, increasing quality and agility.*** Third-party systems are blackboxes, even if they are based on crowdsourcing strategies. When we work with localization service providers, quality issues usually arise. Many times this is due to the fact that translators are not familiarized with our products or even with the IT domain. When this happens, we have to send back translations to the vendor and this is very time-consuming. With this platform, we are able to make the process much more agile, since workers are exclusively trained through our products and documents and we can resubmit work directly to them, if necessary.
- ***We eliminate middleman costs.*** Existing translation service providers get a margin for their services.
- ***We have full control of all the real costs, specifically of the per-word rate paid to translators.*** Due to the socioeconomic situation of the countries speaking a certain

target language, it is a common practice that the per-word rate varies from language to language. For example, the per-word cost for a translation into German is higher than the per-word rate for a translation into Russian.

– ***We gain control in terms of confidentiality.*** Since we are sending our information to the network, security is one of the issues of crowdsourcing. By using the proposed platform, it is possible to establish your own security measures. For instance, we decide how to partition the documents and who we send information to. On the contrary, with third-party services, we rely on security measures that are not under our control, but information is still sent to the people distributed around the world.

By using this system costs are reduced, CA Technologies will gain immediate capacity to translate to any language in the world and time-to-market of CA Technology products will be significantly reduced.

5 Conclusions and Future Work

It is time for industry to start thinking about real-time and real-world crowd computing. Managing human beings automatically and helping them to collaborate with parallel machine processes will be a way to reduce costs and a competitive advantage for all those companies adopting the power of the crowd. In our dynamic world, elasticity becomes essential, human intervention is unavoidable and quality a key requirement. These three components converge to make crowdsourcing one of the most promising ideas to leverage the power of social networks. Nevertheless, new ethical issues arise. In existing systems workers are vulnerable to the whims of employers. Therefore, new legislations will have to be developed in order to create a fair work marketplace.

In the near future, new algorithms will have to be devised and more sophisticated ranking systems will be designed in order to improve the quality provided by the crowd. Several aspects such as confidentiality preservation in crowd computing systems will also become essential. In the specific area of localization, new methods have to be devised in order to preserve the style coherence on large documents.

Acknowledgements. We would like to thank Invest in Spain society, the Ministerio de Economa y Competitividad of Spain and the EU through FEDER funds for his support through grant C2_10_18.

References

1. Banerjee, S., Lavie, A.: METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In: Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization, pp. 65–72. Association for Computational Linguistics, Ann Arbor (2005), http://www.aclweb.org/anthology/W/W05/W05-0909
2. Bentivogli, L., Federico, M., Moretti, G., Paul, M.: Getting expert quality from the crowd for machine translation evaluation. In: MT Summit XIII, pp. 521–528 (2011)
3. Bernstein, M.S., Little, G., Miller, R.C., Hartmann, B., Ackerman, M.S., Karger, D.R., Crowell, D., Panovich, K.: Soylent: a word processor with a crowd inside. In: Procs. of the 23nd Annual ACM Symposium on User Interface Software and Technology, UIST 2010, New York, USA, pp. 313–322 (2010), http://dx.doi.org/10.1145/1866029.1866078

4. Denkowski, M., Lavie, A.: Exploring normalization techniques for human judgments of machine translation adequacy collected using amazon mechanical turk. In: Procs of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk, CSLDAMT 2010, pp. 57–61. Association for Computational Linguistics, Stroudsburg (2010), http://dl.acm.org/citation.cfm?id=1866696.1866705
5. Gao, Q., Vogel, S.: Consensus versus expertise: a case study of word alignment with mechanical turk. In: Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk, CSLDAMT 2010, pp. 30–34. Association for Computational Linguistics, Stroudsburg (2010), http://dl.acm.org/citation.cfm?id=1866696.1866700
6. Geiger, D., Seedorf, S., Schulze, T., Nickerson, R.C., Schader, M.: Managing the crowd: Towards a taxonomy of crowdsourcing processes. In: AMCIS (2011)
7. Harris, C.: You're Hired! An Examination of Crowdsourcing Incentive Models in Human Resource Tasks. In: Lease, M., Carvalho, V., Yilmaz, E. (eds.) 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), Hong Kong, China, pp. 15–18 (February 2011)
8. Howe, J.: Wired 14.06: The Rise of Crowdsourcing, http://www.wired.com/wired/archive/14.06/crowds.html
9. Kittur, A., Smus, B., Khamkar, S., Kraut, R.E.: Crowdforge: crowdsourcing complex work. In: Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, UIST 2011, pp. 43–52. ACM, New York (2011), http://doi.acm.org/10.1145/2047196.2047202
10. Lease, M., Yilmaz, E.: Crowdsourcing for information retrieval. SIGIR Forum 45(2), 66–75 (2012), http://doi.acm.org/10.1145/2093346.2093356
11. Negri, M., Bentivogli, L., Mehdad, Y., Giampiccolo, D., Marchetti, A.: Divide and conquer: crowdsourcing the creation of cross-lingual textual entailment corpora. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP 2011, pp. 670–679. Association for Computational Linguistics, Stroudsburg (2011), http://dl.acm.org/citation.cfm?id=2145432.2145510
12. Papineni, K., Roukos, S., Ward, T., Zhu, W.J.: Bleu: a method for automatic evaluation of machine translation. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, ACL 2002, Stroudsburg, PA, USA, pp. 311–318 (2002), http://dx.doi.org/10.3115/1073083.1073135
13. Silberman, M.S., Irani, L., Ross, J.: Ethics and tactics of professional crowdwork. XRDS 17(2), 39–43 (2010), http://doi.acm.org/10.1145/1869086.1869100
14. Yan, T., Kumar, V., Ganesan, D.: CrowdSearch: exploiting crowds for accurate. In: Intl. Conf. on Mobile Systems, Applications, and Services, pp. 77–90. ACM, New York (2010), http://dx.doi.org/10.1145/1814433.1814443
15. Yuen, M.C., King, I., Leung, K.S.: A Survey of Crowdsourcing Systems. In: Proceedings of the IEEE Third International Conference on Social Computing (SocialCom), pp. 766–773. IEEE (October 2011), http://dx.doi.org/10.1109/PASSAT/SocialCom.2011.203
16. Zaidan, O.F., Callison-Burch, C.: Crowdsourcing Translation: Professional Quality from Non-Professionals. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, Oregon, USA, pp. 1220–1229 (June 2011), http://www.aclweb.org/anthology/P11-1122.pdf