Nichesourcing: Harnessing the Power of Crowds of Experts

Victor de Boer[1], Michiel Hildebrand[1], Lora Aroyo[1], Pieter De Leenheer[1,2], Chris Dijkshoorn[1], Binyam Tesfa[1], and Guus Schreiber[1]

[1] The Network Institute, Dep. of Computer Science, VU University, Amsterdam, The Netherlands
{v.de.boer,m.hildebrand,l.m.aroyo, p.g.m.de.leenheer,c.r.dijkshoorn,guus.schreiber}@vu.nl, b.m.tesfa@student.vu.nl
[2] Collibra nv/sa, Brussels 12, Belgium

Abstract. In this position paper we identify nichesourcing, a specific form of human-based computation that harnesses the computational efforts from niche groups rather than the "faceless crowd". We claim that nichesourcing combine the strengths of the crowd with those of professionals, optimizing the result of human-based computation for certain tasks. We illustrate our claim using scenarios in two domains: cultural heritage and regreening in Africa. The contribution of this paper is to provide a definition of the main characteristics of nichesourcing as a natural extension of crowdsourcing and to outline research challenges for realizing nichesourcing applications.

1 Introduction

In 2005 Luis Von Ahn coined *"a paradigm for utilizing human processing power to solve problems that computers cannot yet solve"* [1]. Since then it has grown into a research field called *human-based computation* applied successfully in various domains for a great number of tasks that rely on the advantage that humans have over machines in skills such as visual recognition and human communication [2]. A popular and well-described form of human-based computation is *crowdsourcing* where the computational task is delegated to a *crowd*, i.e. a very large and redundant workforce of people. Digitization, classification, translation and annotation are usually split into many simple tasks, that can be performed by anonymous people without any specific skills. The crowd is gathered via a more or less open call [3] in a (social) network [4]. In some cases, such as the Amazon's Mechanical Turk, participants are paid small fees for their effort [5]. However, as most initiatives do not provide financial reward, other motivational incentives are deployed in order to attract a sufficient amount of people [6].

Many institutions and companies are currently utilizing the knowledge of the crowd as an alternative to professional efforts. Initially, these crowdsourcing initiatives were centered around users performing *simple tasks* and their overall target was geared towards *achieving quantity* rather than quality. Currently, we observe [7] that there is (i) a growing demand for solving also *complex knowledge-intensive tasks* [8], and (ii) a natural expectation to focus on the quality of the final result [9]. In this paper, we argue that *nichesourcing* is the next natural step in the evolution of crowdsourcing to address

A. ten Teije et al. (Eds.): EKAW 2012, LNAI 7603, pp. 16–20, 2012.

those two demands. Nichesourcing is a specific type of crowdsourcing where complex tasks are distributed amongst a small crowd of amateur experts (for example art enthusiast or African ex-pats) rather than the "faceless" crowd. A niche is gathered from either distributed experts on a specific topic or from an existing network centered around the same culture, location or topic. In both cases the members have domain knowledge and an intrinsic motivation to contribute and provide high quality results [10].

The contribution of this paper is to provide a definition of the main characteristics of nichesourcing as a natural extension of crowdsourcing (Sec. 2), and to identify the research challenges for its realization (Sec. 4). We describe two use cases for which we are currently implementing nichesourcing solutions (Sec. 3).

2 Harnessing the Crowd and the Niche: Comparison

In this section, we provide a basic framework to compare three main aspects of crowdsourcing and nichesourcing. We identify (from a process-centric perspective) three main dimensions in this framework, namely *the task* and its complexity, *the product* that is targeted as a result of the task, and *the resource pool* of users needed in order to realize this task. The purpose of this comparison is to explain the fundamental differences between crowdsourcing and nichesourcing, and to identify for what types of tasks, for what desired result and for what type of "crowd" each of them would be most suitable. This framework supports making an informed decision whether to apply crowdsourcing or nichesourcing for a specific human computation problem.

The Atomic Task: Simple vs. Knowledge-intensive. Crowdsourcing deals with large complex tasks by dividing it up into smaller, *atomic tasks*, the latter which do not require specific knowledge or skills from the crowd members. Consider, e.g., different peers tagging the same image for triangulation purposes. Crowdsourcing typically focuses on repeatable atomic tasks and can be pitched easily through broad syndication channels to random pools of resources. For some complex tasks, however, the atomic tasks are too hard to execute for any random crowd member for two reasons: either (i) the atomic task itself requires specific background knowledge that cannot be assumed to be present in the crowd; or (ii) it will take so much effort that crowdsourcing motivation techniques are not sufficient to engage a significantly sized crowd. Atomic tasks that require a specific level of knowledge or effort can be outsourced to niches in which the members possess particular knowledge or motivation is present.

Product: Quantity versus Quality. The success of the crowdsourced product is determined by quantity. As there is no specific qualification for the targeted resource pool, the quality of the product can only be expressed (i) either globally (e.g., accuracy of the results) or locally (e.g., specificity of individual tags). Hence, typical crowdsourcing solutions require careful application design as well as post-processing of the produced results. For example, by awarding users producing good results. Post-processing typically involves aggregating the results and exploiting redundancy to determine the most likely correct and/or most specific results. Complex nichesourcing tasks strategically target socially-trusted *communities of practice* [11] (see further). The rationale for such strategy is that an acceptable number of appropriately skilled resources are assumed to have the intrinsic courtesy to provide higher quality individual results. This removes much of the need for statistical processing. Secondly, experts typically provide more specific input than non-experts. If the task requires a more controllable level of quality, nichesourcing is a good alternative. This gain in quality control comes at a cost of quantity, since we assume a community of practice to be smaller than any subset of Web users that can be considered a crowd.

Resource Pool: Crowd vs. Community of Practice. Quinn et al. [2] identify five methods for motivation in human-based computation applications: *payment, altruism, enjoyment, reputation* and *implicit work*. Building and maintaining a dedicated crowd is essential to crowdsourcing applications that use altruism, reputation and (to a lesser extent) enjoyment as motivation. A key measure of success is the size of this crowd, and to a certain extent, the level of redundancy (necessary for statistical processing). Such a crowd is usually anonymous and heterogenous, i.e., there is no shared goal or any social affinity required whatsoever. In nichesourcing, composite and complex tasks are distributed within existing communities. Communities, in contrast to crowds, have a common purpose, peers have an identity and affinity with this purpose, and their regular interactions engenders social trust and reputation. These niches can correspond to the notion of a community of practice or interest[11]. Although communities, in contrast to crowds, provide smaller pools to draw resources from, their specific richness in skill is suited for the complex tasks with high-quality product expectations found in nichesourcing. Moreover, the peer resources receptive to complex tasks may exploit their own social trust relations to transitively trigger other communities that may offer reinforcing resource pools.

3 Two Use Cases for Nichesourcing

Rijksmuseum Prints Annotation. The Print Room of the Rijksmuseum in Amsterdam has a collection of about 700 000 prints, drawings and photographs. Within the project *Print Room Online* they register the basic properties of each print, such as the object ID, storage location, title, creator and measurements. In addition, they describe the subject matter of the prints. The annotation is performed by eleven professional cataloguers which are expected to describe 20% of the collection during the 3 years of the project.

Clearly, there is a need for more human effort in order to describe the entire collection of the Print Room. Crowdsourcing seems like the natural solution to the problem. Indeed, many museums and cultural heritage archives have already embraced the notion of human-computing in order to solve the same problem. Typically, they apply crowdsourcing techniques [7], [12]. As a result they receive large quantities of metadata, but not necessarily of sufficient quality [13]. For the Rijksmuseum, the quality of the produced annotations is a prime concern. They are interested in highly domain specific annotations, such as the names of the people, places and events depicted on the prints, and they prefer specific annotations of the objects and concepts, e.g. “the symbolic meaning of a frog in a Japanese print”, over generic annotations, e.g. “frog”. Therefore, the “faceless crowd” is not the right target and instead the Rijksmuseum is in need of hobbyists, self-thought experts and retired professionals that can perform this knowledge-intensive task.

Digitizing Pluvial Data from the Sahel. Governments in the African Sahel region have recorded 1000s of pages of data about rainfall and crop harvest in their region. This data is very useful when aggregated over multiple regions for analysis supporting decisions in re-greening initiatives. For this goal, low resolution digital scans have been made of handwritten documents containing tabular as well as textual data[1]. The data in these documents is to be converted into digitized structured data. Automated techniques digitization are error-prone and require a lot of configuration. Although crowdsourcing has been used for digitizing handwritten documents (e.g. [14]), this specific task is fairly complex in the sense that (i) the semantics of the tables is often not easily understood; and (ii) decoding the handwriting does require specific language and domain knowledge (e.g., familiarity with the regional geography and villages). We expect that the level of quality that the faceless crowd can provide is not sufficient for the goals of the re-greening initiative and that therefore this task is very well suited for nichesourcing. The niche being targeted is the so-called *African Diaspora*: African expatriates who now reside in better connected parts of the world. Members of the diaspora are very much affiliated with local issues in their region of origin. This intrinsic motivation can be exploited through nichesourcing. Existing Web communities set up by members of the Diaspora (e.g., on Facebook) can be addressed. The network connections can be used to distribute human computation tasks as well as reinforce motivation through reputation. Furthermore, the domain knowledge of the niche members (including the local language and names of villages) may guarantee to produce a higher level of quality than which could be obtained by a general crowd.

4 Nichesourcing Challenges

To achieve a systematic, sustainable and efficient nichesourcing process, an institution needs to (1) employ mechanisms to actively engage and support the experts in the task; and (2) define quality measures for both individual results and the overall production.

Task Distribution. Finding the appropriate niche containing people that are most suited to perform complex tasks is not straightforward. This niche identification and matching challenge requires concise descriptions of the task itself as well as descriptions of the type and level of expertise of the niche required for the task. Additionally, the individual tasks that need be performed might require specific types of domain knowledge. For example, in the Rijksmuseum use case the annotation of one print might require knowledge about specific types of castles while others might require knowledge about historic political issues. The research challenge here is to match tasks with the most appropriate experts within a niche. We can benefit from the extensive research done in the field of *expert finding* to automate this process [15]. In existing communitiesocial connections can be exploited to (re-)distribute tasks among niche members.

Quality Assurance. Although in crowdsourcing reputation plays a role in quality control, for complex tasks in nichesourcing the reputation within the social network of contributing peers is key. *Trust* emerges from an history of collaboration on other tasks, and may be based on similarly evolving interests of the peers. These parameters are

[1] Samples can be viewed at `http://www.few.vu.nl/~vbr240/pluvialdata/`

difficult to measure because this data is distributed across different platforms peers use to work and collaborate.

Even though expert contributions can be expected to be of a higher quality than those produced by a crowd, it is most likely necessary to be able to identify the quality of the individual contributions of members. Where complex tasks are performed within peer networks, quality measurements should explore the social roles, responsibilities and interactions that led to the overall result. When a task is published, there are certain product quality expectations attached to its description. As discussed in Sect. 2, for complex tasks this is not straightforward. However, a poor description may impede the receptivity by the communities. To this end, we could learn from service level agreements which provide frameworks for. e.g., outsourcing, to define and agree on complicated considerations of expected quality of a results.

References

1. von Ahn, L.: Human Computation. PhD thesis, School of Computer Science. Carnegie Mellon University (2005)
2. Quinn, A.J., Bederson, B.B.: Human computation: a survey and taxonomy of a growing field. In: Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems, CHI 2011, pp. 1403–1412. ACM, New York (2011)
3. Howe, J.: The rise of crowdsourcing. Wired Magazine 14(6) (June 2006)
4. Hand, E.: Citizen science: People power. Nature (2010)
5. Huber, A.J., Lenz, G.A., Michael, G.S., Alvarez, R.E.: Evaluating online labor markets for experimental research: Amazon.com's mechanical turk. Political Analysis (2012)
6. Raddick, J., Szalay, A.S., Vandenberg, J., Bracey, G., Gay, P.L., Lintott, C.J., Murray, P., Schawinski, K.: Galaxy zoo: Exploring the motivations of citizen science volunteers. Astronomy Education Review 9(1) (2010)
7. Oomen, J., Aroyo, L.: Crowdsourcing in the cultural heritage domain: opportunities and challenges. In: Proceedings of the 5th International Conference on Communities and Technologies, C&T 2011, pp. 138–149. ACM, New York (2011)
8. Oded Nov, O.A., Anderson, D.: Dusting for science: motivation and participation of digital citizen science volunteers. In: Proceedings of the 2011 iConference, iConference 2011 (2011)
9. Ipeirotis, P.G., Provost, F., Wang, J.: Quality management on amazon mechanical turk. In: Proceedings of the ACM SIGKDD Workshop on Human Computation, HCOMP 2010, pp. 64–67. ACM, New York (2010)
10. Schroer, J., Hertel, G.: Voluntary engagement in an open web based encyclopedia: Wikipedians, and why they do it. Media Psychology 12, 1–25 (2009)
11. Wenger, E., Richard, McDermott, W.S.: Cultivating communities of practice: a guide to managing knowledge. Harvard Business School Press, Cambridge (2002)
12. Kalfatovic, M., Kapsalis, E., Spiess, K., Van Camp, A., Edson, M.: Smithsonian team flickr: a library, archives, and museums collaboration in web 2.0 space. Archival Science 8, 267–277 (2008), doi: 10.1007/s10502-009-9089-y
13. Gligorov, R., Hildebrand, M., van Ossenbruggen, J., Schreiber, G., Aroyo, L.: On the role of user-generated metadata in audio visual collections. In: K-CAP, pp. 145–152 (2011)
14. Holley, R.: Many Hands Make Light Work: Public Collaborative OCR Text Correction in Australian Historic Newspapers. National Library of Australia (2009)
15. Balog, K., Azzopardi, L., de Rijke, M.: Formal models for expert finding in enterprise corpora. In: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2006, pp. 43–50. ACM, New York (2006)