Skill ontology-based model for Quality Assurance in Crowdsourcing

Kinda El Maarry[1], Wolf-Tilo Balke[1], Hyunsouk Cho[2],
Seung-won Hwang[2], Yukino Baba[3]

[1]Institut für Informationssysteme, TU Braunschweig
{elmaarry, balke}@ifis.cs.tu-bs.de
[2]Department of Computer Science and Engineering, POSTECH, Korea
{prory, swhwang}@postech.ac.kr
[3]The University of Tokyo, Japan
yukino_baba@mist.i.u-tokyo.ac.jp

Abstract.

Crowdsourcing continues to gain more momentum as its potential becomes more recognized. Nevertheless, the associated quality aspect remains a valid concern, which introduces uncertainty in the results obtained from the crowd. We identify the different aspects that dynamically affect the overall quality of a crowdsourcing task. Accordingly, we propose a skill ontology-based model that caters for these aspects, as a management technique to be adopted by crowdsourcing platforms. The model maintains a dynamically evolving ontology of skills, with libraries of standardized and personalized assessments for awarding workers skills. Aligning a worker's set of skills to that required by a task, boosts the ultimate resulting quality. We visualize the model's components and workflow, and consider how to guard it against malicious or unqualified workers, whose responses introduce this uncertainty and degrade the overall quality.

Keywords: crowdsourcing, quality assurance, skill ontology, uncertain data.

1 Introduction

The hope of being able to somehow benefit from "the wisdom of the crowd" [1] is the main driver for the rising popularity of crowdsourcing [2], coupled with the information flood and the flexible and relatively cheap solution that today's crowdsourcing platforms offer. Thus, more and more companies and organizations are turning to crowdsourcing. Some notable names include: NASA, Threadless, iStockphoto, InnoCentive, etc [3]. Yet, the question of automatically assuring the returned quality of results [4] and the uncertainty that is associated with it, remains an unsolved question [5]. This is because checking every single submitted response is costly, time consuming and threatens to invalidate most of the crowdsourcing gains. This in turn encourages unethical workers to submit low quality results.

Most requestors end up relying on redundancy or repeated labeling as means of verification of user performance. A common approach tests the reliability of users by blending a set of questions for which the answers are known, so-called gold questions, into the workload. This instantly raises the question of how many gold questions should be included [6]. Another possibility is the assignment of multiple workers to the same task and then aggregating their responses. While both approaches pose problems for tasks where the comparison between individual workers' results is difficult (see [7] for a detailed discussion), the redundancy approach usually incurs monetary costs and places the costs for quality control at the task provider's doorstep. Moreover, popular techniques for aggregation like e.g., majority voting, have been shown to suffer from severe limitations [8].

To correctly tackle the issue of uncertainty, various factors affecting the quality of the responses were investigated [9]. While results show monetary incentives having an effect on quality in contrast to the experimental results in [10], it's still rather tricky; low paid jobs yield sloppy work, and highly paid jobs attract unethical workers. Another investigated factor was workers' qualification, where not only was it shown that qualified workers produce better quality and strive to maintain their qualification level, but in a setup that relies on qualifications for task assignments, unqualified workers are pushed to diligently work on improving their own qualifications.

To that end, we investigate a skill ontology-based model to be adopted by crowdsourcing platforms, which aims at identifying those qualified workers and assigning them to the tasks they're eligible to. This can be realized through identifying the skills required to adequately work on a task, and aligning it to the skills a certain worker has. Consequently, by excluding non-qualified workers or non-ethical workers who falsely try to build up their qualifications, the model would be practically excluding the sources of uncertainty introduced to the data altogether.

The rest of the paper is organized as follows. We start off by reviewing the current related work. Next, we define what quality stands for in a crowdsourcing setup and identify the different types of quality that our model needs to realize. Section 4 presents in details the proposed skill-ontology model. This is followed in section 5 by an overview of the model's workflow. Finally, in the last section, we provide a summary and an outlook on future work.

2   Related Work

In recent years, many web-based collaboration platforms and marketplaces are relying on that same "wisdom of the crowd" ideolody, where anonymous users' contributions are in some way combined to provide innovative and diverse services. Threadless (online t-shirt design contest) [11] and istockPhoto, are two prominent examples exploiting that ideology. [12] presents an audio document retrieval service "Pod-Castle", which collects anonymous transcriptions of podcast speech data to train an acoustic model. This was followed two years later by an alternative crowdsourcing-based approach [13]. These examples support the main argument in [14], i.e. that the way people collaborate and interact on the web has been so far poorly leveraged through the existing service-oriented computing architectures.

So instead, a mixed service-oriented system i.e. service-oriented crowdsourcing, is desirable, enabling a more seamless approach, which would also exploit the on-demand allocation of flexible workforces. This steers the trend ever more towards crowdsourcing, which is now being offered by many platforms: Amazon Mechanical Turk, Samasource, Crowdflower, etc. However, every chance needs to overcome challenges, and the main challenge here is that crowdsourcing results are often questionable in terms of their quality, and the associated uncertainty introduced in aggregated results becomes an issue.

This is actually very similar to the missing confidence in third-party services which posed serious issues in the web services community, see e.g., [15]. One solution here was to adopt credentials proving the eligibility of each discovered service. Simply put, a service is eligible if it meets certain quality requirements (in functionality, as well as typical QoS parameters like availability or response time). When composing complex workflows out of individual services these quality requirements can be interpreted as mutual agreements. Such agreements can be expressed for example by Web Service Level Agreement Language (WSLA) [16], or Web Service Management Language (WSML) [17]. In our context, a service provider is none other than a worker who has some skills and a task provider's confidence in results would be based upon the worker's credentialed skills. These credentials can be attained by passing a standardized test or a personalized test that the provider designs for that particular task. An agreement is reached, when a worker's credentialed skills matches those listed by the task provider (requestor) as the exact skills required for the corresponding task.

A lot of work in crowdsourcing literature has already been devoted to mitigate such quality concerns. The solution of redundancy and repeated labeling was first expanded by Dawid and Skene [18], who took into consideration the response's quality based on the workers. Through applying an expectation maximization algorithm, the overall error rate for each worker can be computed. Other approaches that estimate these error rates includes: a Bayesian version of the expectation maximization algorithm approach [19] and a probabilistic approach that takes into account both the worker's skill and the difficulty of the task at hand [20]. A further step was taken in [21] with an algorithm separating the unrecoverable error rates from recoverable bias.

Here, rather than looking at the worker's error rates, we aim at identifying the workers who are a good match for the corresponding task. Each worker has a skill profile, and every skill in the ontology is associated with a library of assessments. These assessments validate whether a worker indeed possesses the necessary skill or not. Both can be viewed analogously to the competence profiles provided by learning objects – entities that are used for task-focused training or learning in the IEEE 1484.12.1 – 2002 Standard for Learning Object Metadata.

These skills can be managed and referred to in a skill ontology. This fortunately leads us to a rich literature to derive and adapt from, which has been devoted to building competencies models, see [22] and [23]. Competency covers: knowledge, experience, skill and willingness to achieve a task. Such models have been used for quite a long time in organizations to help identify and attract suitable workers, as well as to help the workers acquire the needed skills. In order to identify the skills required for a task, skill gap analysis can be used to create the task's corresponding competency map [24]. Workers having the corresponding skills in a task's competency map could be then identified through competency matching. [25] Formalizes another approach

that focuses on Ontology-based semantic matchmaking between demanded skills (skills required by a task in our case) and supply (the workers possessing that skill). However, competency models still have their own challenges. Given their complexity, competencies have to be precisely defined within the different specific domains. Moreover, developing assessments that can truly capture one worker's competency level is unfortunately very often underestimated [26].

But of course, assigning the right worker for a task involves much more than just choosing the workers based on their skills. A worker maybe be highly competent relative to the task he's assigned to, yet his work ethics may earn him a bad reputation This might simply boil down to wanting to finish a task as fast as possible and with the least effort incurred. So the overall quality is in fact affected by both the workers' skills and reputation. This elicits the need for deploying quality control measures, whether in design time, run time or both, see [27] for a more comprehensive list of these measures. Computing workers' reputations poses a real challenge, and many reputation approaches have been investigated whether it's based on a reputation model [28], on feedback and overall satisfaction [29], or on deterministic approaches [30], etc.

3 Types of Quality

Upon addressing the data uncertainty that arise with crowdsourcing tasks, different aspects of quality can be identified. This breakdown allows us to identify the corresponding quality assurance mechanisms, which needs to be addressed by the proposed model. A detailed description of each of those quality aspects follows next.

3.1 Result's quality

Comes first to mind, and covers both the requester's expectations and the usefulness of the results. In terms of requestor's expectations, the structure of the returned results will be heavily influenced by what the requester wants and expects. Accordingly, the crowdsourcing task should be designed in a way that elicit that specific structure in the returned results (factual correctness in the form of a yes/no answer, consensus, opinion diversity, opinion quality, etc.). In terms of usefulness, requestors may also measure the quality in terms of how the results are consistent or abiding to the task description, or whether they are transparent and traceable .i.e. there exists a logical pattern the worker followed to give that response.

3.2 Platform's quality

Refers to the usability of the platform, where a platform's interface and offered tools should equally support both workers and requesters. For workers, the platform should promote a fair working environment. Fairness encompasses: 1) guaranteed payments, 2) nondiscriminatory conduct, 3) payments matching the corresponding load of work. For the requesters, the platform should offer an adequate set of tools to easily and efficiently: 1) upload data and download results, 2) design tasks, 3) automatically assign qualified workers, 4) block spammers, 5) train workers.

3.3 Task's quality

At a lower granularity, the quality of the task directly affects the results' quality. A requestor should: 1) identify the set of skills required to accomplish a task, 2) describe the task clearly, 3) define the expected effort in terms of complexity or time required to finish, 4) design the task's interface to support an easier workflow for the worker.

3.4 Worker's quality

Refers to how fit a work is for the task at hand. Namely, how qualified and prepared they are to do the task. On one hand, qualified can be mapped to skill levels and how relevant these skills are to the task. On the other hand, prepared can be translated into willingness to complete the task to the best of ones skills. Other contributing factors are: 1) workers' availability, 2) flexibility of working hours (Both can be easily monitored through activity logs), 3) workers' reputation. (Can be based on history and average satisfaction score attained upon the completion of a task).

These different aspects will often in reality be interleaved. For instance, the clarity of a task is not only related to a task's quality, but might also fall under the platform's quality, where a platform ensures that the workers get clear task description that helps them avoid getting penalized if they do the task incorrectly due to vague guidelines.

4 Skill Ontology-based model

Following the quality aspects we identified in section 3, we propose a skill ontology-based model to be adopted by crowdsourcing platforms. The model aims to capture the different aspects of quality that helps diminish the resulting uncertainty by eliminating one of its major sources: unqualified workers. The skill ontology-based model roughly comprises of: 1) skill ontology, 2) ontology merger 3) skill's library of assessments, 3) Skill aligner, 4) reputation system and a 5) task assigner.

4.1 Basic and temporary skill ontologies

At the model's core lies the skill ontology. The model maintains a dynamic ontology, which evolves with the crowdsourcing platform's demands. While some skills will be often required for many tasks e.g. language skills for translating tasks, other skills will be highly specific and tailored for a specific task e.g. identifying the family, genus and species a fish belongs to. Accordingly, two ontologies are maintained: a basic and a temporary one. The basic skill ontology retains those skills that are highly demanded by many tasks. The temporary skill ontology retains newly added skills. Later on, only those skills that were frequently required by many tasks are transferred from the temporary ontology to the basic one.

A requestor is always presented with a single consolidated ontology (or an automatically generated taxonomy as presented in subsection 4.4), in which he can browse the skills required for the task he's designing. When the required skill isn't available in the ontology, the requestor can define a new skill.

4.2 Ontology merger

A new skill, which has been newly defined by a requestor is initially added to the temporary skill ontology. Every defined skill must be associated with at least one assessment. The new skill resides in the temporary ontology until it: 1) has proven to be popular 2) has been verified. Popular skills are skills that were required not only by many tasks, but also by many different requestors. Verified skills are skills that are associated with at least one verified assessment as will be further explained next.

4.3 Skill's library of assessments

Identifying whether a worker has a certain skill or not, can be ascertained through an assessment. A skill's library of assessments may comprise two types of assessments: standardized and personalized assessments. For standardized assessments like: TOEFL for the English language, or MOOC (Massive Open Online Course) certificates, most requestors will approve and conclusively trust them. However, when a requestor doesn't, or when there is simply no standardized test for the required skill, the requestor can create a personalized assessment. We assume here that requestors, who´re requiring more specific skills or who´re not satisfied with the available assessments, are willing to invest time to enforce higher quality as per their own standards.

Standardized assessments are inherently verified, since their legitimacy are already proven. On the other hand, personalized assessments, require further investigation for verification. Consider the following scenario: A worker posing as a requestor, creates a new personalized assessment and uploads it for the skill he\she wants to attain. Providing the perfect answers for these personalized assessments becomes then trivial, and the worker can accumulate endless skills in this manner. Assessments' verification can be done manually or automatically (platform-wise or crowd-wise).

1. Manual verification: entails hiring an expert to look over the assessment, this however costs both time and money. Accordingly, as a rule of thumb, this should be limited to cases where a popular skill has only one personalized assessment or multiple personalized assessments from the same requestor.
2. Automatic verification: serves as an alternative to manual verification, when the skill has: at least one verified personalized assessment, or one standardized assessment.

   - Automatic platform-wise verification: The platform creates a new personalized assessment, merging the original questions with those from different verified assessments available in the corresponding skill's library of assessments. If workers can also answer the newly merged questions, the assessment is verified and can be later on used on its own.
   - Automatic crowd-wise verification: Workers who have the corresponding skill in their skill profile, can verify the assessment and earn a higher reputation. Note that, extra measures need to be taken, to avoid workers who maliciously aim at boosting their reputation by creating spam assessments and reporting them later as spam. Accordingly, unlike the task assignment, the workers are automatically assigned a random assessment, rather than choosing one.

Until a personalized assessment is verified, workers are allowed to take. If the workers suspect the assessment to be a spam, they must report it. If not, they may take it and acquire a pending-verification skill in their profile upon passing the assessment. When the assessment is verified, all the corresponding pending-verifications skills are updated. If the assessment was merely spam, workers' who failed to report the assessment as such are penalized, and the corresponding pending-verification skill is revoked.

4.4   Skill aligner

Upon creating a task with a set of prerequisite skills, a requestor can choose to either use one of the available skills in the ontology or define a new one. Choosing one of the skills in the ontology can be a tiresome job, especially since the ontology grows with the needs of the crowdsourcing platform. Ideally, a requestor should be able to quickly see whether the required skill is available in the ontology or not. To that end, a taxonomy can be maintained on top of the ontology, which the requester can quickly traverse. This taxonomy can be automatically built from the skill description and keywords the requestor inputs upon adding the new skill [31], and validated by the crowd. Given the set of prerequisite skills that the requester specifies, the skill aligner should be able align those skills to the similar available skills in the taxonomy and present those to the requestor. This could be based on matching the skill keywords and descriptions.

4.5   Reputation system

To ensure high quality, only qualified workers should be assigned to the corresponding task. Qualified workers are those workers who: 1) have the required skills, 2) are willing and motivated to complete the tasks, 3) are available, and 4) are highly reputable. Each of those can be respectively measured as follows.

1. **Skill profile:** The skill profile holds the worker's list of skills. The profile acts as a primary filter, where only those workers having a task's prerequisite set of skills are considered. A Skill profile can hold two types of skills: 1) verified skills, 2) pending-verification skills. For every skill in the worker's profile, a list of all the tasks that the worker utilized the corresponding skill in are compiled. Furthermore, an accompanying score is attached, reflecting this experience. This score can be derived from the compiled lists of tasks. Only completed tasks with a positive feedback are listed i.e. requestor was satisfied with the worker. Completed tasks with a negative feedback, are only reflected in the skill's score. This gives a chance for the worker to improve his skill, without having a permanent black spot in their skill profile.
2. **Willingness:** A worker's willingness can be captured from his crowdsourcing platform activity, the following can be observed: 1) time needed to finish a job versus that set by the requestor as the optimal processing time for the task to be done, 2) ratio of completed to aborted tasks.
3. **Availability:** A worker's availability can also be captured from the worker's activity log on the crowdsourcing platform, by specifically noting the number of hours

the worker logs per day or month. The time zone a worker is in plays an important role, for urgent tasks i.e. assigning workers with different time zones to the requestor's saves time, where requested tasks can be simply finished overnight.

4. **Reputation:** A worker's reputation can be derived from the average requestor's satisfaction. Moreover, the worker's reputation is penalized, when a pending-verification skill proved to be spam. In addition to such a penalizing system, a reward system can also be in place e.g. Workers contributing in the automatic crowd-wise verification of assessments. To promote fairness and protect the workers from malicious requestors, a similar reputation could be built for the requestors, even if it´s spanning only over one job. A malicious requestor will tend to be malicious in general, which can be easily tracked. In such a case, the requestor´s feedback can be ignored.

4.6   Task assigner

Initially only those workers with the required skills are considered for a task. A ranking based on the combination of the willingness, availability and reputation measures is then provided. The three measures are by default equally weighted. The requestor can however choose to give higher weight for any of those measures. E.g. availability is more critical than willingness. A requestor may also choose to completely disregard any of the measures e.g. availability is of no importance. Ultimately, workers exceeding the quality threshold defined by the requestor are assigned to the task. Furthermore, responses of workers with higher ranking are given a higher weight.

5   Workflow of the skill-ontology based model

The skill-ontology based model's workflow can be functionally broken down into: requestor-side, platform-side and worker-side for ease of illustration as follows. Figure 1 gives a graphical overview of the various components of the model as well as the system's interactions.

1. **Requestor-side:** After the requestor designs the task according to his needs, the list of skills required for that task has to be specified. To that end, the requestor checks the taxonomy of skills provided by the platform. When the required skill is found, the requestor simply adds it in the task's list of required skills. Checking the skill's library of assessments, the requestor chooses the assessments he approves and deems eligible for the task's requirements. If no such assessment is found, the requestor is free to design an assessment of his own, which is then added to the skill's library of assessments as an unverified assessment. On the other hand, if the requestor never finds the required skill from the start, he can add a new one along with at least one assessment. The new skill is initially added to the temporary Ontology. If the defined assessment is a standard assessment, no verification is needed, otherwise it's added as an unverified assessment. In addition to the list of required skills, the requestor defines a threshold for the worker's quality to be employed, as well as the measures of quality (willingness, availability, reputation) he wants to consider and their corresponding weights of importance.

2. Platform-side: The platform maintains at the back-end two ontologies: Temporary and Basic ontology. On the requestor's front end, a view that combines both ontologies is provided. The front-end ontology may or may not reflect the basic ontology at a given time, and may include both verified and unverified skills. Popular unverified skills that are in the temporary ontology are merged with the basic ontology upon verification. Every skill is associated with a library of assessments that holds either standardized assessments and/or personalized assessments. Furthermore, the platform maintains a database of workers, associating each work with a profile of skills (verified, pending verification) along with their computed measures of quality.

3. Worker-side: A worker is free to choose the tasks he wants to be considered for. Only when his skill profile contains the required skills for the corresponding task is he considered for the task. A worker can at any time expand his skill profile, by sitting assessments and attaining new skills. Workers may also boost their reputation by: 1) verifying personalized assessments 2) validating the platform's generated skill taxonomy.
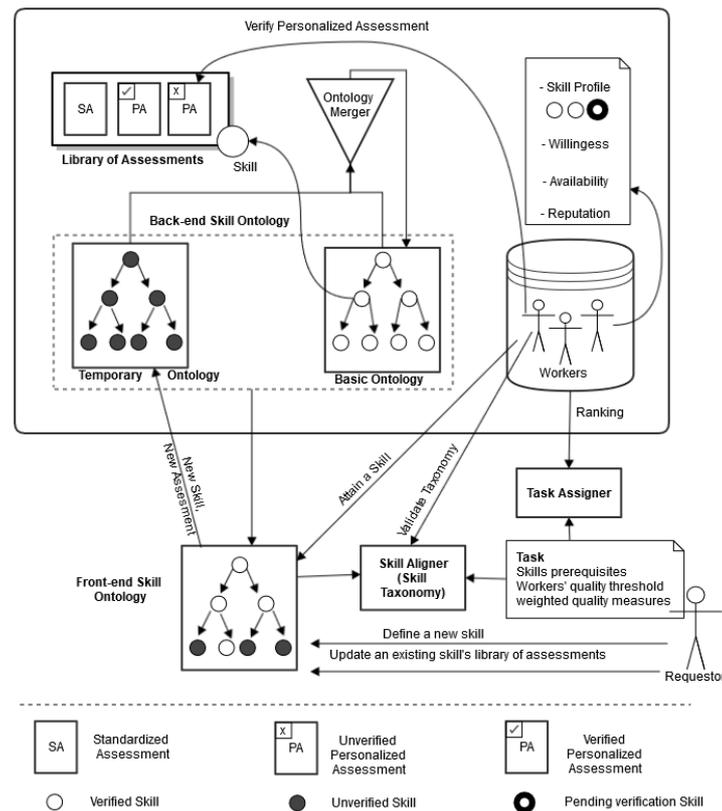

Fig. 1. Skill-Ontology based model workflow

6   Summary & Outlook

Uncertainty is inevitable when dealing with crowdsourcing results. We defined different aspects of quality to identify the corresponding quality assurance measures that should be present. Next, we proposed a skill ontology-based model to be adopted by crowdsourcing platform as a management technique. At its core, the model diminishes the existing uncertainty by eliminating unqualified workers. This is attained by maintaining a dynamically evolving ontology of skills, with libraries of standardized and personalized assessments for awarding credentialed skills. After aligning a worker's set of skills to that required by a task, the resulting quality is improved, where only qualified workers are assigned to the task. Furthermore, in such a setup, qualified workers strive to maintain their qualification level, and unqualified workers are pushed to diligently work on improving their own qualifications. We investigated the model and its workflow on a top level, however, the feasibility of maintaining such a model needs to be further investigated. As examined in the related work section, our model is closely related to web services, reputation-based systems and competency models. Further literature needs to be thoroughly examined, and accordingly adapted to leverage the current model. Furthermore, the proposed workers' quality measures that's to be computed should be formally defined.

References

[1]   J. Surowiecki, The Wisdom of Crowds. Anchor, 2005, p. 336.

[2]   J. Howe, "The Rise of Crowdsourcing," North, vol. 14, no. 14, pp. 1–5, 2006.

[3]   D. C. Brabham, "Crowdsourcing as a Model for Problem Solving: An Introduction and Cases," Convergence: The International Journal of Research into New Media Technologies, vol. 14, no. 1. pp. 75–90, 2008.

[4]   J. Kamps, S. Geva, C. Peters, T. Sakai, A. Trotman, and E. Voorhees, "Report on the SIGIR 2009 workshop on the future of IR evaluation," ACM SIGIR Forum, vol. 43, no. 2. p. 13, 2009.

[5]   D. Zhu and B. Carterette, "An Analysis of Assessor Behavior in Crowdsourced Preference Judgments," in SIGIR 2010 Workshop on Crowdsourcing for Search Evaluation, 2010, no. Cse, pp. 17–20.

[6]   "Scoring Workers in Crowdsourcing: How Many Control Questions are Enough?.pdf," 2013. .

[7]   C. Lofi, J. Selke, and W.-T. Balke, "Information Extraction Meets Crowdsourcing: A Promising Couple," Datenbank-Spektrum, vol. 12, no. 1, 2012.

[8] L. I. Kuncheva, C. J. Whitaker, C. A. Shipp, and R. P. W. Duin, "Limits on the majority vote accuracy in classifier fusion," Pattern Anal. Appl., vol. 6, no. 1, pp. 22–31, Apr. 2003.

[9] G. Kazai, "In Search of Quality in Crowdsourcing for Search Engine Evaluation," SIGIR Forum, vol. 44, no. 2, pp. 165–176, 2011.

[10] W. Mason and D. J. Watts, "Financial incentives and the 'performance of crowds,'" ACM SIGKDD Explorations Newsletter, vol. 11, no. 2. p. 100, 2010.

[11] D. C. Brabham, "MOVING THE CROWD AT THREADLESS," Information, Communication & Society, vol. 13, no. 8. pp. 1122–1145, 2010.

[12] "PodCastle: Collaborative Training of Acoustic Models on the Basis of Wisdom of Crowds for Podcast Transcription - INTERSPEECH2009ogata.pdf," 2009. .

[13] M. Goto and J. Ogata, "Podcastle: recent advances of a spoken document retrieval service improved by anonymous user contributions," in Proceedings of the 12th Annual Conference of the International Speech Communication Association (Interspeech 2011), 2011, no. August, pp. 3073–3076.

[14] D. Schall, Service-Oriented Crowdsourcing: Architecture, Protocols and Algorithms. Springer, 2012, p. 105.

[15] C. Lai, "Endorsements, Licensing, and Insurance for Distributed System Services," The Journal of Electronic Publishing, vol. 2, no. 1. 1996.

[16] H. Ludwig, A. Keller, A. Dan, and R. King, "A service level agreement language for dynamic electronic services," Proc. Fourth IEEE Int. Work. Adv. Issues E-Commerce Web-Based Inf. Syst. (WECWIS 2002), 2002.

[17] V. M. Akhil Sahai, Anna Durante, "Towards Automated SLA Management for Web Services - HPL-2001-310R1.pdf," 2002. .

[18] A. P. Dawid and A. M. Skene, "Maximum likelihood estimation of observer error-rates using the EM algorithm," J. R. Stat. Soc. Ser. C Appl. Stat., vol. 28, no. 1, pp. 20–28, 1979.

[19] V. C. Raykar, S. Yu, L. H. Zhao, G. H. Valadez, C. Florin, L. Bogoni, and L. Moy, "Learning From Crowds," J. Mach. Learn. Res., vol. 11, pp. 1297–1322, 2010.

[20] J. Whitehill, P. Ruvolo, T. Wu, J. Bergsma, and J. Movellan, "Whose Vote Should Count More: Optimal Integration of Labels from Labelers of Unknown Expertise," Adv. Neural Inf. Process. Syst., vol. 22, no. 1, pp. 1–9, 2009.

[21] P. G. Ipeirotis, F. Provost, and J. Wang, "Quality Management on Amazon Mechanical Turk," in New York, 2010, pp. 0–3.

[22] M. a. Campion, A. a. Fink, B. J. Ruggeberg, L. Carr, G. M. Phillips, and R. B. Odman, "Doing Competencies Well: Best Practices in Competency Modeling," Pers. Psychol., vol. 64, no. 1, pp. 225–262, 2011.

[23] J. S. Shippmann, R. A. Ash, M. Battista, L. Carr, L. D. Eyde, B. Hesketh, J. Kehoe, K. Pearlman, E. P. Prien, and J. I. Sanchez, "The practice of competency modeling," Pers. Psychol., vol. 53, pp. 703–740, 2000.

[24] J. L. De Coi, E. Herder, A. Koesling, C. Lofi, D. Olmedilla, O. Papapetrou, and W. Siberski, "A model for competence gap analysis," in WEBIST 2007:

Proceedings of the Third International Conference on Web Information Systems and Technologies, 2007, pp. 304–312.

[25] S. Colucci, T. Di Noia, E. Di Sciascio, F. M. Donini, M. Mongiello, and M. Mottola, "A formal approach to ontology-based semantic match of skills descriptions," J. Univers. Comput. Sci., vol. 9, no. 12, pp. 1437–1454, 2003.

[26] K. Koeppen, J. Hartig, E. Klieme, and D. Leutner, "Current Issues in Competence Modeling and Assessment," Zeitschrift für Psychologie / Journal of Psychology, vol. 216, no. 2. pp. 61–73, 2008.

[27] M. Allahbakhsh, B. Benatallah, A. Ignjatovic, H. R. Motahari-Nezhad, E. Bertino, and S. Dustdar, "Quality Control in Crowdsourcing Systems: Issues and Directions," IEEE Internet Comput., vol. 17, no. 2, pp. 76–81, 2013.

[28] N. F. Mohammad Allahbakhsh, Aleksandar Ignjatovic, Boualem Benatallah, Seyed-Mehdi-Reza Beheshti, Elisa Bertino, "Reputation management in crowdsourcing systems.," 2012. .

[29] A. Ignjatovic, N. Foo, and C. T. L. C. T. Lee, "An Analytic Approach to Reputation Ranking of Participants in Online Transactions," 2008 IEEE/WIC/ACM Int. Conf. Web Intell. Intell. Agent Technol., vol. 1, 2008.

[30] Z. Noorian and M. Ulieru, "The State of the Art in Trust and Reputation Systems: A Framework for Comparison," Journal of theoretical and applied electronic commerce research, vol. 5, no. 2. 2010.

[31] X. Liu, Y. Song, S. Liu, and H. Wang, "Automatic taxonomy construction from keywords," Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp. 1433–1441, 2012.