

Control de calidad en sistemas crowdsourcing: un mapeo sistemático

Quality control in crowdsourcing systems: a systematic mapping.

Andrés Rey Piedrahita, Paulo Andrés Vélez Ángel

Facultad de Ingeniería, Uceva, Tuluá, Colombia - Centro de Bioinformática y Biología computacional BIOS
arey@uceva.edu.co - paulo.velez@bios.co

Resumen— En el ámbito de la Web 2.0 se ha facilitado la implementación de un modelo efectivo en la solución de problemas denominado crowdsourcing que pretende aprovechar las capacidades y competencias de un grupo grande de individuos para resolver una determinada tarea. El crowdsourcing no es un concepto nuevo, pero gracias a las nuevas tecnologías de la información y la comunicación está evolucionando a medida que se emplea en nuevas investigaciones y aplicaciones. Actualmente, el crowdsourcing es utilizado, entre otros, en procesos de adquisición y tratamiento de información estableciendo un nuevo enfoque en el campo de la gestión del conocimiento. Este artículo presenta los resultados de un mapeo sistemático realizado a la literatura con el propósito de conocer los diferentes mecanismos propuestos para asegurar la calidad de los resultados en los sistemas crowdsourcing y las oportunidades de investigación se encuentran en torno a este tema.

Palabras clave— Aseguramiento de calidad, Control de calidad, Crowdsourcing, Mapeo sistemático.

Abstract— The web 2.0 have been easing the implementation of an effective model denominated crowdsourcing to find solutions for given situations. This model pretend to take the advantages, capacities and proficiency of a group of people (from few ones to thousands of peoples) to solve some tasks. The crowdsourcing is not a new concept, but thanks to information and communications technologies (ICT) it is evolving in the same way as the new ICT have been develop and applied. Currently, the crowdsourcing is using in data acquiring process and information generation, establishing a new perspective in the knowledge management field. The present work shows the results of a systematic mapping did to literature review, with the purpose to recognize or learn about several proposed mechanisms to ensure quality results within crowdsourcing systems and research opportunities around this topic.

Key Word — Crowdsourcing, Quality assurance, Quality control, Systematic mapping.

I. INTRODUCCIÓN

El crowdsourcing puede ser visto como un nuevo modelo o

paradigma emergente para la solución de problemas, donde se usan las capacidades y competencias de un grupo grande de individuos (multitud) para resolver una determinada tarea [1]. A diferencia de la externalización (*outsourcing*) no se requiere de una contratación formal y en algunos casos las soluciones aportadas por la multitud superan las dadas por pequeños grupos de expertos [2], por ello se considera particularmente útil en tareas que requieren un gran número de puntos de vista o diferentes soluciones a un mismo problema [3].

El término crowdsourcing vio la luz por primera vez en el año 2006 en la revista Wired donde Jeff Howe lo definió como “el acto de externalización por parte de una empresa o institución, de una función realizada normalmente por un empleado, a un grupo indefinido (y por lo general grande) de personas mediante una convocatoria abierta” [1]. Desde ese momento han surgido nuevas definiciones que han ido recogiendo la evolución y aplicación del concepto que puede llegar a confundirse con otros procesos relacionados pero no exactamente iguales como la innovación abierta, la co-creación, la inteligencia colectiva; también se utiliza como sinónimo de computación humana o computación social que son otros paradigmas emergentes de computación [4].

En la actualidad, el crowdsourcing se lleva a cabo generalmente a través de Internet que ha resultado determinante en su desarrollo y evolución, al proveer la base tecnológica sobre la que se asienta la Web 2.0 que permite a este tipo de iniciativas tener acceso a cientos de miles de individuos de cualquier parte del mundo abarcando un público más amplio y diverso [5]. Aplicando éste concepto diversas aplicaciones se han desarrollado en los últimos años incrementando el número de participantes y plataformas existentes como Amazon Mechanical Turk, Crowdfunder y Threadless [6]. Una forma común de emplear crowdsourcing consiste en publicar las tareas en forma de preguntas [7] que son respondidas por los trabajadores de la multitud motivados por algún tipo de recompensa. Para motivar la participación de las personas en

estos nuevos modelos de participación abierta se pueden tener motivaciones extrínsecas como por ejemplo incentivos financieros o aludir a motivaciones intrínsecas como el reconocimiento de los participantes [8] [9] p.1126.

El crowdsourcing es utilizando cada vez más en diferentes dominios y recientemente se ésta empleando en la adquisición de datos y generación de información estableciendo un enfoque relativamente nuevo en el proceso de adquisición de conocimiento [10]. El rápido crecimiento del número de aplicaciones crowdsourcing crea una variedad de desafíos técnicos dentro de los cuales se tiene como uno de los más importantes el control de la calidad de los resultados obtenidos, porque dada la naturaleza abierta del crowdsourcing los datos recogidos a través de un proceso de este tipo son potencialmente ruidosos [11].

En este artículo se presentan los resultados de un mapeo sistemático de la literatura realizado con el propósito de conocer los diferentes mecanismos propuestos para controlar la calidad en los sistemas crowdsourcing y las oportunidades de investigación que se encuentran en torno al tema. La estructura del artículo es la siguiente: la sección II presenta trabajos similares realizados previamente sobre el crowdsourcing, la sección III describe el método de investigación empleado para realizar el mapeo sistemático de la literatura, la sección IV muestra los resultados de la revisión y la sección V expone las conclusiones del estudio.

II. TRABAJOS RELACIONADOS

El crowdsourcing es un concepto que evoluciona rápidamente gracias a investigaciones y nuevas aplicaciones que se desarrollan en torno a sus diferentes aspectos. Es por esto que parte de la literatura referente al crowdsourcing evidencia una falta de consenso sobre su significado denotando que la parte conceptual no está del todo clara. Evidencia de lo anterior se encuentra en [12] donde se realiza un aporte a la claridad conceptual presentando la siguiente definición integradora de crowdsourcing elaborada a partir de analizar más de 40 definiciones publicadas:

El *crowdsourcing* es un tipo de actividad en línea participativa en la que una persona, institución, organización sin ánimo de lucro, o empresa, propone a un grupo de individuos (de diferentes conocimientos, heterogeneidad y número), mediante una convocatoria abierta y flexible, la realización libre y voluntaria de una tarea. La realización de la tarea, de complejidad y modularidad variable, y en la que la multitud debe participar aportando su trabajo, dinero, conocimiento, y/o experiencia, siempre implica un beneficio mutuo. El usuario recibirá la satisfacción de una necesidad concreta, ya sea esa

económica, de reconocimiento social, de auto-estima, o de desarrollo de aptitudes personales, mientras que el *crowdsourcer* obtendrá y utilizará en su beneficio la aportación del usuario, cuya forma dependerá del tipo de actividad realizada [12].

Posteriormente, buscando proponer un modelo de referencia para el estudio del crowdsourcing en [13] se realiza un mapeo sistemático a la literatura donde a partir de los resultados se propone una taxonomía compuesta por cuatro pilares fundamentales: (1) *The crowd*, que es la multitud se compone de las personas que participan en una actividad de crowdsourcing; (2) *The crowdsourcer*, que es la entidad (una persona, una organización con fines de lucro, una organización sin ánimo de lucro, etc.) que busca el poder y la sabiduría de la multitud para una tarea en cuestión; (3) *The task*, tarea o actividad en la que la gente participa; (4) *The platform*, la plataforma de crowdsourcing que es el sistema (software o no software) dentro del cual se lleva a cabo una tarea de crowdsourcing [14].

La taxonomía anterior integra características extraídas de 113 trabajos seleccionados sobre crowdsourcing que fueron publicados desde su nacimiento en 2006. Los resultados del estudio muestran que existe un creciente interés en el crowdsourcing tanto en el campo científico como en diferentes dominios donde está siendo utilizado. Es de resaltar que no son muchos los mapeos y revisiones sistemáticas de la literatura realizadas en torno al crowdsourcing. Por el lado de las revisiones sistemáticas de la literatura se destacan el trabajo presentado en [15] realizado con el objetivo conocer el estado actual del uso del crowdsourcing como fuente de información para la gestión de desastres en los sistemas de información geográfica voluntaria (*Volunteered Geographic Information, VGI*), [16] que investiga sobre el uso del crowdsourcing en el desarrollo de software y [17] enfocado a conocer el estado actual de las aplicaciones de crowdsourcing en dispositivos móviles.

Los trabajos mencionados anteriormente no han indagado los diferentes mecanismos propuestos para controlar la calidad en los sistemas crowdsourcing y las oportunidades de investigación que se encuentran en torno a este tema que es el tema que cubre el presente estudio.

III. MÉTODO DE INVESTIGACIÓN

Para conocer los diferentes mecanismos propuestos para controlar la calidad en los sistemas crowdsourcing y las oportunidades de investigación que se encuentran en torno a este tema, se realizó un mapeo sistemático de la literatura (*Systematic Mapping*) que es una metodología que se ha

utilizado ampliamente en investigaciones del área médica y recientemente se empieza aplicar al campo de ingeniería con el objetivo clasificar y estructurar los resultados de investigación que han sido publicados [18]; para conocer los avances en un tema e identificar vacíos en la investigación existe también la metodología conocida como revisión sistemática de la literatura, en donde buscando identificar las mejores prácticas (basándose en la evidencia empírica) se realiza una exploración a fondo de los estudios describiendo su métodos y resultados, la diferencia con el mapeo sistemático se tiene en que éste último busca proporcionar una visión más general de grano grueso [19].

En el presente estudio se toma como guía la metodología para revisión sistemática de la literatura descrita en [19] pero se aplica la adaptación que se propone en [18], donde se ajusta para realizar un mapeo sistemático de la literatura. El procedimiento de revisión a seguir está compuesto por cinco etapas: a) definir preguntas de investigación b) realizar la búsqueda literaria, c) seleccionar estudios, d) clasificar estudios y e) extraer y realizar la síntesis de datos. A continuación se describe cada una de las etapas y en la sección IV se presentan los resultados obtenidos.

A. Definir preguntas de investigación.

Para obtener un conocimiento más detallado y una visión integral del tema de interés, primero se plantean las preguntas de investigación (*Research Questions*, RQ) partiendo de lo que se desea conocer:

RQ.1. ¿Qué mecanismos para controlar la calidad de los resultados se han propuesto en el crowdsourcing?

RQ.2. ¿Qué oportunidades de investigación se encuentran en torno a los mecanismos de control de calidad en el crowdsourcing?

B. Realizar la búsqueda literaria

El procedimiento continúa con la delimitación del espacio literario relevante definiendo las cadenas o términos de búsqueda, los tipos de documentos a incluir, las áreas, los lugares de búsqueda y periodos de observación. En el ejercicio de revisión realizado las cadenas de búsqueda presentadas en la tabla 1 fueron definidas a partir de las preguntas de investigación a responder teniendo en cuenta la terminología referente al crowdsourcing y la gestión de información en el idioma inglés, en el cual se encuentran escritos la mayor parte de los trabajos. Como lugar de búsqueda de los términos definidos en el documento se precisó el título. En cuanto a los tipos de documentos a incluir se definió aceptar artículos de

revistas, conferencias y revisiones que provengan de todas las áreas debido al carácter multidisciplinar del Crowdsourcing y teniendo en cuenta que se trata de un término que ha sido potenciado por las nuevas tecnologías de la información y la comunicación no se limitó el periodo de observación buscando identificar toda su evolución. Por otro lado la base de datos referencial utilizada para realizar la búsqueda aplicando los criterios definidos fue Scopus seleccionada por su importancia en la comunidad científica a nivel internacional; la búsqueda literaria en Scopus fue llevada a cabo el 20 de junio de 2016.

C. Seleccionar Estudios.

Para el proceso de selección de estudios se establecen dos criterios para determinar la relevancia de los mismos: inclusión, aquellos trabajos que se centren en la utilización del crowdsourcing y que aborden aspectos en torno al control de calidad; exclusión, trabajos que no contengan los términos de búsqueda presentados en la Tabla 1. También se descartaran los resultados de algunas de las cadenas de búsqueda que presenten problemas con criterios como el HR (High Recall), en el cual se obtiene la información relevante pero además una cantidad enorme de información que no es de utilidad para el estudio y por otro lado, LP (Low Precision) que se presenta al obtener muy pocos trabajos relevantes en comparación con el total de trabajos encontrados [20].

D. Clasificar estudios.

Para clasificar y estructurar los resultados de investigación que han sido publicados se definen las siguientes categorías que corresponden a los diferentes mecanismos de control de calidad propuestos en la literatura: agregación de resultados, votación. Datos de oro (*standard gold*), revisiones multinivel, revisión de expertos, reputación y pre-selección, y otros métodos. Este esquema de clasificación se identificó al revisar cada uno de los trabajos seleccionados y fue evolucionando a medida que se realizaba el proceso, apareciendo nuevas categorías y/o requiriéndose la división o fusión de las mismas.

E. Extracción de Datos y Síntesis.

Para extraer la información de interés de cada trabajo se emplea una planilla de extracción de datos por cada estudio, en esta planilla se consigna la información principal como el título, el resumen, palabras clave, métodos empleados, y lo más importante los aportes, ventajas y falencias que derivan al final en las brechas relacionadas con el problema de investigación.

Cadena de Búsqueda	No. Trabajos Scopus	No. Trabajos. Seleccionados	Precisión de la Cadena
Crowdsourcing	2150	HR, LP	
Crowdsourcing AND Quality Control	21	9	42.8%
Crowdsourcing AND Quality Assurance	5	5	100%
Crowdsourcing AND Quality	94	30	31.9%
Crowdsourcing AND Quality Model	0	0	0%
Crowdsourcing AND Quality Methods	0	0	0%

Tabla 1. Cadenas de búsqueda y precisión obtenida.

IV. RESULTADOS

Aplicando los criterios de selección establecidos anteriormente (inclusión y exclusión), se obtuvo un conjunto de 47 trabajos que se encuentran directamente relacionados con el tema de investigación. A continuación se presentan los resultados del mapeo sistemático que permite dar respuesta a las preguntas de investigación que se plantearon previamente.

A. Mecanismos de control de calidad en Crowdsourcing

Dada la naturaleza abierta del crowdsourcing los datos manejados a través de un proceso de este tipo son potencialmente ruidosos. Para superar este desafío en la literatura se han propuesto los siguientes mecanismos de control de calidad: agregación de resultados, votación, datos de oro (*estándar gold*), revisiones multinivel, revisión de expertos, reputación y pre-selección y otros métodos. A continuación se describe cada uno de ellos:

Agregación de resultados. Es un enfoque muy prometedor que consiste en introducir redundancia pidiendo a varios trabajadores realizar la misma tarea y luego se agregan los resultados buscando obtener un resultado más fiable, en otras palabras, el objetivo de la agregación es encontrar la verdad oculta en el conjunto de respuestas dadas por la multitud para la tarea o problema propuesto [21].

Votación. Consiste en presentar a varios trabajadores los posibles resultados de una determinada tarea, cada trabajador indica la respuesta que considera correcta y posteriormente se

selecciona como respuesta correcta el resultado más votado [22].

Datos de oro. Este mecanismo se basa en preguntas llamadas de oro (cuyas respuestas son conocidos de antemano) que se pueden evaluar automáticamente para comprobar las respuestas dadas buscando estimar la calidad del trabajo realizado. Esta validación es sencilla y no requiere de tiempo adicional o actividades humanas. El porcentaje de respuestas correctas de oro se compara con algún umbral definido por la crowdsourcer para decidir si aceptar o rechazar las respuestas entregadas por el trabajador [23].

Revisiones multinivel. Llamada también opinión de varios niveles, consiste en disponer de un segundo grupo de trabajadores para verificar las respuestas proporcionadas por el primer grupo (se puede aplicar este enfoque teniendo más de dos niveles) [24].

Revisión de expertos. Se trata de un enfoque donde los resultados presentados por los trabajadores son revisados antes de aceptarlos y si un resultado es rechazado el pago puede ser retenido para esa tarea. Es un método muy efectivo pero tiene la desventaja es ser un método costoso que limita la escalabilidad del crowdsourcing debido a que expertos en el tema deben verificar la calidad [22].

Reputación y pre-selección. En esta técnica los trabajadores que se quieren para una determinada tarea se pueden pre-seleccionar con base en su reputación y características específicas. Igualmente se les puede pedir responder a preguntas definidas previamente para evaluar su idoneidad antes de la asignación de tareas [25].

Otros métodos. Se han explorado otros métodos como por ejemplo la capacitación de los participantes en las tareas crowdsourcing que deben llevar a cabo con el fin de recopilar datos de calidad, para lo cual resulta importante proporcionar apoyo a los usuarios mediante la creación de formularios, folletos, guías de referencia y consulta, o incluso un canal de comunicación directa como forma de facilitar el trabajo de los participantes y minimizar la complejidad de la tarea [26]. Otras propuestas no son tan simples y exploran el uso de estadísticas sofisticadas [27] o el aprendizaje automático para aprender y monitorear el desempeño de los trabajadores al realizar las tareas [28].

En la tabla 2 se presenta la agrupación de los trabajos realizada de acuerdo con el mecanismo de control de calidad principal que trata el estudio. En primer lugar se ubican los mecanismos de control de calidad que se basan en la reputación de los trabajadores corroborando lo planteado en [22] donde se plantea que la calidad se puede caracterizar desde dos dimensiones y una de ellas es el perfil de los trabajadores conformado por la reputación y la experiencia. El enfoque de control de calidad basado en reputación busca identificar los siguientes tipos de trabajadores que se pueden tener en

crowdsourcing: (1) los trabajadores fiables, que tienen un conocimiento profundo acerca del dominio específico requerido para la tarea lo que les permite responder a las preguntas con muy alta fiabilidad; (2) los trabajadores normales, que tienen conocimientos generales para dar respuestas correctas, pero cometen errores de vez en cuando; (3) los trabajadores poco ruidosos (*sloppy workers*) que tienen muy poco conocimiento y por lo tanto a menudo dan respuestas incorrectas, pero sin querer; (4) los *uniform spammers* (sin traducción) que son los que intencionalmente dan la misma respuesta para todas las preguntas; (5) los *random spammers* (sin traducción) que descuidadamente dan respuestas aleatorias para todas las preguntas [74].

La segunda dimensión que se menciona en [22] corresponde al diseño de las tareas que se aplica sobre todo para tareas complejas y consiste en descomponerlas en otras más simples conformando flujos de trabajo. En la Tabla 2 los trabajos que aplican este enfoque se clasificaron en lo que se denominó “Otros métodos” ocupando el segundo lugar, como ejemplo se puede resaltar [42] donde se aplica la descomposición de tareas en un proceso de traducción llevado a cabo mediante crowdsourcing (con 60 citas este estudio ocupa el primer puesto de todos los trabajos revisados ordenados en torno a las citas que han tenido). En esta categoría se emplean también otros enfoques como el aprendizaje automático (*Machine Learning*, ML), un buen ejemplo se tiene en [28] donde en tareas crowdsourcing de transcripción de información se emplea el ML de dos formas: primero, empleando una máquina de vectores de soporte (*Support Vector Machine*, SVM) para clasificar las transcripciones de mala calidad en base a las señales acústicas y patrones de lenguaje presentes en las transcripciones; segundo, con base en la puntuación de confianza asignada a los trabajadores de acuerdo a los resultados de sus tareas se puede predecir la calidad de sus transcripciones.

Propuesta Control de Calidad	No. de Trabajos	Trabajos	Porcentaje
Reputación y pre-selección	14	[22] [29] [30] [31] [32] [33] [34] [35] [36] [37] [38] [39] [40] [41]	29.8%
Otros métodos	12	[42] [43] [28] [44] [45] [46] [47] [48] [49] [50] [27] [51]	25.5%

Datos de oro	5	[23] [52] [53] [54] [55]	10.6%
Agregación de resultados	4	[56] [57] [58] [59]	8.5%
Votación	4	[60] [61] [62] [63]	8.5%
Revisiones multinivel	4	[64] [65] [66] [67]	8.5%
Revisión de expertos	4	[68] [69] [70] [71]	8.5%

Tabla 2. Clasificación de los trabajos.

En tercer lugar se encuentran los mecanismos basados en datos de oro que presentan mayor facilidad de implementación que los métodos anteriores, pero presentan la desventaja de generar trabajo adicional al tener que crearlos cuando no existen, Por ejemplo en [52] se aprovechan, como datos de oro, anotaciones existentes de datos clínicos generadas por expertos que permiten evaluar la calidad de la anotaciones crowdsourcing realizadas por los trabajadores sobre los registros clínicos. Compartiendo los últimos lugares con igual porcentaje se encuentran la votación que es hasta ahora el método más fácil de implementar y presenta buenos resultados al permitir seleccionar una respuesta de alta calidad a partir de un conjunto de respuestas, en [61] por ejemplo se demuestran sus bondades empleando la votación en tareas crowdsourcing de audición y traducción. Otra propuesta es la agregación de resultados que busca aprovechar la inteligencia colectiva que resulta del concurso de la multitud, en este enfoque se busca llegar a un consenso a partir de las diferentes respuestas entregadas por los trabajadores. Aquí se espera que al agregar las respuestas se mejore la calidad de los resultados [62], para el caso de magnitudes numéricas esto es sencillo pero muy complejo cuando se deben agregar entradas de texto. El enfoque de revisiones multinivel, que como se indica en [65] es también un mecanismo que presenta buenos resultados y es de fácil implementación, en dicho estudio se emplea el crowdsourcing en la transcripción de información y para el control de la calidad se define un nivel llamado “arbitraje” donde llegan los registros y las entradas que transcriben independientemente dos voluntarios para que se seleccione una de ellas. Por último se encuentra la revisión de expertos que tiene como desventaja el costo que genera tener que disponer de expertos lo que también limita la escalabilidad del crowdsourcing, en los trabajos revisados se observa que este método se utiliza cuando se desea evaluar si el crowdsourcing resulta adecuado para la ejecución de una determinada tarea en un dominio específico, en [68] por ejemplo se emplea para medir el desempeño del crowdsourcing en el campo de la recuperación de información.

B. Desafíos y líneas de investigación abierta.

El crowdsourcing se encuentra todavía en una etapa temprana donde se afrontan varios desafíos y se tienen líneas de investigación abierta dentro de las cuales se pueden mencionar: las plataformas crowdsourcing, donde se abordan aspectos de arquitectura y funcionalidad; los procesos y tareas, indagando como descomponer tareas complejas en otras más simples que resulten más adecuadas para resolver mediante crowdsourcing; la motivación de los participantes; y los mecanismos de control de calidad en los sistemas crowdsourcing.

En lo que respecta a los mecanismos de control de calidad los resultados del mapeo sistemático permiten identificar que gran parte de los estudios (31 trabajos que representan el 66%) que abordan el control de calidad en los sistemas crowdsourcing corresponden a métodos concretos que se basan en alguno de los enfoques descritos en el punto anterior. En la tabla 3 se pueden observar las diferentes propuestas para el control de calidad identificadas en la literatura, clasificadas de acuerdo al tipo de contribución (modelo/métrica, marco de trabajo/herramienta, o método) que realizan para realizar el control de calidad.

Propuesta	Tipo de Aporte		
	Modelos/ Métricas	Marcos de Trabajo / Herramientas	Métodos
Agregación Resultados		[56]	[57] [58] [59]
Votación		[60]	[61] [62] [63]
Datos de Oro		[54]	[23] [52] [53] [55]
Revisiones multinivel	[65]		[64] [66] [67]
Revisión expertos			[67] [69][70] [71]
Reputación y pre- selección	[22] [32] [35] [36] [37]	[33]	[71] [30] [31] [34] [38] [39] [73] [41]
Otros métodos	[42] [43] [28] [44] [45]	[46]	[46] [48] [49] [49] [27] [27]

Tabla 3. Clasificación de los trabajos de acuerdo al tipo de aporte

En una medida importante con 11 trabajos que representan el 23.4% empiezan aparecer estudios que enfocan el control de calidad en los sistemas crowdsourcing hacia la definición de modelos de calidad y métricas. En menor medida con 5 trabajos y un 10.6% otros apuntan al desarrollo de marcos de trabajo y/o herramientas que faciliten el control de calidad en éstos sistemas siendo éstos tópicos las que constituyen las mejores oportunidades para nuevas investigaciones.

V. CONCLUSIONES

Controlar la calidad siempre ha sido un desafío por tratarse de un tema subjetivo en general, actualmente éste aspecto es uno de los factores más importantes que afectan el crecimiento de la industria crowdsourcing [72]. En la revisión adelantada a la literatura en el presente estudio se pudo identificar siete métodos habituales utilizados para realizar el control de calidad en los sistemas crowdsourcing. Los métodos identificados fueron la agregación de resultados, la votación, los datos de oro, las revisiones multinivel, la revisión de expertos, la reputación y pre-selección, y otros métodos.

Estos métodos se pueden combinar de diferentes formas buscando obtener mejores resultados, no obstante, lograr la calidad de los resultados del crowdsourcing es difícil por varias razones dentro de las cuales se encuentra que los trabajadores de la multitud tienen diferentes niveles de experiencia y motivación, además las preguntas o tareas propuestas varían en diferentes grados de dificultad, lo que puede conducir a una alta contradicción e incertidumbre en el conjunto de respuestas dadas [73]. Es por ello que se plantea que la calidad de los resultados de una tarea crowdsourcing obedece en gran medida el rendimiento de los trabajadores multitud, por lo cual la reputación es el enfoque más utilizado para realizar el control de calidad.

En último lugar, es de resaltar el crowdsourcing ha demostrado ser un enfoque eficaz y escalable en especial para superar problemas que son computacionalmente costosos o imposibles de resolver para las máquinas, pero que son más bien sencillos para los seres humanos, de allí que las tareas propuestas a la multitud sean llamadas tareas de inteligencia humana (*Human Intelligence Task, HITs*). En estos momentos el crowdsourcing se encuentra todavía en una etapa temprana donde se afrontan varios desafíos y se tienen líneas de investigación abiertas como son: las plataformas crowdsourcing, donde se abordan aspectos de arquitectura y funcionalidad; los procesos y tareas, indagando como descomponer tareas complejas en otras más simples que resulten más adecuadas para resolver mediante crowdsourcing; la motivación de los participantes; y en lo que respecta al control de calidad se plantean la optimización de los métodos existentes y la formulación de nuevos métodos, pero al mismo tiempo el trabajo se empieza a enfocar hacia un tratamiento más integral con la definición de modelos y métricas de calidad, y el desarrollo de marcos de trabajo o herramientas que faciliten el control de calidad en los sistemas crowdsourcing.

REFERENCIAS

- [1] J. Howe, "The rise of crowdsourcing," *Wired Mag.*, vol. 14, no. 6, pp. 1–4, 2006.
- [2] J. Surowiecki, *The wisdom of crowds*. Anchor, 2005.
- [3] D. C. Brabham, "Crowdsourcing as a model for problem solving an introduction and cases," *Converg. Int. J. Res. into new media Technol.*, vol. 14, no. 1, pp. 75–90, 2008.
- [4] E. Estellés-Arolas and F. González-Ladrón-De-Guevara, "Clasificaci{ó}n de iniciativas de crowdsourcing basada en tareas," *El Prof. la Inf.*, vol. 21, no. 3, pp. 283–291, 2012.
- [5] T. Erickson, "Geocentric crowdsourcing and smarter cities: Enabling urban intelligence in cities and regions," in *1st Ubiquitous Crowdsourcing Workshop at UbiComp*, 2010.
- [6] M. Hosseini, A. Shahri, K. Phalp, and R. Ali, "Recommendations on adapting crowdsourcing to problem types," in *Research Challenges in Information Science (RCIS), 2015 IEEE 9th International Conference on*, 2015, pp. 423–433.
- [7] K. Benouaret, R. Valliyur-Ramalingam, and F. Charoy, "Crowdsc: Building smart cities with large-scale citizen participation," *Internet Comput. IEEE*, vol. 17, no. 6, pp. 57–63, 2013.
- [8] J. Rogstadius, V. Kostakos, A. Kittur, B. Smus, J. Laredo, and M. Vukovic, "An Assessment of Intrinsic and Extrinsic Motivation on Task Performance in Crowdsourcing Markets.," *ICWSM*, vol. 11, pp. 17–21, 2011.
- [9] D. C. Brabham, "Moving the crowd at Threadless: Motivations for participation in a crowdsourcing application," *Information, Commun. Soc.*, vol. 13, no. 8, pp. 1122–1145, 2010.
- [10] C.-A. Papadopoulou and M. Giaoutzi, "Crowdsourcing as a tool for knowledge acquisition in spatial planning," *Futur. Internet*, vol. 6, no. 1, pp. 109–125, 2014.
- [11] A. J. Quinn and B. B. Bederson, "@article{oleson2011programmatic, title={Programmatic Gold: Targeted and Scalable Quality Assurance in Crowdsourcing.}, author={Oleson, David and Sorokin, Alexander and Laughlin, Greg P and Hester, Vaughn and Le, John and Biewald, Lukas}, journal={Human co.} in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2011, pp. 1403–1412.
- [12] E. Estellés-Arolas and F. González-Ladrón-de-Guevara, "Towards an integrated crowdsourcing definition," *J. Inf. Sci.*, vol. 38, no. 2, pp. 189–200, 2012.
- [13] M. Hosseini, A. Shahri, K. Phalp, J. Taylor, and R. Ali, "Crowdsourcing: A taxonomy and systematic mapping study," *Comput. Sci. Rev.*, vol. 17, pp. 43–69, 2015.
- [14] M. Hosseini, K. Phalp, J. Taylor, and R. Ali, "The four pillars of crowdsourcing: A reference model," in *Research Challenges in Information Science (RCIS), 2014 IEEE Eighth International Conference on*, 2014, pp. 1–12.
- [15] F. E. A. Horita, L. C. Degrossi, L. F. G. de Assis, A. Zipf, and J. P. de Albuquerque, "The use of volunteered geographic information (VGI) and crowdsourcing in disaster management: a systematic literature review," 2013.
- [16] N. Leicht, D. Durward, I. Blohm, and J. M. Leimeister, "Crowdsourcing in Software Development: A State-of-the-Art Analysis."
- [17] F. Mahmud and H. Aris, "State of mobile crowdsourcing applications: A review," in *Software Engineering and Computer Systems (ICSECS), 2015 4th International Conference on*, 2015, pp. 27–32.
- [18] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, "Systematic mapping studies in software engineering," in *12th international conference on evaluation and assessment in software engineering*, 2008, vol. 17, no. 1, pp. 1–10.
- [19] B. Kitchenham, R. Pretorius, D. Budgen, O. P. Brereton, M. Turner, M. Niazi, and S. Linkman, "Systematic literature reviews in software engineering--a tertiary study," *Inf. Softw. Technol.*, vol. 52, no. 8, pp. 792–805, 2010.
- [20] H. Zhang, M. A. Babar, and P. Tell, "Identifying relevant studies in software engineering," *Inf. Softw. Technol.*, vol. 53, no. 6, pp. 625–637, 2011.
- [21] L. Duan, S. Oyama, H. Sato, and M. Kurihara, "Separate or joint? Estimation of multiple labels from crowdsourced annotations," *Expert Syst. Appl.*, vol. 41, no. 13, pp. 5723–5732, 2014.
- [22] 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.*, no. 2, pp. 76–81, 2013.
- [23] D. Oleson, A. Sorokin, G. P. Laughlin, V. Hester, J. Le, and L. Biewald, "Programmatic Gold: Targeted and Scalable Quality Assurance in Crowdsourcing.," *Hum. Comput.*, vol. 11, no. 11, 2011.
- [24] A. J. Quinn and B. B. Bederson, "Human

- computation: a survey and taxonomy of a growing field,” in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2011, pp. 1403–1412.
- [25] G. Paolacci, J. Chandler, and P. G. Ipeirotis, “Running experiments on amazon mechanical turk,” *Judgm. Decis. Mak.*, vol. 5, no. 5, pp. 411–419, 2010.
- [26] J. C. Tweddle, L. D. Robinson, M. J. O. Pocock, and H. E. Roy, *Guide to citizen science: developing, implementing and evaluating citizen science to study biodiversity and the environment in the UK*. NERC/Centre for Ecology & Hydrology, 2012.
- [27] Y. Baba and H. Kashima, “Statistical quality estimation for general crowdsourcing tasks,” in *AAAI Workshop - Technical Report*, 2013, vol. WS-13–18, pp. 8–9.
- [28] C.-Y. Lee and J. Glass, “A transcription task for crowdsourcing with automatic quality control,” in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2011, pp. 3041–3044.
- [29] A. . Kulkarni, P. . Gutheim, P. . Narula, D. . Rolnitzky, T. . Parikh, and B. . Hartmann, “Mobileworks: Designing for quality in a managed crowdsourcing architecture,” *IEEE Internet Comput.*, vol. 16, no. 5, pp. 28–35, 2012.
- [30] A. J. Mashhadi and L. Capra, “Quality control for real-time ubiquitous crowdsourcing,” in *UbiCrowd’11 - Proceedings of the 2nd International Workshop on Ubiquitous Crowdsourcing*, 2011, pp. 5–8.
- [31] G. . Kazai, J. . Kamps, and N. . Milic-Frayling, “The face of quality in crowdsourcing relevance labels: Demographics, personality and labeling accuracy,” in *ACM International Conference Proceeding Series*, 2012, pp. 2583–2586.
- [32] J. Feng, G. Li, H. Wang, and J. Feng, “Incremental quality inference in crowdsourcing,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8422 LNCS, no. PART 2, pp. 453–467, 2014.
- [33] D. Schall, “Automatic quality management in crowdsourcing [Leading Edge],” *IEEE Technol. Soc. Mag.*, vol. 32, no. 4, pp. 9–13, 2013.
- [34] J. . Jarrett, L. F. D. . Silva, L. . Mello, S. . Andere, G. . Cruz, and M. B. . Blake, “Self-generating a labor force for crowdsourcing: Is worker confidence a predictor of quality?,” in *Proceedings - 3rd Workshop on Hot Topics in Web Systems and Technologies, HotWeb 2015*, 2015, pp. 85–90.
- [35] A. Fremuth, V. Adzic, and H. Kalva, “Parameterized framework for the analysis of visual quality assessments using crowdsourcing,” in *Proceedings of SPIE - The International Society for Optical Engineering*, 2015, vol. 9394.
- [36] B. . Nicholson, V. S. . Sheng, J. . Zhang, Z. . Wang, and X. . Xian, “Improving label accuracy by filtering low-quality workers in crowdsourcing,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9413, pp. 547–559, 2015.
- [37] M.-C. Yuen, I. King, and K.-S. Leung, “Probabilistic matrix factorization with active learning for quality assurance in crowdsourcing systems,” in *Proceedings of the 14th International Conference WWW/Internet 2015*, 2015, pp. 127–138.
- [38] M. . Ashikawa, T. . Kawamura, and A. . Ohsuga, “Development and evaluation of quality control methods in a microtask crowdsourcing platform,” *Trans. Japanese Soc. Artif. Intell.*, vol. 29, no. 6, pp. 503–515, 2014.
- [39] M. . Allahbakhsh, S. . Samimi, H.-R. . d Motahari-Nezhad, and B. . Benatallah, “Harnessing implicit teamwork knowledge to improve quality in crowdsourcing processes,” in *Proceedings - IEEE 7th International Conference on Service-Oriented Computing and Applications, SOCA 2014*, 2014, pp. 17–24.
- [40] G. Zhang and H. Chen, “Quality control of massive data for crowdsourcing in location-based services,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8286 LNCS, no. PART 2, pp. 112–121, 2013.
- [41] J. Jo, A. Stevens, and C. Tan, “A quality control model for trustworthy crowdsourcing in collaborative learning,” *Adv. Intell. Syst. Comput.*, vol. 208 AISC, pp. 85–90, 2013.
- [42] O. F. Zaidan and C. Callison-Burch, “Crowdsourcing translation: Professional quality from non-professionals,” in *ACL-HLT 2011 - Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 2011, vol. 1, pp. 1220–1229.
- [43] Z.-Q. Zhang, J.-S. Pang, X.-Q. Xie, and Y. Zhou, “Research on crowdsourcing quality control strategies and evaluation algorithm,” *Jisuanji Xuebao/Chinese J. Comput.*, vol. 36, no. 8, pp. 1636–1649, 2013.
- [44] Y. . Zhengy, J. . Wang, G. . Li, R. . Chengy, and J. . Feng, “QASCA: A Quality-Aware task assignment system for crowdsourcing applications,” in *Proceedings of the ACM SIGMOD International*

- Conference on Management of Data*, 2015, vol. 2015-May, pp. 1031–1046.
- [45] K. E. . Maarry, W.-T. . Balke, H. . Cho, S.-W. . Hwang, and Y. . Baba, “Skill ontology-based model for quality assurance in crowdsourcing,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8505 LNCS, pp. 376–387, 2014.
- [46] M. Hirth, S. Scheuring, T. Hossfeld, C. Schwartz, and P. Tran-Gia, “Predicting result quality in Crowdsourcing using application layer monitoring,” in *2014 IEEE 5th International Conference on Communications and Electronics, IEEE ICCE 2014*, 2014, pp. 510–515.
- [47] S. . Han, P. . Dai, P. . Paritosh, and D. . Huynh, “Crowdsourcing human annotation on web page structure: Infrastructure design and behavior-based quality control,” *ACM Trans. Intell. Syst. Technol.*, vol. 7, no. 4, 2016.
- [48] D. . Iren and S. . Bilgen, “Cost models of Quality assurance in crowdsourcing,” in *2014 IEEE 5th International Conference on Communications and Electronics, IEEE ICCE 2014*, 2014, pp. 504–509.
- [49] H. . Su and S. . Liao, “Efficient crowdsourcing via Bayes quality control,” *Adv. Mater. Res.*, vol. 998–999, pp. 1576–1580, 2014.
- [50] T. Schulze, D. Nordheimer, and M. Schader, “Worker perception of quality assurance mechanisms in crowdsourcing and human computation markets,” in *19th Americas Conference on Information Systems, AMCIS 2013 - Hyperconnected World: Anything, Anywhere, Anytime*, 2013, vol. 5, pp. 4046–4056.
- [51] T. . Matsui, Y. . Baba, T. . Kamishima, and H. . c Kashima, “Crowdsourcing quality control for item ordering tasks,” in *AAAI Workshop - Technical Report*, 2013, vol. WS-13–18, pp. 52–53.
- [52] H. Zhai, T. Lingren, L. Deleger, Q. Li, M. Kaiser, L. Stoutenborough, and I. Solti, “Web 2.0-based crowdsourcing for high-quality gold standard development in clinical natural language processing,” *J. Med. Internet Res.*, vol. 15, no. 4, 2013.
- [53] G. . Kaza and I. . Zitouni, “Quality management in crowdsourcing using gold judges behavior,” in *WSDM 2016 - Proceedings of the 9th ACM International Conference on Web Search and Data Mining*, 2016, pp. 267–276.
- [54] J. . Zhang, V. S. . Sheng, J. . Wu, X. . Fu, and X. . Wu, “Improving label quality in crowdsourcing using noise correction,” in *International Conference on Information and Knowledge Management, Proceedings*, 2015, vol. 19–23-Oct-, pp. 1931–1934.
- [55] T. Kubota and M. Aritsugi, “How Many Ground Truths Should We Insert? Having Good Quality of Labeling Tasks in Crowdsourcing,” in *Proceedings - International Computer Software and Applications Conference*, 2015, vol. 2, pp. 796–805.
- [56] F. . Ribeiro, D. . Florencio, and V. . Nascimento, “Crowdsourcing subjective image quality evaluation,” in *Proceedings - International Conference on Image Processing, ICIP*, 2011, pp. 3097–3100.
- [57] M. . Acosta, A. . Zaveri, E. . Simperl, D. . Kontokostas, S. . Auer, and J. . Lehmann, “Crowdsourcing linked data quality assessment,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8219 LNCS, no. PART 2, pp. 260–276, 2013.
- [58] D. . Kontokostas, A. . Zaveri, S. . Auer, and J. . Lehmann, “Triplecheckmate: A tool for crowdsourcing the quality assessment of linked data,” *Commun. Comput. Inf. Sci.*, vol. 394, pp. 265–272, 2013.
- [59] O. . Figuerola Salas, V. . Adzic, H. . Kalva, and A. . Shah, “Assessing Internet video quality using crowdsourcing,” in *CrowdMM 2013 - Proceedings of the 2nd ACM International Workshop on Crowdsourcing for Multimedia*, 2013, pp. 23–28.
- [60] S. Lee, S. Park, and S. Park, “A quality enhancement of crowdsourcing based on quality evaluation and user-level task assignment framework,” in *2014 International Conference on Big Data and Smart Computing, BIGCOMP 2014*, 2014, pp. 60–65.
- [61] S. . Feng, X. . Li, and H. . Ou, “Quality control method in crowdsourcing platform for professional dictionary compilation process (PDCCP),” *Commun. Comput. Inf. Sci.*, vol. 495, pp. 75–92, 2015.
- [62] D. Yungn, M.-L. Li, and S. Chang, “Evolutionary approach for crowdsourcing quality control,” *J. Vis. Lang. Comput.*, vol. 25, no. 6, pp. 879–890, 2014.
- [63] D.-J. . Yue, G. . Yu, D.-R. . Shen, and X.-C. . Yu, “Crowdsourcing quality evaluation strategies based on voting consistency,” *Dongbei Daxue Xuebao/Journal Northeast. Univ.*, vol. 35, no. 8, pp. 1097–1101, 2014.
- [64] Q. . Xu, Q. . Huang, and Y. . Yao, “Online crowdsourcing subjective image quality assessment,” in *MM 2012 - Proceedings of the 20th ACM International Conference on Multimedia*, 2012, pp. 359–368.
- [65] D. L. . Hansen, P. . Schone, D. . Corey, M. . Reid, and J. . Gehring, “Quality control mechanisms for crowdsourcing: Peer review, arbitration, & expertisat

- familysearch indexing,” in *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, 2013, pp. 649–660.
- [66] S. . Liao, C. . Wu, and J. . Huerta, “Evaluating human correction quality for machine translation from crowdsourcing,” in *International Conference Recent Advances in Natural Language Processing, RANLP*, 2011, pp. 598–603.
- [67] T. Xia, C. Zhang, J. Xie, and T. Li, “Real-time quality control for crowdsourcing relevance evaluation,” in *Proceedings - 2012 3rd IEEE International Conference on Network Infrastructure and Digital Content, IC-NIDC 2012*, 2012, pp. 535–539.
- [68] J. B. P. . Vuurens and A. P. . De Vries, “Obtaining high-quality relevance judgments using crowdsourcing,” *IEEE Internet Comput.*, vol. 16, no. 5, pp. 20–27, 2012.
- [69] H. . b Gelas, S. T. . Abate, L. . Besacier, and F. . Pellegrino, “Quality assessment of crowdsourcing transcriptions for African languages,” in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2011, pp. 3065–3068.
- [70] C. F. . b Salk, T. . Sturn, L. . See, S. . Fritz, and C. . Perger, “Assessing quality of volunteer crowdsourcing contributions: lessons from the Cropland Capture game,” *Int. J. Digit. Earth*, vol. 9, no. 4, pp. 410–426, 2016.
- [71] A. Kulkarni, P. Gutheim, P. Narula, D. Rolnitzky, T. Parikh, and B. Hartmann, “MobileWorks: Designing for quality in a managed crowdsourcing architecture (Extended abstract),” in *AAAI Workshop - Technical Report*, 2012, vol. WS-12–08, pp. 9–10.
- [72] A. Aker, M. El-Haj, M.-D. Albakour, U. Kruschwitz, and others, “Assessing Crowdsourcing Quality through Objective Tasks.,” in *LREC*, 2012, pp. 1456–1461.
- [73] J. Ross, L. Irani, M. Silberman, A. Zaldivar, and B. Tomlinson, “Who are the crowdworkers?: shifting demographics in mechanical turk,” in *CHI’10 extended abstracts on Human factors in computing systems*, 2010, pp. 2863–2872.
- [74] G. Kazai, J. Kamps, and N. Milic-Frayling, “Worker types and personality traits in crowdsourcing relevance labels,” in *Proceedings of the 20th ACM international conference on Information and knowledge management*, 2011, pp. 1941–1944.