

**UNIVERSIDAD POLITÉCNICA DE VALENCIA**

**Departamento de Organización de Empresas**



**UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA**

**TESIS DOCTORAL:**

Relación entre el crowdsourcing y la inteligencia colectiva: el caso de los sistemas de etiquetado social

**Presentada por:**

D. Enrique Estellés Arolas

**Dirigida por:**

Dr. D. Fernando González Ladrón de Guevara

Valencia, Julio de 2013

*A mi esposa Noemí,  
por su amor, ayuda y paciencia.*

*A mis hijos Enrique, Clara, Francisco y Elena (mi pequeña multitud)  
por su amor e inspiración.*

*A mis padres Eduardo y Cristina,  
por su amor y confianza.*

**Título:** "Relación entre el crowdsourcing y la inteligencia colectiva: el caso de los sistemas de etiquetado social"

**Autor:** Enrique Estellés Arolas

**Director:** Dr. Fernando González Ladrón de Guevara

### **RESUMEN**

El crowdsourcing es un término acuñado recientemente que hace referencia a un tipo de iniciativas que se dan en Internet. En estas iniciativas, alguien, ya sea una empresa, una persona o una institución, propone a la multitud de Internet la realización de una tarea a cambio de una recompensa.

Para que estas iniciativas se puedan llevar a cabo, Internet, y más concretamente, el desarrollo de la Web 2.0, ha sido fundamental. Internet, además de suponer la base tecnológica sobre la que se asienta el crowdsourcing, permite a este tipo de iniciativas tener acceso a cientos de miles de individuos de cualquier parte del mundo.

Al haber sido un término acuñado recientemente, la literatura existente es escasa, realidad que va subsanándose paulatinamente. Además, las fronteras conceptuales del término son difusas. Por esta razón, muchas veces se confunde el crowdsourcing con procesos relacionados aunque no exactamente iguales, como la innovación abierta, la co-creación o la inteligencia colectiva.

La presente tesis tiene como objetivo clarificar cual es exactamente la relación existente entre el crowdsourcing y uno de estos fenómenos: la inteligencia colectiva. Con este fin, se analizarán los sistemas de etiquetado social, una aplicación Web 2.0 claramente perteneciente al ámbito de la Inteligencia Colectiva, para observar las diferencias y semejanzas entre ésta y el crowdsourcing.

En el camino que se recorre para identificar y analizar esta relación, se alcanzan otros hitos relevantes que ayudan a conseguir el objetivo de la tesis. En lo que al crowdsourcing respecta, se ha definido este término en base a ocho elementos, lo que facilita la identificación de qué es o no crowdsourcing. También se ha desarrollado una tipología de iniciativas de crowdsourcing en base a otras tipologías propuestas por diferentes autores. En cuanto a los sistemas de etiquetado social, se ha analizado y descrito el uso que hacen los usuarios de las etiquetas que describen los recursos de Internet, además de explicar como estos sistemas pueden favorecer los procesos de investigación colaborativos.

**Títol:** "Relació entre el crowdsourcing i la intel·ligència col·lectiva: el cas dels sistemes d'etiquetatge social"

**Autor:** Enrique Estellés Arolas

**Director:** Dr. Fernando González Ladrón de Guevara

### RESUM

El crowdsourcing és un terme encunyat recentment que fa referència a un tipus d'iniciatives que es donen a Internet. En aquestes iniciatives, algú, ja sigui una empresa, una persona o una institució, proposa a la multitud d'Internet la realització d'una tasca a canvi d'una recompensa.

Perquè aquestes iniciatives es puguin dur a terme, Internet, i més concretament, el desenvolupament de la web 2.0, han estat fonamentals. Internet, a més de suposar la base tecnològica sobre la qual s'assenta el crowdsourcing, permet a aquest tipus d'iniciatives tenir accés a centenars de milers d'individus d'arreu del món.

En haver estat un terme encunyat recentment, la literatura existent és escassa, realitat que es va esmenant gradualment. A més, les fronteres conceptuals del terme són difuses. Per aquesta raó, moltes vegades es confon el crowdsourcing amb processos relacionats encara que no exactament iguals, com la innovació oberta, la co-creació o la intel·ligència col·lectiva.

La present tesi té com a objectiu aclarir quina és exactament la relació existent entre el crowdsourcing i un d'aquests fenòmens: la intel·ligència col·lectiva. Amb aquesta finalitat, s'analitzaran els sistemes d'etiquetatge social, una aplicació Web 2.0 clarament pertanyent a l'àmbit de la intel·ligència col·lectiva, per observar les diferències i semblances entre aquesta i el crowdsourcing.

En el camí recorregut per identificar i dibuixar aquesta relació, s'assoleixen altres fites rellevants que ajuden a aconseguir l'objectiu de la tesi. Pel que al crowdsourcing fa, s'ha definit aquest terme en funció de vuit elements, fet que facilita la identificació de què és o no crowdsourcing. També s'ha desenvolupat una tipologia d'iniciatives de crowdsourcing en base a altres tipologies proposades per diferents autors. Pel que fa als sistemes d'etiquetatge social, s'ha analitzat i descrit l'ús que fan els usuaris de les etiquetes que descriuen els recursos d'Internet, a més d'explicar com aquests sistemes poden afavorir els processos de recerca col·laboratius.

**Title:** "Relationship between crowdsourcing and collective intelligence: the social tagging systems case"

**Author:** Enrique Estellés Arolas

**Director:** Dr. Fernando González Ladrón de Guevara

### **ABSTRACT**

Crowdsourcing is a recently coined term that refers to a type of initiatives that exist on the Internet. In these initiatives, someone, whether a company, a person or an institution, offers to the Internet crowd the accomplishment of a task in exchange for a reward.

For these initiatives to be carried out, Internet, and more specifically, the development of Web 2.0 have been critical. Internet, in addition to being the crowdsourcing technological base, allows such initiatives to access to hundreds of thousands of individuals from all over the world.

Because it is a recently coined term, the existing literature is scarce, although it's a reality that is being gradually changing. Furthermore, the conceptual boundaries of the term are blurred. For this reason it is often confused with other processes that, although they're related to crowdsourcing, they're not the same. These processes are for example open innovation, co-creation or collective intelligence.

This thesis aims to clarify which is the relationship between crowdsourcing and one of these phenomena: collective intelligence. To this end, the social tagging systems will be analyzed, a Web 2.0 application clearly within the scope of collective intelligence, to see the differences and similarities with crowdsourcing.

In the way of claryfing this relationship, other significant milestones are reached that help achieve the objective of the thesis. Regarding crowdsourcing, the term has been defined based on eight elements, what facilitates future identifications of what is or is not crowdsourcing. It has also been developed an integrated typology of crowdsourcing initiatives, based on other author typologies. Regarding social tagging systems, these systems has been analyzed and described, showing how can favor collaborative research processes and how users use the different kind of tags used to describe Internet resources.

### **AGRADECIMIENTOS**

En primer lugar quiero dar las gracias a mi director de tesis Fernando González Ladrón de Guevara, que tan sabiamente ha sabido guiarme, orientarme, motivarme y ayudarme. Ha sido un excelente director y amigo. Siempre ha confiado en mí, invitándome a participar en todos los proyectos que ha podido, de uno de los cuales nace la temática de esta tesis. Sin su sabia dirección, esta tesis no se podría haber escrito.

Gracias a los participantes del proyecto "*Metal 2.0: crowdsourcing*". En especial a Santiago Bonet, Javier Megías y Elena Benito. Su profesionalidad, actitud y conocimientos han sido, y son, una inspiración, un ejemplo y una ayuda.

Gracias a Antonio Falcó, por su ayuda, confianza y apoyo tanto durante mis años de estudiante de Ingeniería Informática así como durante los comienzos de mi doctorado.

Gracias a Gracia Prats Arolas y a Hallie Kreitlow, por su ayuda en la traducción y revisión de los artículos escritos en inglés.

Nunca podré dar suficientes gracias a mi mujer Noemí. Sin su complicidad, paciencia, ayuda, cariño y colaboración, sin esas horas dedicadas en exclusiva a nuestros hijos mientras yo me encerraba para estudiar y escribir sobre esa cosa rara ("*¿crowdque?*" solía preguntarme), esta tesis no habría sido posible.

Gracias a mis hijos (mi pequeña multitud), tres de los cuales han nacido mientras yo realizaba esta tesis. Ellos me han inspirado y empujado a trabajar más rápido y mejor.

Gracias a mis padres, que siempre han confiado en mí, me han apoyado y respetado en todas mis decisiones. Han sido, y son, un ejemplo para mí.

Gracias al resto de mi familia y a mis hermanos de comunidad, que tan de cerca han vivido esta aventura del doctorado y que tanto han rezado por mí.

Por último, gracias a Dios, por su misericordia, paciencia y amor.

## Índice

<b>CAPÍTULO 1 - Introducción</b> .....	<b>1</b>
<b>1.1 Introducción</b> .....	<b>2</b>
1.1.1 La inteligencia colectiva .....	2
1.1.2 El crowdsourcing .....	4
1.1.3 Los sistemas de etiquetado social .....	6
<b>1.2 Antecedentes</b> .....	<b>8</b>
<b>1.3 Objetivos</b> .....	<b>12</b>
1.3.1 Objetivo principal .....	12
1.3.2 Objetivos secundarios .....	12
<b>1.4 Metodología</b> .....	<b>14</b>
1.4.1 Metodología utilizada en la línea del crowdsourcing .....	14
1.4.2 Metodología utilizada en el área de los sistemas de etiquetado social .....	15
1.4.3 Metodología utilizada para comprobar la hipótesis de partida .....	16
<b>1.5 Estructura del trabajo de investigación</b> .....	<b>17</b>
<b>1.6 Bibliografía del capítulo</b> .....	<b>19</b>
<b>CAPÍTULO 2 - Hacia una definición integradora del crowdsourcing</b> .....	<b>29</b>
<b>2.1 Introducción</b> .....	<b>23</b>
2.1.1 Resumen del artículo.....	23
2.1.2 Datos de la publicación .....	23
<b>2.2 Artículo</b> .....	<b>27</b>
<b>CAPÍTULO 3 - Tipología del crowdsourcing basada en la actividad de la multitud</b> .	<b>50</b>
<b>3.1 Introducción</b> .....	<b>51</b>
3.1.1 Resumen del artículo.....	51
3.1.2 Datos de la publicación .....	51
<b>3.2 Artículo</b> .....	<b>52</b>
<b>CAPÍTULO 4 - Los sistemas de etiquetado social: el caso de Diigo</b> .....	<b>70</b>
<b>4.1 Introducción</b> .....	<b>71</b>
4.1.1 Resumen del artículo.....	71
4.1.2 Datos de la publicación .....	71
<b>4.2 Artículo</b> .....	<b>73</b>
<b>CAPÍTULO 5 - Estudio y análisis de los diferentes tipos de etiquetas que se pueden utilizar en los sistemas de etiquetado social</b> .....	<b>97</b>
<b>5.1 Introducción</b> .....	<b>98</b>
5.1.1 Resumen del artículo.....	98
5.1.2 Datos de la publicación .....	98
<b>5.2 Artículo</b> .....	<b>100</b>
<b>CAPÍTULO 6 - Relación entre el crowdsourcing y la inteligencia colectiva: el caso de los sistemas de etiquetado social</b> .....	<b>126</b>
<b>5.1 Introducción</b> .....	<b>127</b>

5.1.1	Resumen del artículo.....	127
5.1.2	Datos de la publicación.....	127
<b>5.2</b>	<b>Artículo .....</b>	<b>128</b>
<b>CAPÍTULO 7 - Conclusiones y trabajo futuro .....</b>		<b>196</b>
<b>7.1</b>	<b>Introducción .....</b>	<b>150</b>
<b>7.2</b>	<b>Conclusiones .....</b>	<b>150</b>
<b>7.3</b>	<b>Líneas de trabajo futuras .....</b>	<b>152</b>
7.3.1	Relación entre el etiquetado y el crowdsourcing .....	153
7.3.2	Bases teóricas del crowdsourcing .....	153
<b>7.4</b>	<b>Conclusión final.....</b>	<b>153</b>
<b>CAPÍTULO 8 - Bibliografía general .....</b>		<b>155</b>
8.1	Bibliografía general .....	156



## Índice de tablas

Table 2.1. Consulted databases .....	30
Table 2.2. Summary of documents found .....	31
Table 2.3. Collected definitions of crowdsourcing .....	31
Table 2.4. Verification of the definition .....	43
Table 3.1. Composición del repositorio documental .....	55
Table 3.2. Comparación de tipologías: elementos distintivos faltantes .....	59
Table 3.3. Encaje de la nueva tipología con las tipologías estudiadas .....	62
Table 3.4. Contraste de la tipología planteada con los casos seleccionados.....	64
Table 4.1. Examples of academic use of Diigo .....	85
Table 4.2. Comparing SBS .....	87
Table 4.3. Comparative between Diigo, Traditional Bookmark and Delicious (Adapted from Diigo help).....	89
Table 4.4. SWOT analysis .....	92
Table 5.1. List of rejected SBSs.....	107
Table 5.2. A summary chart of those SBSs that were accepted. ....	108
Table 5.3. Summary of the webs collected and number of tags related to each one. By the authors. ....	110
Table 5.4. Use of languages in the analyzed web pages.....	111
Table 5.5. Quantity of webs according to the tags with which they have been marked.....	112
Table 5.6. Data about the use of tags per web according to each SBS.....	112
Table 5.7. Most frequently used tags. ....	113
Table 5.8. Itemization of the collected urls.....	114
Table 5.9. Percentages of implicit and explicit tags. ....	115
Table 5.10. Summary of the quantity of times that explicit tags are used.....	116
Table 5.11. Summary of the quantity of times implicit tags are used. ....	116
Table 5.12. Explicit tags frequently used.....	117
Table 5.13. Implicit tags frequently used.....	118
Table 5.14. Frequency of appearance of the different tags in the corresponding text. ....	119
Table 5.15. HTML tags frequently used. ....	120
Table 6.1. Identification of the collective intelligence elements that appear in the selected STS. ....	136
Table 6.2. Elements of crowdsourcing in the selected SBS. '+': indicates presence of the characteristic; '-': indicates absence of the characteristic .....	139

## Índice de figuras

Figure 4.1. Diigo toolbar.....	80
Figure 4.2. Diigolet virtual toolbar .....	80
Figure 4.3. Button ‘add to Diigo’ .....	80
Figure 4.4. Enhanced linkrolls .....	81
Figure 4.5. Diigo tagrolls .....	81
Figure 4.6. Diigo users’ database by countries (Adapted from Dataopedia.com).....	90
Figure 4.7. Daily traffic during the year 2009 in Diigo and del.iciou.us (Google Trends) .....	91
Figure 5.1. A box and whisker diagram showing the number of tags per marked resources. Outliers and extreme values are hidden in order to appreciate the graphic.....	113
Figure 5.2. Distribution of explicit tags and implicit tags. ....	115

**CAPÍTULO 1 -**  
**Introducción**

## 1.1 Introducción

El auge de las nuevas tecnologías, en concreto todos aquellos avances relacionados con Internet, han permitido el nacimiento de multitud de nuevos procesos y aplicaciones que tienen como base tecnológica la red de redes. La Web 2.0, evolución de la web tradicional (o 1.0) tras la burbuja de las punto com (O'Really, 2007), y las aplicaciones que representan sus principios, son algunos de estos frutos.

El término "Web 2.0" fue acuñado por O'Really (2007) en el año 2005, y, aunque es un término difícil de definir (Cormode & Krishnamurthy, 2008), lo cierto es que ha supuesto un cambio radical en la forma de utilizar Internet. La Web 2.0 hace referencia tanto a una forma de usar la web como a un paradigma tecnológico. Comprende tanto un conjunto de nuevas tecnologías web, como un conjunto de estrategias de negocio así como una serie de tendencias sociales (Murugesan, 2007).

Como nueva forma de usar la web, la Web 2.0 ha supuesto la utilización de la web como plataforma, dejando de lado las aplicaciones de escritorio. Ha supuesto un catalizador para la colaboración entre usuarios, la interoperabilidad y la compartición de información.

Como paradigma tecnológico, la Web 2.0 ha dado pie a un conjunto de nuevas tecnologías web como son los blogs, los sistemas de RSS (*Really Simple Syndication*), el uso de etiquetas, folksonomías y nubes de etiquetas, las wikis (i.e.: la Wikipedia), los sistemas de etiquetado social (i.e.: Delicious), los mashups (i.e.: HousingMaps) o las redes sociales (i.e.: Facebook) entre otras.

Estas tecnologías comparten muchas características, pero una de las principales es que el usuario de Internet se convierte en su razón de ser, en la pieza fundamental de su funcionamiento (Emory, 2007). Una de las principales consecuencias es que, a diferencia de la web 1.0, donde los usuarios tenían un papel pasivo y eran meros espectadores y consumidores de información, ahora estos mismos usuarios toman un papel activo y se convierten en co-productores y co-creadores del contenido (O'Really, 2007).

### 1.1.1 La inteligencia colectiva

Al permitir, promover y facilitar tanto la colaboración entre los usuarios como su participación en la creación de contenidos, lo que se ha producido es la aparición en Internet de la inteligencia colectiva. Ésta es definida por Lévy (2001) como una inteligencia

distribuida de manera universal, constantemente mejorada, coordinada en tiempo real y que da como resultado la efectiva movilización de las capacidades.

Un aspecto fundamental de este tipo de inteligencia es que considera que el resultado de la combinación del esfuerzo realizado por un grupo suficientemente grande de individuos puede ser mejor que el realizado por un único experto. Es decir, el grupo es más inteligente que cualquiera de sus miembros.

La inteligencia colectiva, que es un fenómeno relativamente nuevo en Internet, realmente ha existido desde hace mucho tiempo y ha ido desarrollándose en las distintas culturas humanas tanto de manera espontánea como de manera intencional (Leimister, 2010; Murty, Paulini & Maher, 2010).

#### **1.1.1.1 Los genes de la inteligencia colectiva**

Existen en Internet cientos de ejemplos de plataformas que se nutren y que funcionan gracias a la inteligencia colectiva, pero que lo hacen de distintas maneras. En este sentido, Malone et al. (2009) estudiaron más de 250 casos de inteligencia colectiva e identificaron un conjunto de elementos que se combinaban en cada uno de esos casos de manera distinta. A este conjunto de elementos los denominaron los genes de la inteligencia colectiva, constituyendo cada una de las combinaciones de estos elementos un *genoma* distinto. Estos elementos se basan en cuatro preguntas básicas: qué acción debe llevarse a cabo, quién realiza dicha acción, por qué la realiza y cómo la realiza.

Respondiendo a la pregunta de quién realiza la acción, se pueden encontrar dos respuestas distintas: la multitud de Internet (grupo grande y heterogéneo) o un grupo más reducido de personas que se organizan de manera jerárquica (Leimister, 2010).

El porqué, hace referencia a los incentivos que mueven a los participantes, que son en general tres: beneficio económico (dinero), amor (entendido como el disfrute derivado de realizar una tarea y que está relacionado con la motivación intrínseca) y gloria (a través del reconocimiento de otros individuos).

En cuanto a la tarea o acción que debe realizar cada uno de los participantes, Malone et al. (2009) identifican dos tareas fundamentales: *crear* (algún tipo de contenido como texto, código fuente, diseños, etc.) o *decidir* (evaluando y seleccionando alternativas) (Pénin, 2008; Leimeister, 2010).

Por último, a la pregunta de cómo realizar las tareas, se puede responder que de manera dependiente o de manera independiente. Relacionado con qué tarea se realiza, Georgi & Jung (2012) indican que si una tarea consiste en crear algo de manera independiente, puede hacerse mediante la elaboración de una colección o mediante una competición. Si la creación es dependiente, la colaboración es la única opción. Si la tarea implica la toma de decisiones de manera dependiente, la votación, el consenso, la media o la predicción son las opciones posibles, mientras que si la toma de decisiones se realiza de manera independiente, ésta se basa en una decisión individual.

### **1.1.2 El crowdsourcing**

Otra de las consecuencias del uso de la Web 2.0 y de la centralidad e importancia que adquiere el usuario de Internet, es la proliferación de distintos procesos que se basan en la multitud de Internet, y por lo tanto en la inteligencia colectiva, como la innovación abierta (Reindhart et al., 2010; Chesbrough, 2003), la co-creación (McLoughlin & Lee, 2007) o el crowdsourcing (Howe, 2006, 2008), por ejemplo. De entre todos estos conceptos, en los últimos años, el crowdsourcing está adquiriendo especial relevancia.

#### **1.1.2.1 Una primera aproximación**

El término "crowdsourcing" fue acuñado en el año 2006 por el periodista norteamericano Jeffrey Howe. Howe (2006) definió entonces el término como "el acto, iniciado por una empresa o institución, que tiene como objetivo externalizar una tarea, normalmente realizada por un empleado, a un grupo de individuos grande e indefinido mediante una convocatoria abierta". A lo largo de esta tesis, a la empresa que propone la tarea se le denominará *crowdsourcer* mientras que a los miembros de la multitud que realicen esa tarea se les denominará *crowdworkers*.

Este acto definido por Jeff Howe (2006) puede aplicarse de distintas maneras en función de la tarea a externalizar. Este mismo autor (Howe, 2008) desarrolló una primera tipología del crowdsourcing en la que diferenciaba un total de cuatro tipos de iniciativas, cada una con su propia denominación. De esta manera, Howe diferenciaba entre iniciativas de:

- 1 *Crowdwisdom*, relacionadas con la inteligencia colectiva, que divide en:
  - a Predicción de mercados, que consiste en averiguar el resultado de eventos a partir del conocimiento de la multitud plasmado físicamente en la compra-venta de acciones.

- b Competiciones o *crowdcasting*, donde, tras proponer un problema o reto a la multitud, se premia únicamente a aquel que primero lo resuelva.
  - c Tormenta de ideas en línea o *crowdstorming*, que es una tormenta de ideas pero con una participación masiva.
- 2 *Crowdproduction*, relacionadas con tareas creativas y que buscan la obtención de un producto (ya sea en forma de diseño de un coche, en forma de un post de un blog, o de un logotipo, por ejemplo).
  - 3 *Crowdvoting*, destinadas a recoger la opinión de los usuarios sobre un diseño, una prenda de ropa, etc.
  - 4 *Crowdfunding*, que son tareas que buscan la obtención de fondos económicos.

Aunque esta tipología ha quedado obsoleta ya que algunos de estos tipos se superponen, hecho destacado ya por el propio Howe (2008), supuso un primer acercamiento hacia los distintos tipos de iniciativas de crowdsourcing. De hecho, alguno de los tipos propuestos ha quedado afianzado como un tipo claro de crowdsourcing, tal es el caso de las iniciativas de *crowdfunding*. En este tipo de iniciativas, la tarea propuesta a los *crowdworkers* es la de realizar una aportación económica de una cantidad determinada a un proyecto propuesto por un *crowdsourcer*. A cambio de esta aportación, el *crowdsourcer* dará una recompensa acorde a la cantidad desembolsada (a mayor desembolso, mejor recompensa).

### **1.1.2.2 La evolución del crowdsourcing**

Desde que el término se acuñó en 2006, el crowdsourcing ha ido evolucionando y creciendo rápidamente. A nivel empresarial, cientos de plataformas centradas en alguno de los tipos de crowdsourcing han comenzado a surgir en todo el mundo. En Estados Unidos la plataforma de crowdfunding Kickstarter, fundada en 2009, consiguió reunir \$99.344.382 para financiar 11.836 proyectos (Kickstarter, 2012).

España no ha sido una excepción en este caso, y en estos últimos años se ha dado un incremento muy grande en la creación de plataformas, sobretodo de crowdfunding (Estellés-Arolas, 2012a). A modo de ejemplo, y salvando las distancias, la plataforma española de crowdfunding Lánzanos ha conseguido en sus dos primeros años de funcionamiento la financiación de 175 proyectos con un desembolso de 1.200.000 € (Lánzanos, 2012).

A nivel científico, el estudio y el interés del crowdsourcing ha ido plasmándose en la publicación de artículos en revistas especializadas, conferencias, libros, prensa, etc. tratando

el crowdsourcing desde distintos enfoques: desde el punto de vista de la empresa (Vukovic & Bartolini, 2010), de la biblioteconomía (Oomen & Aroyo, 2011), del marketing (Parvanta, Roth & Keller, 2013) o del humanitario (Sutherlin, 2013), por ejemplo.

En muchos de estos artículos, debido a la juventud del término, el crowdsourcing se ha confundido e identificado de manera unívoca con algunos de los procesos surgidos bajo el paraguas de la Web 2.0 antes mencionados. Aunque no es posible realizar una identificación unívoca entre estos y el crowdsourcing, sí que es cierto que el crowdsourcing se nutre de los mismos, y en función de su forma de aplicación, puede adoptar esas formas.

Por otro lado, al ser la Web 2.0 la base tecnológica sobre la que descansa el crowdsourcing (Vukovic, Mariana & Laredo, 2009), en muchos casos también se ha tendido a identificar de manera errónea determinadas plataformas Web 2.0 como plataformas de crowdsourcing, tal es el caso de Delicious (Geiger, Seedorf & Schader, 2011) o YouTube (Huberman et al., 2009).

### **1.1.3 Los sistemas de etiquetado social**

Los sistemas de etiquetado social son aplicaciones web en las que los usuarios pueden subir, etiquetar y compartir recursos (ya sean páginas web, videos, fotos, etc.) con otros usuarios (Marinho et al., 2012).

Estas etiquetas, que adquieren la forma de metadatos (Subramanya & Liu, 2008), son cadenas de texto generadas libremente por los usuarios que forman palabras, frases o combinaciones de símbolos y caracteres alfanuméricos (Millen et al., 2007). Cuando un conjunto de estas etiquetas es asignado a un recurso por parte de un grupo de usuarios, estas etiquetas pasan a formar lo que se denomina *folksonomía* (Illig et al., 2011; Mathes, 2004).

Las etiquetas utilizadas por los distintos sistemas de etiquetado social, independientemente del tipo de contenido que etiqueten, pueden tener distintas funcionalidades. Golder and Huberman (2005) identifican siete:

- 1 Identificar sobre qué o quién trata el recurso.
- 2 Identificar de qué recurso se trata.
- 3 Identificar quién ha marcado dicho recurso.
- 4 Refinar categorías.
- 5 Identificar o resaltar cualidades o características.
- 6 Organización de tareas.



7 Uso personal (i.e. miscosas, miscomentarios).

Otros autores, como Millen et al. (2007) o Kořner et al. (2010), hacen una clasificación más general indicando que las etiquetas tienen básicamente dos funcionalidades: categorizar y describir.

Las ventajas que implica el uso del etiquetado social son múltiples: permite categorizar contenidos de una manera flexible (a diferencia de las taxonomías), establece relaciones entre el contenido que etiqueta y la persona que lo etiqueta, permite un nuevo tipo de búsqueda más efectiva (el *pivot browsing*), por ejemplo. Además, una de las ventajas más importantes del etiquetado social, es el hecho de que un usuario no necesita formación previa para utilizarlo. Sin embargo, el uso de etiquetas por parte de distintos usuarios da lugar a una falta de homogeneidad y una falta de acuerdo en cómo definir dichas etiquetas.

Esta situación lleva inevitablemente a la aparición de ambigüedades (Mathes, 2004) que se manifiestan de dos maneras. En primer lugar a través de la redundancia de información. Esta redundancia se produce principalmente debido a que distintos usuarios etiquetan con diferentes palabras un mismo recurso a través de la sinonimia (múltiples etiquetas del mismo concepto), la homonimia (misma etiqueta con diferente significado) y la polisemia (misma etiqueta con múltiples significados relacionados) (Golder & Huberman, 2005).

La segunda manifestación de ambigüedades se corresponde con el uso de etiquetas excesivamente específicas (i.e., “!fic”, “#cm10conf” o “#mn1010”). Estas etiquetas no serán comprensibles para otros usuarios y limitarán la efectividad del etiquetado colaborativo en la descripción y recuperación de documentos (Yeung, Gibbins & Shadbolt, 2009).

## 1.2 Antecedentes

La temática de esta tesis surge a partir del proyecto denominado “*Metal 2.0: Crowdsourcing*” (<http://www.metal20.org/>), nacido en el Instituto Tecnológico Metalmecánico de Valencia (AIMME), en el cual colaboraron tanto el Departamento de Organización de Empresas (DOE) de la Universidad Politécnica de Valencia (UPV) como la empresa GMV.

Este proyecto, continuación del proyecto “*Metal 2.0: Viabilidad de las herramientas Web 2.0 en el sector del metal*” (2008-2009), pretendía profundizar en cómo la aplicación de tecnologías de la información basadas en la colaboración, las tecnologías y aplicaciones Web 2.0, podrían aumentar la competitividad de las empresas. En concreto, se centró en el análisis, difusión y experimentación de las relaciones entre las empresas y su entorno a través de estas herramientas (Metal 2.0, 2011a).

El proyecto comenzó con un estudio del estado del arte del crowdsourcing en el año 2011. Durante este estudio se recogieron más de 200 referencias. Al ser un término tan joven, se recogieron documentos tanto científicos (proceedings de congresos, artículos de revistas y libros), como de carácter más general, así como entradas de blogs o publicaciones en revistas y periódicos. Posteriormente, esta colección ha seguido alimentándose llegando a alcanzar un total de 536 documentos.

Los resultados de este estudio del estado del arte fueron presentados en la jornada “*Metal 2.0 CS: Aplicación del crowdsourcing en las empresas*” el 10 y 11 de noviembre de 2011, donde el doctorando, junto con su director, realizó una presentación sobre el crowdsourcing desde el punto de vista científico y sobre sus elementos básicos (Metal 2.0, 2011b).

Mientras que tanto el doctorando como el director de tesis siguieron colaborando con este proyecto, también comenzaron a investigar de manera paralela sobre el crowdsourcing. Una vez leídos y trabajados muchos de los documentos recogidos en primera instancia, se comenzaron a vislumbrar algunas lagunas y algunos aspectos del crowdsourcing que se consideraron importantes y sobre los que se ha investigado dejando constancia en esta tesis doctoral:

- La existencia de multitud de definiciones del crowdsourcing, realizadas desde distintos enfoques: marketing, negocios, etc.

- La existencia de tipologías de las actividades de crowdsourcing que, aun estando basadas en el mismo criterio, tenían diferente estructura.
- La confusión generalizada, sobre todo en Internet, en la relación existente entre el crowdsourcing y otros términos afines como la innovación abierta o la co-creación.

Fruto del trabajo derivado de esta tesis son, en primer lugar, una serie de artículos publicados en revistas científicas:

- Estellés-Arolas, E., and González-Ladrón-de-Guevara, F. (2013) Relationship between Collective Intelligence and Crowdsourcing: the social tagging systems case. *Computer Supported Cooperative Work* (under review).
- Estellés-Arolas, E., and González-Ladrón-de-Guevara, F. (2012) Towards an integrated crowdsourcing definition. *Journal of Information science*, 38 (2), 189-200.
- Estellés-Arolas, E., and González-Ladrón-De-Guevara, F. (2012). Clasificación de iniciativas de crowdsourcing basada en tareas. *El profesional de la información*, 21(3), 283-291.
- Arolas, E. E. and Ladrón-de-Guevar, F. G. (2012), Uses of explicit and implicit tags in social bookmarking. *Journal of the American Society for Information Science and Technology*, 63(2), 313-322.
- Estellés, E., del Moral, E., & González, F. (2010). Social bookmarking tools as facilitators of learning and research collaborative processes: The Diigo case. *Interdisciplinary Journal of E-Learning and Learning Objects*, 6(1), 175-193.

Además de estos artículos, existen otras producciones científicas como:

- Estellés, E., González, F. (2012) La Colaboración para Innovar en las Organizaciones: Crowdsourcing. In *Organizaciones Virtuales*. Libro editado por la Universidad San Martín De Porres (Perú)
- Estellés Arolas, E., González, F. (2011) Crowdsourcing desde el punto de vista de la empresa: ventajas y desventajas de su aplicación en la resolución de problemas, III Congreso Iberoamericano SOCOTE y VIII Congreso SOCOTE “Tu + TIC = Innovación + Competitividad + Sostenibilidad” Universidad Politécnica de Valencia, 11-12 Noviembre de 2011

Por otro lado, los conocimientos adquiridos han permitido realizar otro tipo de publicaciones como:

- Bonet, S., González, F., Estellés, E. & Megías, J. (2011) El arte del crowdsourcing: Es fácil obtener ayuda a través de Internet si sabes cómo. Ed. Aimme Instituto Tecnológico Metalmecánico
- Entrevista para el artículo “Innovación en red”, en revista de Antiguos Alumnos de la IAE Business School nº 27 en diciembre 2012 (AAIAE, 2012).
- Entrevista para el artículo “Creación colectiva”, en revista “Vídeo Popular” nº 150 para el número de noviembre-diciembre de 2012 (Vídeo Popular, 2012).
- Post invitado “We’re Sitting on a Definition Problem here”, en Daily Crowdsourc el 2 de julio de 2012 (Estellés-Arolas, 2012b).
- Post invitado “El crowdfunding para Pymes y Startups”, en Lance Talent el 21 de enero de 2013 (Estellés-Arolas, 2013a).
- Post invitado "El crowdsourcing y la informática", en el blog de informática de la Universidad Cardenal Herrera - CEU el 30 de mayo de 2013 (Estellés-Arolas, 2013b)
- Blog sobre crowdsourcing con más de 35 entradas (Estellés-Arolas, 2013c)

También se han realizado distintas charlas o conferencias sobre el crowdsourcing y la inteligencia colectiva:

- Moderador y ponente de la mesa redonda "Caleidoscopio del movimiento CROWD" en la sección CROWDFEST, dedicada al crowdfunding, en las jornadas ZINC SHOWER en Madrid (12 de abril de 2013).
- “La Inteligencia Colectiva, ¿y esto qué es?”, en las jornadas “El v@lor de Internet. Comunicación Digital y Nueva Evangelización” en la Universidad Cardenal Herrera CEU (24 de marzo de 2012).
- Conferencia magistral online “Crowdsourcing para la innovación educativa”, dentro del I Congreso de Educación Virtual “Más allá de la educación digital” de la Universidad San Martín de Porres (Perú).
- Ponencia online “Aplicación del Crowdsourcing en contextos educativos” en la Universidad Nacional de Villarica del Espíritu Santo (Paraguay).

- “Crowdsourcing desde el punto de vista científico”, dentro de la jornada Metal 2.0 Crowdsourcing: aplicación del Crowdsourcing en las empresas llevadas a cabo en AIMME (11 de noviembre de 2010).

## 1.3 Objetivos

### 1.3.1 Objetivo principal

Como se verá en la sección de metodología, el primer paso que se llevó a cabo para la realización de esta tesis doctoral fue una primera revisión bibliográfica. Este primer acercamiento al término *crowdsourcing* permitió vislumbrar importantes lagunas: la existencia de múltiples definiciones del término con un grado de complementariedad variable, múltiples tipologías, opiniones contradictorias de distintos autores sobre qué era crowdsourcing y qué no, etc.

De esta manera, surge una hipótesis de trabajo. Existen distintos procesos cercanos y relacionados con el crowdsourcing que distintos autores equiparan en distinta medida con el propio crowdsourcing. Algunos de estos procesos son la innovación abierta, la co-creación o la inteligencia colectiva entre otros. Centrándonos en este último caso, se plantea la hipótesis de qué relación existe entre ambos procesos. Éste sería el objetivo principal: ¿es cualquier tipo de iniciativa de crowdsourcing una manifestación de inteligencia colectiva? ¿existe una relación entre ambos? Si es así, ¿qué tipo de relación y en qué grado?

Para poder determinar esta relación, se ha procedido al estudio y análisis comparativo de un tipo de plataforma Web 2.0 perteneciente, de forma clara, al ámbito de la inteligencia colectiva que suele confundirse con una plataforma destinada al uso del crowdsourcing: los sistemas de etiquetado social.

Dentro del estudio de los sistemas de etiquetado social, es importante destacar la profundización realizada en el uso que se hace de las etiquetas implícitas y explícitas por parte de los usuarios. En este sentido se ha profundizado en las características que suelen presentar las etiquetas explícitas: dentro de qué etiquetas HTML suelen aparecer, qué tipo de palabras son (adjetivos, sustantivos, etc.), cuál es el idioma predominante, etc.

### 1.3.2 Objetivos secundarios

Dado que el crowdsourcing se encuentra en su infancia, en el proceso de diferenciarlo de la inteligencia colectiva, nos hemos visto obligados a plantear algunos objetivos intermedios, que han dado lugar a varias investigaciones previas con sus correspondientes publicaciones:

- Debido a la falta de consenso entre los autores en una definición común para el crowdsourcing, se ha procedido a desarrollar una definición que integre las definiciones

ya propuestas por distintos autores, y que además permita diferenciar el crowdsourcing de cualquier otro tipo de actividad.

- Esta falta de consenso entre los autores, no se da tan sólo en lo referente a la definición del término, sino que también se da en otros aspectos teóricos como el desarrollo de una tipología del crowdsourcing. De esta manera, a partir de las tipologías existentes, se ha procedido a integrar en una única tipología aquellas que se basan en el tipo de tarea a realizar por los *crowdworkers*.

En relación a los sistemas de etiquetado social, se ha profundizado en el uso, características y la opinión de los usuarios sobre uno de los sistemas de etiquetado social más utilizados en el ámbito educativo: Diigo.

## **1.4 Metodología**

En la realización de esta tesis se ha utilizado metodología tanto cualitativa como cuantitativa, utilizando una u otra según el objetivo del artículo lo requiera.

### **1.4.1 Metodología utilizada en la línea del crowdsourcing**

En lo referente al crowdsourcing, ha primado la investigación cualitativa ya que, tal como indica Sampieri (2010), este es el método que debe utilizarse cuando el tema de estudio ha sido poco explorado.

Como se ha introducido en la sección de objetivos, el primer paso dentro de la investigación del crowdsourcing fue una revisión bibliográfica general. Esta primera revisión bibliográfica, dio pie a la elaboración de una hipótesis de trabajo: ¿es todo crowdsourcing un caso de inteligencia colectiva?

El hecho de que la hipótesis surja tras un primer acercamiento al concepto y tras la realización de una revisión de la literatura existente es normal en las aproximaciones cualitativas, tal como afirman Williams, Unrau & Grinnell (2005).

De las otras lagunas identificadas, destacaron dos en concreto que era necesario cubrir para poder confirmar la hipótesis. La primera es la elaboración de una definición clara del crowdsourcing que permitiera diferenciarlo de cualquier otro tipo de proceso y la segunda es la elaboración de una tipología basada en algún criterio claro y concreto que permita clasificar las distintas iniciativas del crowdsourcing.

#### **1.4.1.1 Elaborando una definición integradora**

Con respecto a la definición del término, aunque es cierto que Jeffrey Howe (2006) había definido en un primer momento el concepto, múltiples autores procedieron a redefinirlo posteriormente desde distintas perspectivas, añadiendo nuevos matices. El resultado es la falta de una definición integradora que permita a aquellos que se acercan al crowdsourcing diferenciarlo de cualquier otro concepto. Tras consultar algunos manuales, no se encontró ninguna metodología que permitiera la elaboración de una definición a partir de otras ya existentes. Sin embargo, en el área de la filosofía, se encontró un procedimiento usado por el filósofo e historiador del arte polaco Władysław Tatarkiewicz (1886-1980), el cual desarrolló una definición del concepto “arte” a partir de las definiciones creadas por otros autores (capítulo 2).



Para comprobar la validez de esta definición, se han seguido dos planteamientos. Por un lado el de Aliakbarian et al. (2007), que comprueban la validez de una posible definición de una red P2P verificando que todos los elementos de su definición se cumplen en cinco redes de este tipo. Por otro lado, el de Vukovic (2009), que identifica los requisitos necesarios para desarrollar un servicio de crowdsourcing y posteriormente comprueba que estos requisitos se cumplen en distintos servicios de este tipo para proponer una taxonomía.

#### **1.4.1.2 Elaborando una tipología integradora**

En cuanto a la necesidad de una tipología clara, tanto en la literatura científica como en plataformas de Internet y blogs, han aparecido multitud de tipologías basadas en distintos criterios. El resultado es, de nuevo, multitud de tipologías que se superponen en cierto grado y en función de cual es el criterio aplicado.

En este caso se ha realizado una revisión de la literatura existente buscando las distintas tipologías desarrolladas hasta el momento. De éstas se seleccionaron solo aquellas que se basan en la acción que debe realizar la multitud. Para integrar estas tipologías, en base a Pinto-Molina et al. (2004), se ha utilizado una tabla de doble entrada donde cada tipología es comparada con las demás, indicando qué partes de las distintas tipologías coinciden.

Para comprobar si esta tipología funciona, se seleccionaron distintos casos de plataformas de manera aleatoria del listado que aparece en la wikipedia. La tipología trata de aplicarse, con éxito, en esta selección de plataformas valorando como válida la tipología siempre que agrupe todas las plataformas seleccionadas.

### **1.4.2 Metodología utilizada en el área de los sistemas de etiquetado social**

El interés en este tipo de plataformas, y las investigaciones y publicaciones consecuentes, se deben a que este tipo de sistemas se basan en la inteligencia colectiva y a que distintos autores (Howe, 2008; Bernstein et al., 2010; Geiger et al., 2011; Hirth, Hoßfeld & Tran-Gia, 2010; Huberman et al., 2009) lo identifican como ejemplo de crowdsourcing. En esta área del trabajo de tesis se ha podido proceder con un enfoque un poco más cuantitativo.

#### **1.4.2.1 Conociendo los sistemas de etiquetado social**

El estudio de uno de los sistemas de marcado y etiquetado social más utilizados en el ámbito educativo en Estados Unidos, Diigo, busca profundizar de manera cualitativa en las características que definen este tipo de plataformas.

Este estudio se ha llevado a cabo mediante la identificación de las características propias de estos sistemas, de una descripción de la herramienta en concreto, su comparación con otro sistema de etiquetado social muy popular y mediante un análisis DAFO que indica, según una muestra de 30 usuarios, los aspectos más relevantes y mejorables del sistema.

#### **1.4.2.2 Uso de las etiquetas por parte de los usuarios**

Para estudiar más en profundidad estos sistemas, tras un primer acercamiento, se procedió a estudiar el uso que los distintos usuarios hacen de las etiquetas, elemento fundamental dentro de los sistemas de etiquetado y marcado social.

Mediante el uso de estadística descriptiva, se analizó una muestra de más de 50.000 etiquetas identificando algunas características de las mismas como el tipo de etiqueta más utilizado (si era explícita - aparecía en el contenido marcado y etiquetado - o implícita - no aparecía en el contenido marcado y etiquetado), el idioma predominante, en qué partes de las páginas web suelen aparecer estas etiquetas cuando son explícitas, etc.

#### **1.4.3 Metodología utilizada para comprobar la hipótesis de partida**

Por último, para resolver la hipótesis de trabajo, se ha procedido a comprobar si los sistemas de etiquetado social, herramientas arquetípicas de inteligencia colectiva, pueden considerarse plataformas de crowdsourcing. Con este fin se han tratado de identificar todos los elementos tanto del crowdsourcing (Estellés-Arolas y González, 2012) como de la inteligencia colectiva (Malone, 2009) en tres sistemas de marcado social. Estos tres sistemas de etiquetado social, diferentes en base al contenido que etiquetan, se escogieron por haber sido utilizados por distintos autores como ejemplo de plataforma o herramienta de crowdsourcing (Howe, 2008; Bernstein et al., 2010; Geiger et al., 2011; Hirth, Hoßfeld & Tran-Gia, 2010; Huberman et al., 2009).

## 1.5 Estructura del trabajo de investigación

La presente tesis se ha desarrollado por la modalidad de recopilación de artículos científicos. Cada uno de estos puede ser leído de manera independiente del resto, aunque entre todos existe un nexo común que vertebra este trabajo.

Tras la lectura de estos artículos en el orden presentado en la tesis, el lector será capaz de entender qué es el crowdsourcing, en qué consiste y qué tipos existen, qué son los sistemas de etiquetado social, cuáles son sus principales funcionalidades y cuales son las características de las etiquetas que se utilizan y por último, el lector será capaz de entender porqué un sistema de etiquetado social, como Diigo o Delicious, es un ejemplo de inteligencia colectiva pero no lo es *per se* de crowdsourcing.

De esta manera, la tesis se divide en siete capítulos, cinco de los cuales se corresponden con cinco artículos:

1. Introducción.
2. Definiendo el crowdsourcing (artículo 1).
3. Tipología del crowdsourcing basada en la actividad de la multitud (artículo 2).
4. Los sistemas de etiquetado social: ¿qué son y para qué sirven? (artículo 3).
5. Estudio y análisis de los diferentes tipos de etiquetas que se pueden utilizar en los sistemas de etiquetado social (artículo 4).
6. Relación entre el crowdsourcing y la inteligencia colectiva: el caso concreto de los sistemas de etiquetado social (artículo 5).
7. Conclusiones y trabajo futuro.

El capítulo de introducción es el presente capítulo, donde se plantea la tesis de manera general y se introducen los términos sobre los que versa la tesis doctoral.

El segundo capítulo se corresponde con el artículo *Towards an integrated crowdsourcing definition*. En este artículo, a partir de una colección de definiciones recogidas de distintos artículos, *proceedings* y libros, se construye una nueva definición de carácter integrador de crowdsourcing. En este artículo se podrá obtener una idea precisa de qué es el crowdsourcing y qué elementos implica.

El tercer capítulo se basa en el artículo *Clasificación de iniciativas de crowdsourcing basada en tareas*. En este artículo se realiza la misma tarea integradora que en el primero, pero en este caso con una serie de tipologías distintas de las actividades de crowdsourcing.

El cuarto capítulo se corresponde con el artículo *Social bookmarking tools as facilitators of learning and research collaborative processes: The Diigo case*. Aquí se estudia en profundidad las características y funcionalidades del sistema de marcado y etiquetado social Diigo. Dentro de estas funcionalidades, se hace hincapié en el papel que pueden tener estas aplicaciones Web 2.0 en los procesos colaborativos tanto de aprendizaje como de investigación.

A continuación, en el quinto capítulo se estudian las características de los dos tipos de etiquetas que se utilizan en los sistemas de etiquetado social: las explícitas (etiquetas que aparecen en el contenido textual marcado) y las implícitas (etiquetas que no aparecen en el contenido textual marcado). Este capítulo se corresponde con el artículo *Uses of explicit and implicit tags in social bookmarking*.

El penúltimo capítulo de la tesis, el sexto, se basa en el artículo *Relationship between Collective Intelligence and Crowdsourcing: the social tagging systems case*, que ha sido enviado a una revista y está en proceso de revisión. En este caso, tras haber definido y delimitado el crowdsourcing, y tras haber estudiado los sistemas de etiquetado social y las características de las etiquetas utilizadas, se procede a estudiar la relación entre la inteligencia colectiva y el crowdsourcing, utilizando los sistemas de etiquetado social como ejemplo.

Por último, en el séptimo capítulo, se procederá a enumerar algunas conclusiones extraídas de todo el trabajo realizado y plasmado en esta tesis. También se comentarán tanto los trabajos en proceso que el doctorando lleva actualmente a cabo así como aquellas líneas de investigación en las que se podría trabajar.

## 1.6 Bibliografía del capítulo

- Aliakbarian, S., Rahimabadi, A.M., Sadeghi, P.H. and Mirsatari, N.S. (2006) Neighbor Definition in P2P Networks. In: *Proceedings of 2006 International Conference on Communications, Circuits and Systems*. (Guilin, 2007) 1562-1565.
- Bernstein, M.S., Tan, D., Smith, G., Czerwinski, M. and Horvitz, E. (2010) Personalization via friendsourcing. *ACM Transactions on Computer-Human Interaction*, 17(2): 1-28.
- Chesbrough, H. (2003). *Open Innovation: The new Imperative for Creating and Profit from Technology*. Boston: Harvard Business School Press.
- Cormode, G. and Krishnamurthy, B. (2008) Key differences between Web 1.0 and Web 2.0. *First Monday*, 13(6).
- Emory, M. C., (2007) Changing paradigms: managed learning environments and Web 2.0. *Campus-Wide Information Systems*, 24 (3), 152-161.
- Estellés-Arolas, E. (2012a) *Situación del crowdsourcing en España*. Crowdsourcing Blog. Recuperado el 10 de abril de 2013, de <http://www.crowdsourcing-blog.org>
- Estellés-Arolas, E. (2012b) *We're Sitting on a Definition Problem here*. Daily Crowdsourc. Recuperado el 10 de abril de 2013, de <http://dailycrowdsourc.com/crowdsourcing/articles/opinions-discussion/1180-we-re-sitting-on-a-definition-problem-here>
- Estellés-Arolas, E. (2013a) *El crowdfunding para Pymes y Start-ups*. LanceTalent. Recuperado el 10 de abril de 2013, de <http://www.lancetalent.com/blog/el-crowdfunding-para-pymes-y-startups/>
- Estellés-Arolas, E. (2013b) *El crowdsourcing y la informática*. Blog de informática de la Universidad CEU-Cardenal Herrera. Recuperado el 30 de mayo de 2013, de <http://blog.uchceu.es/informatica/que-es-el-crowdsourcing/>
- Estellés-Arolas, E. (2013c) Crowdsourcing Blog, things about crowdsourcing. Recuperado el 30 de mayo de 2013, de <http://www.crowdsourcing-blog.org>
- Geiger, D., Seedorf, S. and Schader, M. (2011) Managing the Crowd: Towards a Taxonomy of Crowdsourcing Processes. In: *Proceedings of the Seventeenth Americas Conference on Information Systems*, Detroit, Michigan August 4th-7th 2011.
- Georgi, S. and Jung, R. (2012) Collective Intelligence Model: How to Describe Collective Intelligence. In: J. Altmann, U. Baumöl, & B. J. Krämer (Eds.), *Advances in*

Collective Intelligence 2011, *Advances in Intelligent and Soft Computing*. 113: 53–64). Springer Berlin / Heidelberg.

- Golder, S. A. and Huberman, B. A. (2005) *The Structure of Collaborative Tagging Systems*. HP Labs technical report, 2005.
- Grinnell, R.M., Unrau, Y.A. and Williams, M. (2005) *The Qualitative Research Approach*. In: Grinnell, R.M. & Unrau, Y.A. (eds) *Social work research and evaluation. Quantitative and qualitative approaches (7th ed)*. Oxford: University Press.
- Hernández Sampieri, R., Fernández Collado, C., and Baptista Lucio, P. (2007). *Fundamentos de metodología de la investigación*. Editorial Mcgraw-Hill.
- Hirth, M., Hoßfeld, T., and Tran-Gia, P. (2010) *Cheat-detection mechanisms for crowdsourcing*. Technical report, University of Würzburg.
- Howe, J. (2006) *The rise of crowdsourcing*. *Wired*, 14(6).
- Howe, J. (2008) *Crowdsourcing: How the Power of the Crowd is Driving the Future of Business*. Great Britain: Business Books.
- Huberman, B.A., Romero, D.M. and Wu F. (2009) *Crowdsourcing, Attention and Productivity*. *Journal of Information Science*, 35(6), 758–765.
- Illig, J., Hotho, A., Jäschke, R., & Stumme, G. (2011). *A comparison of content-based tag recommendations in folksonomy systems*. In *Knowledge Processing and Data Analysis* (pp. 136-149). Springer Berlin Heidelberg.
- Kickstarter (2012) *2011: The Stats*. Recuperado el 10 de abril de 2013, de <http://www.kickstarter.com/blog/2011-the-stats>
- Körner, C., Benz, D., Hotho, A., Strohmaier, M., and Stumme, G. (2010). *Stop thinking, start tagging: tag semantics emerge from collaborative verbosity*. In: *Proceedings of the 19th international conference on World wide web*, pages 521-530, New York, NY, USA. ACM.
- Lánzos (2012) *Lánzos en cifras*. Recuperado el 10 de abril de 2013, de <http://www.lanzanos.com/blog/entry/26/Lanzanosencifras/>
- Leimeister, J. (2010) *Collective Intelligence*. *Business & Information Systems Engineering*, 2(4), 245–248.
- Lévy, P. (2001) *Collective intelligence*. *Reading digital culture*, 4: 253
- Malone, T. W., Laubacher, R. and Dellarocas, C. N. (2009) *Harnessing Crowds: Mapping the Genome of Collective Intelligence*. MIT Sloan; Research Paper No. 4732-09.

- Marinho, L. B., Nanopoulos, A., Schmidt-Thieme, L., Jäschke, R., Hotho, A., Stumme, G., and Symeonidis, P. (2012). Social tagging recommender systems. In *Recommender systems handbook* (pp. 615-644). Springer US.
- Mathes, A. (2004) *Folksonomies - Cooperative Classification and Communication Through Shared Metadata*. Computer Mediated Communication - LIS590CMC. Graduate School of Library and Information Science, University of Illinois Urbana-Champaign
- McLoughlin, C. and Lee, M. J. (2007). Social software and participatory learning: Pedagogical choices with technology affordances in the Web 2.0 era. In *Proceedings ASCILITE*, Singapore 2007.
- Metal 2.0 (2011a) *METAL 2.0 CROWDSOURCING - Web 2.0, redes sociales y crowdsourcing aplicados al sector del metal*. Recuperado el 1 de marzo de 2013, de <http://www.metal20.org/proyecto>
- Metal 2.0 (2011b) Video "*El crowdsourcing desde un punto de vista científico*". Recuperado el 1 de marzo de 2013, de [http://youtu.be/khBTMi2\\_4XA](http://youtu.be/khBTMi2_4XA)
- Millen, D. R., Yang, M., Whittaker, S. and Feinberg, J. (2007). Social bookmarking and exploratory search. In: *Proceedings of the ECSCW 2007* (pp. 21-40). Springer London.
- Murty, P., Paulini, M. and Maher, M.L. (2010) Collective Intelligence and Design Thinking. In: *Proceedings of the Design Thinking Research Symposium, DTRS'10*; Sydney, Australia. 2010.
- Murugesan, S. (2007) Understanding Web 2.0. *IT Professional*, 9(4), 34-41.
- Oomen, J. and Aroyo, L. (2011) Crowdsourcing in the cultural heritage domain: opportunities and challenges. In: *Proceedings of the 5th International Conference on Communities and Technologies* (pp. 138-149). New York, NY, USA: ACM.
- O'Reilly, T. (2007) What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. *Communications & Strategies*, 1,17.
- Parvanta, C., Roth, Y. and Keller, H. (2013). Crowdsourcing 101 A Few Basics to Make You the Leader of the Pack. *Health promotion practice*, 14(2), 163-167.
- Pénin, J. (2008) More open than open innovation? Rethinking the concept of openness in innovation studies. Working papers of BETA, Bureau d'Economie Théorique et Appliquée, UDS, Strasbourg.
- Pinto Molina, M., Alonso Berrocal, J. L., Cerdón García, J. A., Fernández Marcial, V., García Figuerola, C., García Marco, J., ... and Doucet, A. V. (2004). Análisis cualitativo

de la visibilidad de la investigación de las universidades españolas a través de sus páginas web. *Revista española de documentación científica*, 27(3).

- Reinhardt, M., Frieß, R., Groh, G., Wiener, M., and Amberg, M (2010) Web 2.0 driven Open Innovation Networks - A social Network Approach to Support the Innovation Context within Companies. In: *Schumann, M., Kolbe, L., Breiner, M., Frerichs, A. (Eds) Proceedings of the Multikonferenz Wirtschaftsinformatik (MKWI)*, p. 1177-1190, Gottingen, Shipton, H., West, M., Dawson, J.
- Sutherlin, G. (2013). A voice in the crowd: Broader implications for crowdsourcing translation during crisis. *Journal of Information Science*.
- Vukovic, M. and Bartolini C. (2010) Towards a Research Agenda for Enterprise crowdsourcing. In: M. Tiziana and S. Bernhard (eds) *Leveraging Applications of Formal Methods, Verification, and Validation* (Springer, Berlin/Heidelberg, 2010) 425-434 [Lecture Notes in Computer Science 6415].
- Vukovic, M. (2009) Crowdsourcing for enterprises. In: *Proceedings of the 2009 Congress on Services – I*, IEEE Computer Society (Washington, DC, USA 2009). 686-692.
- Vukovic M., Mariana L. and Laredo J. (2009) PeopleCloud for the Globally Integrated Enterprise. In: *D. Asit et al. (eds) Service-Oriented Computing*. (Springer-Verlag, Berlin/Heidelberg, 2009)
- Yeung, C., Gibbins, N., and Shadbolt, N. (2009). Contextualising tags in collaborative tagging systems. In: *Proceedings of the 20th ACM conference on Hypertext and hypermedia*, pages 251-260, New York, NY, USA. ACM.



**CAPÍTULO 2 -**

**Hacia una definición integradora  
del crowdsourcing**

## 2.1 Introducción

Este capítulo se corresponde con el artículo “*Towards an integrated Crowdsourcing definition*” publicado en la revista *Journal of Information Science*.

### 2.1.1 Resumen del artículo

Debido a la juventud del crowdsourcing y a que posee unos límites no siempre claramente delimitados, el término ha recibido multitud de definiciones por parte de diferentes autores. Con el objetivo de conseguir una definición de crowdsourcing que integre al resto, facilitando así el acercamiento al término, en el siguiente artículo se analizan un total de 40 definiciones extrayendo de todas ellas un total de 8 características que definen el crowdsourcing y lo diferencian de cualquier otro fenómeno.

### 2.1.2 Datos de la publicación

El artículo ha sido publicado en la revista *Journal of Information Science*, revista internacional que se ocupa de temas de interés para todos aquellos que investigan y trabajan las ciencias de la información y la gestión del conocimiento. La revista está indexada tanto en *Social Science Citation Index*, como en *Science Citation Index*, así como en *Scopus*. Se encuentra en distintas bases de datos como *Academic Search Premier*, *Francis*, *Business Source Elite*, *Information Science and Technology Abstracts*, *Library and Information Abstracts* y *Library Literature and Information Science*.

Esta revista tuvo en 2011 un índice de impacto JCR de 1.299, encontrándose la revista, según el JCR Science Edition, en la posición 46/135 en la categoría de “Informática y Sistemas de Información”, y, según el JCR Social Science Edition, en la posición 24/86 en la categoría de “Ciencias de la Información y Biblioteconomía”. Ocupa en ambos casos el segundo cuartil.

Los autores del artículos son, en orden de aparición, Enrique Estellés-Arolas y Fernando González Ladrón-de-Guevara.

- Nombre de la revista: *Journal of Information Science*
- Editor: SAGE Journals
- ISSN: 0165-5515
- Fecha: Abril de 2012
- Volumen: 38
- Nº: 2
- Páginas: 189-200

Se considera relevante el hecho de que, desde su publicación en abril de 2012, el artículo ha sido citado por 33 fuentes distintas (21 en 2012 y 12 en 2013), entre artículos, proceedings, tesis de doctorado y máster e informes técnicos. Por otro lado, la definición ha sido utilizada en distintos sitios web y conferencias.

Artículos que lo citan:

1. Hetmank, L. (2013) Components and Functions of Crowdsourcing Systems – A Systematic Literature Review. In: *Wirtschaftsinformatik Proceedings 2013*. Paper 4. Recuperado el 25 de abril de 2013, de <http://aisel.aisnet.org/wi2013/4>
2. Allwood, J., and Dhakhwa, S. (2013) Working prototype of a multimodal dictionary. *Nepalese Linguistics*, 27, 1-8
3. Jones, P., Comfort, D., and Hillier, D. (2013). Crowdsourcing corporate sustainability strategies. *International Journal of Business and Globalisation*, 10(3), 345-356.
4. Evaldsson, J., Ljungdahl, T., & Suter, F. (2013) The emergence of crowdsourcing and open source models in drug development. Master's Thesis, Blekinge College of Technology, Karlskrona, Sweden.
5. Simula, H. (2013). The Rise and Fall of Crowdsourcing?. In *System Sciences (HICSS), 2013 46th Hawaii International Conference on* (pp. 2783-2791). IEEE.
6. Zapico, J. L. (2013). Hacking for sustainability. Doctoral Thesis, KTH Royal Institute of Technology. Stockholm, Sweden.
7. Kuokkanen, S. (2013). Joukkoistaminen ja sen hyödyntäminen huonekalualan yrityksessä: case: Isku-Yhtymä Oy. Bachelor's Thesis. Lahti University of Applied Sciences. Lahti, Finland.
8. Hansen, D. L., Schone, P. J., Corey, D., Reid, M., & Gehring, J. (2013). Quality control mechanisms for crowdsourcing: peer review, arbitration, & expertise at familysearch indexing. In *Proceedings of the 2013 conference on Computer supported cooperative work* (pp. 649-660). ACM.
9. Ei Chew, H., Sort, B., & Haddawy, P. (2013). Building a crowdsourcing community: how online social learning helps in poverty reduction. In *Proceedings of the 3rd ACM Symposium on Computing for Development* (p. 21). ACM.
10. Papakonstantinou, A., & Bogetoft, P. (2013). Crowd-sourcing with uncertain quality-an auction approach. MPRA paper n° 44236
11. Shapiro, D. N., Chandler, J., & Mueller, P. A. (2013). Using Mechanical Turk to Study Clinical Populations. *Clinical Psychological Science*.
12. Kingston, L. N., & Stam, K. R. (2013). Online Advocacy: Analysis of Human Rights NGO Websites. *Journal of Human Rights Practice*, 5(1), 75-95.
13. Fernie, K., & Kouloumpis, T. (2013) D1. 4 Final State of Art Monitoring Report. EU PROJECT “Personalised Access To Cultural Heritage Spaces”, No. ICT-2009-270082

14. Ching, A., Zegras, C., Kennedy, S., & Mamun, M. (2013). A User-Flocksourced Bus Experiment in Dhaka: New Data Collection Technique with Smartphones. *Transportation Research Record: Journal of the Transportation Research Board*.
15. Saengkhattiya, M., Sevandersson, M., & Vallejo, U. (2013) Quality in Crowdsourcing. *Organization*, 17(6), 903-914.
16. Bittencourt, N., & Runeberg Schultz, G. (2012). Crowdsourcing: hur motiveras deltagande och vad innebär det för innovation?: En kvalitativ studie baserad på Etsy. com. Doctoral dissertation, Umeå University. Umeå, Sweden.
17. Yu, Z., Zhang, D., Yang, D., & Chen, G. (2012). Selecting the Best Solvers: Toward Community Based Crowdsourcing for Disaster Management. In *Services Computing Conference (APSCC), 2012 IEEE Asia-Pacific* (pp. 271-277). IEEE.
18. Cruz, A. (2012). Social BPM—The role of social networks in business process management. Tesis de master. Universidad Nova de Lisboa. Lisboa, Portugal.
19. McKinley, D. (2012) Practical management strategies for crowdsourcing in libraries, archives and museums. Technical Report. School of Information Management, Faculty of Commerce and Administration. Victoria University of Wellington. Recuperado el 25 de abril de 2013, de <http://www.digitalglam.org/crowdsourcing/crowdsourcing-strategies/>
20. Garner, K. (2012). Ripping the pith from the Peel: Institutional and Internet cultures of archiving pop music radio. *Radio Journal: International Studies in Broadcast & Audio Media*, 10(2), 89-111.
21. Dhakhwa, S., & Allwood, J. (2012). Self documentation of endangered languages. In: *Chinese Spoken Language Processing (ISCSLP), 2012 8th International Symposium on* (pp. 392-395). IEEE.
22. Korthaus, A., & Dai, W. (2012, September). Crowdsourcing in Heterogeneous Networked Environments-Opportunities and Challenges. In: *Network-Based Information Systems (NBIS), 2012 15th International Conference on* (pp. 483-488). IEEE.
23. Kärkkäinen, H., Jussila, J., & Multasuo, J. (2012, October). Can crowdsourcing really be used in B2B innovation?. In: *Proceeding of the 16th International Academic MindTrek Conference* (pp. 134-141). ACM.
24. Väättäjä, H., Vainio, T., & Sirkkunen, E. (2012). Location-based crowdsourcing of hyperlocal news: dimensions of participation preferences. In *Proceedings of the 17th ACM international conference on Supporting group work* (pp. 85-94). ACM.
25. Riitamaa, T. (2012). Sosiaalinen media hunajakennona—toimivan vuorovaikutuksen edellytykset. Bachelor Thesis. Haaga-Helia, University of Applied Sciences. Helsinki, Finland.
26. Gritti, A. (2012). Crowd outsourcing for software localization. Master's Thesis. Technical University of Catalunya. Barcelona, Spain.

27. Häsel, M., Quandt, T., & Vossen, G. (2012). Social, Supply-Chain, Administrative, Business, Commerce, Political networks: a multi-discipline perspective (Dagstuhl Perspectives Workshop 12182).
28. Skaržauskaitė, M. (2012). The Application of Crowd Sourcing in Educational Activities. *Social Technologies*, 2.
29. Palmer, M., & Nicey, J. (2012). Social Media and the Freedom of the Press: a long-term Perspective from within International News Agencies (AFP, Reuters). *ESSACHESS–Journal for Communication Studies*, 5(1 (9)), 107-124.
30. Cai, Y., Theng, Y. L., Cai, Q., Ling, Z., Ou, Y., & Theng, G. (2012). Crowdsourcing Metadata Schema Generation for Chinese-Style Costume Digital Library. *The Outreach of Digital Libraries: A Globalized Resource Network*, 97-105.
31. Simula, H., & Vuori, M. (2012). Benefits and barriers of crowdsourcing in B2B firms: generating ideas with internal and external crowds. *International Journal of Innovation Management*, 16(06).
32. Zhao, Y., & Zhu, Q. (2012). Evaluation on crowdsourcing research: Current status and future direction. *Information Systems Frontiers*, 1-18.
33. Bazilian, M., Rice, A., Rotich, J., Howells, M., DeCarolis, J., Macmillan, S., ... & Liebreich, M. (2012). Open source software and crowdsourcing for energy analysis. *Energy Policy*, 49, 149-153.

## 2.2 Artículo

### Towards an integrated crowdsourcing definition

Enrique Estellés-Arolas

Department of Management, Technical University of Valencia, Valencia, Spain

Fernando González-Ladrón-de-Guevara

Department of Management, Technical University of Valencia, Valencia, Spain

#### **Abstract**

“Crowdsourcing” is a relatively recent concept that encompasses many practices. This diversity leads to the blurring of the limits of crowdsourcing that may be identified virtually with any type of Internet-based collaborative activity, such as co-creation or user innovation. Varying definitions of crowdsourcing exist and therefore, some authors present certain specific examples of crowdsourcing as paradigmatic, while others present the same examples as the opposite. In this paper, existing definitions of crowdsourcing are analyzed to extract common elements and to establish the basic characteristics of any crowdsourcing initiative. Based on these existing definitions, an exhaustive and consistent definition for crowdsourcing is presented and contrasted in eleven cases.

#### **Keywords**

Crowdsourcing; definition; innovation

## 1. Introduction

As indicated by Jeff Howe [1], the word crowdsourcing is used for a wide group of activities that take on different forms [2, 3]. The adaptability of crowdsourcing allows it to be an effective and powerful practice, but makes it difficult to define and categorize. Moreover, the theoretical knowledge base is still not solid, being developed with works like Brabham’s, in which he defines crowdsourcing [4] and creates a typology of it [5]; Vukovic’s, in which she makes a general overview of various characteristics of crowdsourcing including the kind of crowd that can participate, the incentive schema, the different variants of crowdsourcing initiatives [2], or the requirements of a crowdsourcing initiative [6]; or Geiger’s [7], in which he develops a taxonomy using different examples. Nor is there an agreed definition; instead there are a variety of definitions, which look at crowdsourcing from differing points of view

including problem resolution [8, 9] or innovation applied to business process improvement [10, 4].

Depending upon the perspective and the definition used, certain initiatives classified by some authors as crowdsourcing, are not classified as such by others. For example, Buecheler et al. [11] consider Wikipedia to be an example of crowdsourcing, as Huberman et al. [12] do of YouTube, while Kleeman et al. [13] declare the opposite in both cases. The abundance of definitions also means that crowdsourcing cannot be coherently classified, as occurs in Andriole [14], where crowdsourcing is identified with other Web 2.0 technologies.

In the search for a common definition, an etymological analysis does not prove to be useful. The name crowdsourcing is formed from two words, crowd, making reference to the people who participate in the initiatives, and the word sourcing, which refers to a number of procurement practices aimed at finding, evaluating, and engaging suppliers of goods and services. Following this approach, authors such as Jeff Howe affirm that crowdsourcing “is a business practice that means literally to outsource an activity to the crowd” [15]. However, to adopt the etymological significance as a definition is too discriminatory [1].

The objective of this article is to form an exhaustive and global definition to describe any given crowdsourcing activity. In order to obtain this definition, existing definitions in the literature will be analyzed.

Furthermore, the elements required to obtain a clear idea of the minimum conditions that need to be completed by a crowdsourcing initiative are identified. This definition also allow us to:

1. Distinguish those activities that can be considered crowdsourcing.
2. Formalize an incipient theoretical base for crowdsourcing [16].

## **2. Methodology**

The methodology used to obtain a global definition for crowdsourcing follows three stages: the search for documentation on crowdsourcing via a systematic review of the literature with its corresponding filter, the creation of an exhaustive definition based on commonly detected elements, and the testing of its validity.

## **2.1. Search for information and filtering of documents**

A systematic review of the literature is undertaken, following the Delgado approach [17] based on Petitti and Egger et al. [18,19]. After selecting four databases and establishing concrete search criteria, documents are searched for to form an initial repository. The repository is expanded to include those documents referenced in the most prolific author's articles and those documents that reference the most cited author. For the filtering of the documents, only those with an original definition for crowdsourcing are selected.

This search was conducted between January and August 5, 2011.

## **2.2. Preparation**

To create a cohesive definition, Tatarkiewicz's approach is followed [20]. Tatarkiewicz was a Polish philosopher and historian of art and philosophy who developed a global definition of the concept "art" from definitions created by other authors. After collecting all definitions, Tatarkiewicz set aside all of them that were centered on particular manifestations of art. The reason was that these could not be a total reconstruction of the concept, taking into account only certain features while ignoring the rest. Next, a definition that encompasses all the other definitions was obtained through the union of sentences referring to the intention and effect of the art.

Also taken into account was the work of Cosma and Joy [21] that utilizes a survey to achieve a definition of "source-code plagiarism" by extracting elements that can be later combined to form a definition.

In this paper, from the original definitions of crowdsourcing, the elements designated by Tatarkiewicz as *differentia specifica* are obtained. These include elements whose characteristics differentiate crowdsourcing from other collaborative activities based on ICT.

## **2.3. Integrating crowdsourcing definition**

The elements designated as *differentia specifica* are transformed from the authors' points of view into a conceptual perspective. In this way, the final components of the definition are obtained [19] and the integrating definition is stated.

## **2.4. Verification**

To check the validity of the definition, the approaches of Vukovic [6] and Aliakbarian et al. [22] will be followed. In Aliakbarian et al. [22], to verify the definition proposed for "P2P network", the definition is applied to five cases checking if all the elements of the definition



are satisfied. In Vukovic [6], the requirements for the development of a general-purpose crowdsourcing service in the Cloud are analyzed. Then, a taxonomy is proposed for the categorization of crowdsourcing platforms through the evaluation of cases against the set of identified features.

In this paper, the formulated definition is applied to eleven Internet initiatives (some considered crowdsourcing, others not) to see if the definition discriminates correctly, taking into account in each case the presence of the distinctive characteristics. An initiative will be considered a real crowdsourcing initiative if all the distinctive characteristics are present.

### 3. Results

In this section, the results obtained over the previous stages are described: the information sources consulted, document filter criteria, identified elements and characteristics, formulated definition, and formulated definition verification.

#### 3.1. Information search and filtering of documents

For the information search, six databases are consulted: ACM, IEEE, ScienceDirect, SAGE, SpringerLink, and Emerald using search criteria with “crowdsourcing” as one of the keywords. Of these, SpringerLink is set aside because it was not possible to search solely via keyword. The first search resulted in 132 documents (Table 2.1).

**Table 2.1. Consulted databases**

Document type	ACM	IEEE	Science Direct	SAGE	Emerald	Total
Conference paper	81	30	0	0	0	111
Journal paper	0	6	8	7	34	55
TOTAL	81	36	8	7	34	166

To complete this document repository, all of those documents that made reference to the most cited document [4] are searched, as are all the references of the most prolific author, Maja Vukovic. Of these, those with the word “crowdsourcing” in the title are added to the document repository, with 30 from the first group and 13 from the second. Using this approach, 43 new documents are added to make a final document repository of 209 documents. A summary of these documents can be seen in Table 2.2.

From these 209 documents, 40 original definitions of crowdsourcing were found, which appear in Table 2.3. The most frequently cited definitions are the ones proposed by Howe [1], Brabham [23], and Wikipedia [24].

**Table 2.2. Summary of documents found**

Document type	Search #1	Search #2	Total
Conference paper	111	16	127
Journal paper	55	13	68
Workshop	0	3	3
Book	0	1	1
Technical report	0	4	4
Working paper series	0	4	4
Book chapter	0	1	1
Book	0	1	1
TOTAL	166	43	209

**Table 2.3. Collected definitions of crowdsourcing. Source: author**

Document	Page	Definition: Crowdsourcing is...
Alonso and Lease [25]	1	... the outsourcing of tasks to a large group of people instead of assigning such tasks to an in-house employee or contractor.
Bederson and Quinn [26]	1	... people being paid to do web-based tasks posted by requestors.
Brabham [9]	75	... an online, distributed problem solving and production model already in use by for profit organizations such as Threadless, iStock...
Brabham [4]	79	... a strategic model to attract an interested, motivated crowd of individuals capable of providing solutions superior in quality and quantity to those that even traditional forms of business can.
Buecheler et al. [11]	1	... a special case of such collective intelligence.
Burger-Helmchen and Penin [10]	2	... one way for a firm to access external knowledge.
Chanal and Caron-Fasan [27]	5	... the opening of the innovation process of a firm to integrate numerous and disseminated outside competencies through web facilities. These competences can be those of individuals (for example creative people, scientists, engineers...) or existing organized communities (for example OSS communities).
DiPalantino and Vojnovic [28]	1	... [a set of] methods of soliciting solutions to tasks via open calls to large-scale communities.
Doan et al. [8]	2	... a general-purpose problem-solving method.
Grier [29]	1	... a way of using the Internet to employ large numbers of dispersed workers. ... an industry that's attempting to use human beings and machines in large production systems.
Heer and Bostok [30]	1	... a relatively new phenomenon in which web workers complete one or more small tasks, often for micro-payments on the order of \$0.01 to \$0.10 per task.
Heymann and Garcia-Molina [31]	1	...getting one or more remote Internet users to perform work via a marketplace.
Howe [32]	-	...a web based business pattern, which make best use of the individuals on the internet, through open call, and finally get innovative solutions.
Howe [15]	-	... the application of Open Source principles to fields outside of software.

Relación entre el crowdsourcing y la inteligencia colectiva: el caso de los sistemas de etiquetado social

Document	Page	Definition: Crowdsourcing is...
	-	... the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and general large) network of people in the form of an open call. This can take the form of peer-production (when the job is performed collaborative), but is also often undertaken by sole individual. The crucial prerequisite is the: use of an open call format, and the wide network of potential laborers. ... a business practice that means literally to outsource an activity to the crowd.
Howe [1]	-	... the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call.
	-	... just a rubric for a wide range of activities.
	-	... the mechanism by which talent and knowledge is matched to those of need it.
Kazai [33]	-	... an open call for contributions from members of the crowd to solve a problem or carry out human intelligence tasks, often in exchange for micro-payments, social recognition, or entertainment value.
Kleeman et al. [13]	22	... a form of the integration of users or consumers in internal processes of value creation. The essence of crowdsourcing is the intentional mobilization for commercial exploitation of creative ideas and other forms of work performed by consumer.
	5	... outsourcing of tasks to the general internet public.
	6	... a profit oriented form outsources specific tasks essential for the making or sale of its product to the general public (the crowd) in the form of an open call over the internet, with the intention of animating individuals to make a contribution to the firm's production process for free or significantly less than that contribution is worth to the firm.
La Vecchia and Cisternino [34]	425	... a tool for addressing problems in organizations and business.
Ling [35]	1	... a new innovation business model through internet.
Liu & Porter [36]		... the outsourcing of a task or a job, such as a new approach to packaging that extends the life of a product, to a large group of potential innovators and inviting a solution. It is essentially open in nature and invites collaboration within a community.
Mazzola and Distefano [37]	3	... an intentional mobilization, through Web 2.0, of creative and innovative ideas or stimuli, to solve a problem, where voluntary users are included by a firm within the internal problem solving process, not necessarily aimed to increase profit or to create product or market innovations, but in general, to solve a specific problem.
Oliveira et al. [38]	413	... a way of outsourcing to the crowd tasks of intellectual assets creation, often collaboratively, with the aim of having easier access to a wide variety of skills and experience.
Poetz and Schreier [39]	4	... outsource the phase of idea generation to a potentially large and unknown population in the form of an open call.
Porta et al [40]		... enlisting customers to directly help an enterprise in every aspect of the lifecycle of a product or service.
Reichwald and Piller [41]	58	... interactive value creation: in terms of isolated activity of individual as directed toward one unit of the product, involving a cooperation between firm and users in the development of a new

Document	Page	Definition: Crowdsourcing is...
		product.
Ribiere & Tuggle [42]		... consists of making an open online call for a creative idea, or problem solving, or evaluation or any other type of business issues, and to let anyone (in the crowd) submit solutions.
Sloane [43]		...one particular manifestation of open innovation. It is the act of outsourcing a task to a large group of people outside your organization, often by making a public call for response. It is based on the open source philosophy, which used a large “crowd” of developers to build the Linux operating system.
Vukovic [6]	1	... new on-line distributed problem solving and production model in which networked people collaborate to complete a task.
Vukovic et al. [44]	539	... a new online distributed production model in which people collaborate and may be awarded to complete task.
Wexler [45]	11	... focal entity’s use of an enthusiastic crowd or loosely bound public to provide solutions to problems.
Whitla [46].	15	... a process of outsourcing of activities by a firm to an online community or crowd in the form of an “open call”.
	16	... a process of organising labour, where firms parcel out work to some form of (normally online) community, offering payment for anyone within the ‘crowd’ who completes the tasks the firm has set.
Yang et al. [47]		... the use of an Internet-scale community to outsource a task.

These 40 definitions come from 32 distinct articles published between 2006 and 2011 (2006, 2; 2008, 7; 2009, 4; 2010, 10; 2011, 9). The authors with multiple definitions of the term are Howe, Brabham, Kleeman et al., Grier, Vukovic, and Whitla.

### 3.2. Preparation

From the textual analysis of these definitions and the revision of the literature [1,10,48], three elements are identified (Crowd, 1; Initiator, 2; Process, 3). From which, eight characteristics are extracted constituting the *differentia specifica* [20].

About the crowd:

1. Who forms it. (a)
2. What it has to do. (b)
3. What it gets in return. (c)

About the initiator:

1. Who it is. (d)
2. What they get in return for the work of the crowd. (e)

About the process:

1. The type of process it is. (f)
2. The type of call used. (g)
3. The medium used. (h)

The results obtained for each characteristic are described below, as well as the partial synthesis that will form part of the proposed definition.

### **3.2.1. Who forms the crowd (a)**

The majority of the authors agree in defining the crowd in a general manner, providing information such as composition, type of people, heterogeneity, or the skills possessed.

Reference is made to the crowd as a generic mass of individuals: general Internet public [13], large group of people [1,15,25,39,36,43], individuals [13,27], people [26,44], or members of the crowd [33]. Some authors specify further the origin or grouping of the crowd: users (referring to a firm), consumers [13], customers [40], voluntary users [37], Internet-scale community [47], or organized and online communities [27,46].

Based on the sources consulted, it is possible to distinguish two crowd characteristics: number of people and their typology.

Regarding the number, the majority of the authors make reference to an indeterminate and large group of individuals, a group of people that do not necessarily know each other, and a loosely bound public according to Wexler [45]. The only exception is the online communities, where there is a greater possibility of the people knowing each other.

Regarding the type of people, this is obtained by describing the crowd. Kleeman et al. [13] identify the crowd as users or consumers, considered the essence of crowdsourcing. Schenk and Guittard [3] identify the nucleus of the crowd as amateurs (students, young graduates, scientists or simply individuals), although they do not set aside professionals. Authors such as Grier [29] and Heer and Bostok [30] identify the crowd as web workers. According to Howe [1], Crowdsourcing certainly requires a smart, well-trained crowd.

#### *Who forms the crowd - conclusion*

Fifty percent of the definitions coincide when the crowd is profiled as a large group of individuals. The optimum number of people will depend on the crowdsourcing initiative, due to the fact that the information needs to be filtered and evaluated [34]. There are initiatives,

such as in the case of the Iceland Constitution [49], where the optimal size is approximately 330.000 people, while in others it is a few thousands, like in the Lego case [1]. There are also cases where the size of the crowd is limited, e.g., those within a company, those that deal with confidential information, or those that are directed towards customers of a certain company.

In relation to the knowledge possessed by the individuals within the crowd, each initiative will need a specific one, thus limiting the number of participants. In the case of Amazon Mechanical Turk, a website where any given person can make micropayments in return for generally repetitive work, the proposed tasks do not generally require people with special skills. The same thing occurs in cases where the users have to give an opinion on a given product [50]. However, the tasks proposed on Innocentive or Starmind, websites that allow organizations to propose R&D problems whose resolution implies an economic recompense need a more educated crowd. This is demonstrated by Buecheler et al. [11] and others, who identify 66% of the participants of Starmind as PhD students, postdoctoral, researchers, professors, etc. Similar results were obtained by Brabham with the crowd of iStockPhoto [9] or Threadless [51], whose platforms relate to creative tasks.

The heterogeneity of the crowd will depend upon the type of initiative considered. Some will require the wisdom of crowds like a heterogeneous crowd [52] where each person brings their personal knowledge. In other cases, the heterogeneity will not be so important, such as in the translation tasks proposed by Amazon Mechanical Turk.

Therefore, we can conclude that the crowd will refer to a group of individuals whose characteristics of number, heterogeneity, and knowledge will be determined by the requirements of the crowdsourcing initiative.

### **3.2.2. What the crowd has to do (b)**

In regards to what the crowd has to do, two tendencies are detected; one more general and one more concrete. The general tendency includes two groups of authors. The first considers that the crowd should just undertake tasks [6,25,28,30,38,46,47,36], specifying at times the difficulty or size of these tasks [30], a given characteristic such as being done via the web [26], or of being human intelligent tasks [33]. The second refers to the fact that the crowd has to solve problems [8,9,4,33,34,37], in many cases for companies. The authors also make reference in a general way to what the crowd should undertake: a function or activity [15,32], a job [1], or simply to contribute to the firm [13].

About the specific tendency, authors such as Reichwald and Piller [41] make reference to the development of a new product, Kleeman et al. [13] speak of the exploitation of creative ideas, and Poetz and Schreier [39] contemplate idea generation. Beside the collected definitions, authors such as Giudice [53] are more concrete in the way they propose rating, recommendation, or text comments.

#### *What the crowd has to do - conclusion*

In principle, any non-trivial problem can benefit from crowdsourcing [8]. This includes tasks that range from purely routine poor cognitive tasks, to complicated tasks [13], passing through creative tasks or those related to innovation [41] where uniqueness has value per se [3]. Independent from the complexity of the problem, Vukovic [44] and Herr and Bostok [30] emphasize that a generic crowdsourcing task must be divisible into lower level tasks, each one of which can be accomplished by individual members of the crowd.

It is important to indicate that the tasks undertaken need to have a clear objective. For example, in an online platform called InnoCentive, money is offered in exchange for the solution of problems and in an Internet t-shirt company called Threadless, t-shirt designs are created and selected by users. Therefore, the use of free services, unless there is a secondary purpose, does not imply a crowdsourcing action. In this way, a user uploading a video to YouTube and sharing it is not a crowdsourcing initiative, while it is when a user uploads a video to any given platform to participate in initiatives such as those of Doritos and Pepsi at the Superbowl [54].

In this way, it can be concluded that the crowd will need to carry out the resolution of a problem through the undertaking of a task of variable complexity and modularity that will imply the voluntary contribution of their work, money (in the case of crowdfunding), knowledge, and/or experience. It is considered that a problem is comprised of any given situation of need held by the initiator of the crowdsourcing activity, e.g., the translation of a fragment of text or opinions about products.

### **3.2.3. What does the crowd get in return (c)**

Given that this characteristic is one of the most important in crowdsourcing, it is surprising that few definitions mention it. While Vukovic [44] mentions the existence of recompense, and Kazai [33] talks about social recognition and entertainment value as recompense, the rest of the authors that talks about the recompense identify it with money [13,26,30,33,46].

In reference to the level of recompense, Herr and Bostok [30] and Kleeman et al. [13] specify the recompense to micro-payments of the order of \$0.01 to \$0.10 per task, as occurs in the case of Amazon Mechanical Turk. In other cases such as InnoCentive, the prizes can even reach the level of a million dollars. Kleeman et al. [13] indicate that the task should be done for free or for significantly less than the contribution is worth to the firm.

*What does the crowd get in return - conclusion*

One of the characteristics that differentiates the people included in the crowd is that they have to be compensated because they are acting voluntarily [34]. Some authors suggest that the best situation would be that in which the reward is not material and that instead the motivation to participate is similar to that in Open Source Communities: passionate about the activity and participating for fun [55].

In regards to real motivations of the crowd to participate, various studies have been carried out [9][51][56]. These studies suggest different motivations that fit some of Maslow's individual needs: the financial reward, the opportunity to develop creative skills, to have fun, to share knowledge, the opportunity to take up freelance work, the love of the community and an addiction to the tasks proposed; understanding addiction as an exaggeration to describe the amount of time the crowd spends on the crowdsourcing site and their love to that site.

In this way, the recompense would vary depending on the crowdsourcer, but would always look to satisfy one or more of the individual needs mentioned in Maslow's pyramid [57]: economic reward, social recognition, self-esteem, or to develop individual skills. Although certain authors such as Kazai [33] also speak of entertainment as a type of motivation, it's important to mention that entertainment is present in any of the hierarchial levels proposed by Maslow [58].

On the other hand, it is important to highlight that the use of a free service cannot be considered recompense, as seen in Delicious or YouTube. This is because in those cases the user does not have to undertake a concrete task (except for the registration) to be able to use the services.

It's also important to highlight the reward is always given by the initiator of the crowdsourcing initiative (crowdsourcer). There can be secondary rewards, like social recognition from other crowdsourcing participants, but these rewards are not the main ones, and are not required to be present.



Therefore, it can be concluded that the user will obtain satisfaction of a given necessity, whether it be economic, social recognition, self-esteem, or the development of individual skills.

#### **3.2.4. Who is the initiator (crowdsourcer) (d)**

With respect to the person that initiates crowdsourcing processes (referred to as the crowdsourcer going forward), the majority of authors identify this individual, implicitly or explicitly, as a company [10,13,25,27,32,34-37,41,46,40,43]. Only the definitions of Howe [32] and La Vecchia and Cisternino [34] also include institutions or organizations without specifying if they are companies or not. In this sense, Brabham [9] is much more specific and makes reference to for-profit organizations. Lastly, Bederson and Quinn [26] refer to requestors, without specifying any characteristics.

##### *Who is the initiator (crowdsourcer) - conclusion*

Although it is certain that the crowdsourcer is in many cases a company (Converse, Sony, L'Oreal, etc.), it can also be a public organization, such as the FBI [59] or the European Union [60], writers, such as Jeff Howe who used crowdsourcing to design the cover of one of his books [1], or individuals, such as those cases of crowdfunding where any given type of professional can seek funding. This is to say that crowdsourcing does not only suggest a business model for companies, but is also a potential problem solving tool for the government and the non-profit sector [4].

Therefore, it can be concluded that the crowdsourcer can be any given entity that has the means to carry out the initiative considered, whether it is a company, institution, non-profit organization, or an individual.

#### **3.2.5. What the initiator gets in return (e)**

The majority of the authors agree that crowdsourcers will get the result they seek for a given task [1,15,6,28,30,31,33], with some being more direct and indicating that this result implies the resolution of a problem [8,9,34,37,45]. The rest of the authors can be considered as being a part of one of three groups: those that identify what the crowdsourcer gets with knowledge, those that identify it with ideas, and those that identify it with a given type of added value.

In the first case, Howe [1] indicates that crowdsourcers obtain talent and knowledge, and Burger-Helmchen and Penin [10] indicate that they obtain external knowledge. Other authors also include knowledge, but in an implicit form. For example, Oliveira et al. [38] indicate that

crowdsourcers obtain access to skills and experience, and Chanal and Caron-Faran [27] make reference to disseminated outside competencies. The authors of the second group identify the achieved object with ideas, with Kleeman et al. [13] going further and discussing commercial exploitation of creative ideas and making a sale of its products [13][46]. Kleeman et al. [13] could be also included in the third group, whose authors identify the achieved object with a given type of added value: value creation [47], increased profits, and product and service innovations [44].

#### *What the initiator gets in return - conclusions*

Many authors refer to specific cases, such as Del Giudice [53] who indicates that social feedback is obtained. For this reason, those cases should not be taken into account in the preparation of the definition.

It can be concluded that the crowdsourcer will obtain the solution to the problem via the fulfilment of a given action or task by the crowd. The crowdsourcer will benefit from the work of the crowd, from its experience, from its knowledge, and also, in the case of crowdfunding, from its assets.

#### **3.2.6. What type of process it is (f)**

In regards to the type of process addressed by crowdsourcing, there are authors who identify it as an outsourcing process, such as in the case of Amazon Mechanical Turk [13,38,39,46,36,43] and others as a problem solving process [9,37,40] via a distributed online process [37], such as in the case of InnoCentive. Others indicate that it is a production model [9,44] with an example being Threadless, while there are others who identify it as a business model or practice [15,35] or a strategic model, relating crowdsourcing directly to the business area [4]. There are also authors that identify crowdsourcing as a process of organizing labour [46], as a client integration process [13], or as an open innovation process [27, 43]; understanding open innovation as a paradigm that assumes firms can commercialize both its own ideas as well as innovations from other firms [61].

#### *What type of process it is - conclusion*

From all the previous affirmations various common points can be taken: crowdsourcing is an online process that is distributed by the very nature of the Internet and it always involves the participation of the crowd. The rest of the characteristics depend on the proposed initiative.

In this sense, each one of the definitions makes reference to a distinct type of crowdsourcing initiative: it will be a production of goods model in the case of Threadless, but not in the case of InnoCentive. In a similar way, crowdsourcing will be an open innovation process in InnoCentive but not in the case of Amazon Mechanical Turk, where it is an outsourcing process. The majority of the examples of crowdsourcing suppose a business model, but not always (e.g. FBI, or the European Union).

It can be concluded that crowdsourcing will be a participative distributed online process that allows the undertaking of a task for the resolution of a problem.

### **3.2.7. What type of call to use: Open call (g)**

With respect to the type of call used to propose tasks to the crowd, only ten documents make reference to the use of an open call [1,13,28,32,33,39,46,36,40,43].

*Conclusion - What type of call to use: Open call*

In agreement with the bibliography consulted, there are authors who consider that the call to bring together the potential participants should not be limited to experts or preselected candidates, or that participation should be non-discriminatory [3]. Everybody can answer the call: individuals can participate in addition to firms, non-profit organizations, or communities of individuals [10]. With this in mind, the call should be molded to the concrete crowdsourcing initiative. Whitley [46] clearly explains this by indicating that the call can be of one of three types:

1. A true open call where any given interested party can participate.
2. A call limited to a community with specific knowledge and expertise.
3. A combination of both, where an open call is made, but those who can participate are controlled.

In conclusion, it can be said that to get in touch with the crowd a flexible open call will be used.

### **3.2.8. Which medium is used (h)**

All the authors that mention the utilized medium make reference to the Internet, explicitly [1,9,4,13,6,26,27,29,31,35,44,46,47,42], or implicitly, like Howe [32] when he speaks of a web-based business pattern or Herr and Bostok [30] when they speak of web workers.

*Which medium is used - conclusion*

With respect to this characteristic there is unanimity: the medium used by crowdsourcing is the Internet. In fact, the importance of the Internet in crowdsourcing has been emphasized by a multitude of authors [1,10,13,14], some of them even affirm that Web 2.0 is the technological basis upon which crowdsourcing is developed and operates [2,44] given the level of collaboration that can be achieved [1,2].

### **3.3. Integrating crowdsourcing definition**

From the analysis undertaken, and fusing the previous partial elements, a definition that covers any type of crowdsourcing initiative has been created. It achieves the previously mentioned objectives of the study, discerns whether a given activity is crowdsourcing or not, and formalizes a theoretical base through the reduction of semantic confusion. The definition is as follows:

“Crowdsourcing is a type of participative online activity in which an individual, an institution, a non-profit organization, or company proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task. The undertaking of the task, of variable complexity and modularity, and in which the crowd should participate bringing their work, money, knowledge and/or experience, always entails mutual benefit. The user will receive the satisfaction of a given type of need, be it economic, social recognition, self-esteem, or the development of individual skills, while the crowdsourcer will obtain and utilize to their advantage that what the user has brought to the venture, whose form will depend on the type of activity undertaken.”

### **3.4. Verification**

As can be seen below, the definition will be applied to eleven initiatives present on the Internet, some of them crowdsourcing, others not, assessing the eight characteristics of the definition [6,22]. To this end, ‘+’ will be assigned to a characteristic that clearly appears; and ‘-’ to those characteristics which do not appear.

In Table 3.4, the assessment of each characteristic in each case can be seen. The selected examples are: Wikipedia (collaborative online encyclopedia), InnoCentive (an online platform where money is offered in exchange for the solution of problems), Threadless (an Internet t-shirt company, whose designs are created and selected by users), Amazon Mechanical Turk (a platform where crowdsourcers can propose tasks that are offered in exchange for money), ModCloth (an Internet clothing shop that allows its users to give

opinions on and vote for clothing designs before their sale), YouTube (an Internet video platform), Lánzanos (a Spanish website where people give money for participating in different projects, receiving rewards for their participation), Delicious (a social bookmarking system), Fiat Mio (an initiative begun by Fiat through which a car has been created following the suggestions of users), iStockPhoto (an Internet image sale platform), and Flickr (a platform that allows the uploading and tagging of photographs).

The characteristics of the definition, to be evaluated in each case, have been mentioned previously:

- There is a clearly defined crowd. (a)
- There exists a task with a clear goal. (b)
- The recompense received by the crowd is clear. (c)
- The crowdsourcer is clearly identified. (d)
- The compensation to be received by the crowdsourcer is clearly defined. (e)
- It is an online assigned process of participative type. (f)
- It uses an open call of variable extent. (g)
- It uses the Internet. (h)

According to Table 2.4, some clear cases of crowdsourcing exist including InnoCentive, Threadless, Amazon Mechanical Turk, Lánzanos, iStockPhoto, ModCloth and Fiat Mio. For example, in the case of ModCloth, the crowd can be easily identified (ModCloth customers from any part of the world), a task (to rate dresses), a recompense (recognition given by the company to the opinions of the users and to participate in order to buy clothes that the user likes), a crowdsourcer (the company ModCloth), the compensation (cost saving and efficient use of resources, among others), the participative process (the process implies the conscious participation of the crowd), the open call (using their website) and the use of Internet.

On the other hand, other cases are not identified as crowdsourcing. In the case of Delicious, six characteristics are not identified: a task with a clear goal, the recompense received by the crowd, the crowdsourcer, the benefit it receives, the participative nature of the task and the existence of an open call. Concerning the company behind Delicious, AVOS Systems, it does not act like a crowdsourcer and it does not receive a benefit from the work of the crowd.

Regarding the open call, there is no one; it is a free service usable by anyone. Furthermore, it cannot be said to be a participative process in which all the users are seeking the same end goal. The use of the site is mainly individual; then the platform makes use of the collective intelligence to interconnect and exploit the information. Lastly, For these reasons Delicious cannot be considered a crowdsourcing example.

**Table 2.4. Verification of the definition. Source: author**

	a	b	c	d	e	f	g	h
Wikipedia	+	+	+	-	-	+	-	+
InnoCentive	+	+	+	+	+	+	+	+
Threadless	+	+	+	+	+	+	+	+
Amazon Mechanical Turk	+	+	+	+	+	+	+	+
ModCloth	+	+	+	+	+	+	+	+
YouTube	+	-	-	-	-	-	-	+
Lánzanos	+	+	+	+	+	+	+	+
Delicious	+	-	-	-	-	-	-	+
Fiat Mio	+	+	+	+	+	+	+	+
iStockPhoto	+	+	+	+	+	+	+	+
Flickr	+	-	-	+	-	-	-	+

## 4. Conclusion and future work

The term “crowdsourcing” is a term in its infancy, which, as new applications appear, is undergoing a constant evolution. Following the analysis of a group of scientific articles, it has been shown that distinct definitions of crowdsourcing exist, clearly illustrating the lack of consensus and a certain semantic confusion.

This article provides a wide definition that covers the majority (if not all) of existing crowdsourcing processes. Through the analysis of all the authors’ definitions, eight characteristics common to any given crowdsourcing initiative were found: the crowd, the task at hand, the recompense obtained, the crowdsourcer or initiator of the crowdsourcing activity, what is obtained by them following the crowdsourcing process, the type of process, the call to participate, and the medium. For each one of these elements an analysis based on the collected definitions was undertaken and a conclusion formulated, attempting to make each element as global as possible while trying to maintain the upmost precision as well. The coordination of these conclusions has allowed the creation of a global definition that spans any of the crowdsourcing initiatives compared.

Additionally, it should be noted that the proposed definition encompasses all of the definitions mentioned in Table 3.3 due to its global reach. It also should be noted that these

definitions mentioned in Table 3.3 are very focused on a certain type of crowdsourcing initiative so the proposed definition will represent those cases in a more blurred way. For this reason, each type of concrete crowdsourcing activity (crowdvoting, crowdfunding, etc.) will require a more precise definition of each one of the eight elements. For example, in the case of crowdfunding, the task of the crowd will be to give money, while in the case of crowdvoting, it will be to vote for and give opinions on certain products.

Although the definition obtained is clear and accomplishes its objective, there is a limitation that must be noted. Emerald and SAGE databases, which include business and human science papers, have been consulted but the percentage of documents related to computer science area is higher than those found in other areas. Due to this, some nuances of crowdsourcing may have been lost. It would be important to complete this work trying to describe this evolving concept using similar methodology taking into account the definitions of crowdsourcing from other sources more related to business or human sciences.

About the future lines of investigation, there are other areas in crowdsourcing where little consensus exists, such as in the classification of distinct types of activities within crowdsourcing. With this in mind, some work analyzing, recompiling, and summarizing, with the goal of unifying some of the positions may be of interest.

Another area where consensus does not exist is in the relationship between crowdsourcing and other associated concepts such as Open Innovation, defined previously; Outsourcing, defined as a mean of procuring from external suppliers services or products that are normally part of organization [62]; or Open Source Development, which is understood as a kind of production that involves allowing access to the essential elements of a product to anyone for the purpose of collaborative improvement to the existing product [63]. While some authors unequivocally identify crowdsourcing with Open Innovation [27], others state the exact opposite [3]. Also in this case, it would be interesting to undertake a study of all the terms that are linked regularly with crowdsourcing to establish the similarities and differences with the objective of better profiling the concept of crowdsourcing and defining a theoretical framework, as has been attempted in this article.

## 5. References

- [1] Howe J., *Crowdsourcing: How the Power of the Crowd is Driving the Future of Business*. (Business Books, Great Britain, 2008).

- [2] Vukovic M. and Bartolini C., Towards a Research Agenda for Enterprise crowdsourcing. In: M. Tiziana and S. Bernhard (eds) *Leveraging Applications of Formal Methods, Verification, and Validation* (Springer, Berlin/Heidelberg, 2010) 425-434 [Lecture Notes in Computer Science 6415].
- [3] E. Schenk and C. Guittard, *Crowdsourcing: What can be Outsourced to the Crowd, and Why?* Technical Report (2009) Available from: <http://halshs.archives-ouvertes.fr/halshs-00439256/> (accessed: 1 September 2011).
- [4] Brabham D. C., *Crowdsourcing as a Model for Problem Solving: An Introduction and Cases*, *Convergence: The International Journal of Research into New Media Technologies* 14(1) (2008) 75-90.
- [5] Brabham D. C., *Crowdsourcing: A model for leveraging online communities*. In: A. Delwiche & J. Henderson (Eds.), *The Routledge Handbook of Participatory Culture* (in press).
- [6] Vukovic M., *Crowdsourcing for enterprises*. In: *Proceedings of the 2009 Congress on Services – I*, IEEE Computer Society (Washington, DC, USA 2009). 686-692.
- [7] Geiger D., Seedorf S. and Schader M., *Managing the Crowd: Towards a Taxonomy of Crowdsourcing Processes*. In: *Proceedings of the Seventeenth Americas Conference on Information Systems*, Detroit, Michigan August 4th-7th 2011
- [8] Doan A., Ramakrishnan R. and Halevy A.Y., *Crowdsourcing systems on the World-Wide Web*, *Communications of the ACM* 54(4) (2011) 86-96.
- [9] Brabham D. C., *Moving the crowd at iStockphoto: The composition of the crowd and motivations for participation in a crowdsourcing application*, *First Monday* 13(6) (2008)
- [10] Burger-Helmchen T. and Penin J., *The limits of crowdsourcing inventive activities: What do transaction cost theory and the evolutionary theories of the firm teach us?*. In: *Workshop on Open Source Innovation*, Strasbourg, France (2010).
- [11] Buecheler T., Sieg J.H., Füchslin R.M., Pfeifer R., *Crowdsourcing, Open Innovation and Collective Intelligence in the Scientific Method: A Research Agenda and Operational Framework*. In: H. Fellerman et al (eds), *Artificial Life XII. Proceedings of the Twelfth International Conference on the Synthesis and Simulation of Living Systems*, Odense, Denmark, 19-23 August 2010, 679-686.
- [12] Huberman B.A., Romero D.M. and Wu F. *Crowdsourcing, Attention and Productivity*. *Journal of Information Science* 35(6) (2009) 758–765



- [13] Kleeman F., Voss G.G and Rieder K., Un(der)paid Innovators: The Commercial Utilization of Consumer Work through crowdsourcing, *Science, Technology and Innovation Studies* 4(1) (2008) 5-26.
- [14] Andriole S.J., Business impact of Web 2.0 technologies, *Communications of the ACM* 53(12) (2010) 67-79.
- [15] Howe J., The rise of crowdsourcing, *Wired* 14(6) (2006)
- [16] Denyer D., Tranfield D., Van Aken J.E., Developing design propositions through research synthesis, *Organization Studies* 29(3) (2008): 393.
- [17] Delgado M., *Revisión sistemática de estudios: Metaanálisis* (Signo, Barcelona, 2010).
- [18] Pettiti D. B., *Meta-analysis, Decision Analysis and Cost-Effectiveness Analysis* (Oxford University Press, New York, 2000).
- [19] Egger M., Smith G.D. and Altman D., *Systematic reviews in health care. Meta-analysis in context* (BMJ Books, London, 2001).
- [20] Tatarkiewicz W., *History of Six Ideas: An Essay in Aesthetics* (Springer, 1980)
- [21] Cosma G. and Joy M., Towards a Definition of Source-Code Plagiarism. *IEEE Transactions on Education* 51(2) (2008) 195-200
- [22] Aliakbarian S., Rahimabadi A.M., Sadeghi P.H. and Mirsatari N.S., Neighbor Definition in P2P Networks. In: *Proceedings of 2006 International Conference on Communications, Circuits and Systems*. (Guilin, 2007) 1562-1565
- [23] Brabham D. C., Crowdsourcing the public participation process for planning projects, *Planning Theory* 8(3) (2009) 242-262.
- [24] Wikipedia, Crowdsourcing (2011). Available from: <http://en.wikipedia.org/wiki/Crowdsourcing> (accessed 15 August 2011)
- [25] Alonso O. and Lease M., Crowdsourcing 101: Putting the WSDM of Crowds to Work for You. In: *Proceedings of the fourth ACM international conference on Web search and data mining, WSDM '11* (ACM, New York, 2011) 1-2
- [26] Bederson B. B. and Quinn A.J., Web workers Unite! Addressing Challenges of Online Laborers. In: *Proceedings of the 2011 annual conference extended abstracts on Human Factors in Computing Systems, CHI '11* (Vancouver, 2011)
- [27] Chanal V. and Caron-Fasan M.L., How to invent a new business model based on crowdsourcing: The crowdspirit ® case. In: *EURAM* (Lubjana, Slovenia, 2008)
- [28] DiPalantino D. and Vojnovic M., Crowdsourcing and all-pay auctions. In: *Proceedings of the 10th ACM conference on Electronic commerce, EC '09* (2009) 119–128.

- [29] Grier D. A., Not for All Markets. *Computer* 44(5) (2011) 6-8
- [30] Heer J., and Bostok M., Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In: *Proceedings of the 28th international conference on Human factors in computing systems, CHI'10* (ACM, New York, 2010) 203-212.
- [31] Heymann P. and Garcia-Molina H., Turkalytics: analytics for human computation. In: *Proceedings of the 20th international conference on World wide web, WWW '11* (ACM, New York, 2011) 477-486.
- [32] Howe J., Crowdsourcing: A definition. *Crowdsourcing: Why the Power of the Crowd is Driving the Future of Business*. Weblog, 2 June. Available at [http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing\\_a.html](http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html) (accessed 27-7-2011)
- [33] Kazai G., In Search of Quality in crowdsourcing for Search Engine Evaluation. In *Proceedings of the 33rd European conference on Advances in Information retrieval* (Springer-Verlag, Berlin/Heidelberg, 2011). [Lecture Notes in Computer Science 6611, 165-176]
- [34] La Vecchia G. and Cisternino A., Collaborative workforce, business process crowdsourcing as an alternative of BPO. In: *Proceedings of First Enterprise crowdsourcing Workshop in conjunction with ICWE 2010* (Springer-Verlag, Berlin/Heidelberg, 2010) 425-430
- [35] Ling P., An Empirical Study of Social Capital in Participation in Online crowdsourcing, *Computer* 7(9) (2010) 1-4.
- [36] Liu, E., & Porter, T. (2010). Culture and KM in China. *VINE*, 40(3/4), 326-333
- [37] Mazzola D. and Distefano A., Crowdsourcing and the participacion process for problem solving: the case of BP. In: *VII Conference of the Italian Chapter of AIS. Information technoogy and Innovation trend in Organization*. (Napoles, Italy, 2010)
- [38] Oliveira F., Ramos I., Santos L., Definition of a crowdsourcing Innovation Service for the European SMEs. In: Daniel F. et al. (eds.) *Current Trends in Web Engineering* (Springer, Berlin/Heidelberg, 2010) 412-416
- [39] Poetz M.K. and Schreier M., The Value of crowdsourcing: Can Users Really Compete with Professionals in Generating New Product Ideas?. *Journal of Product Innovation Management* (2009) Forthcoming. Available at SSRN: <http://ssrn.com/abstract=1566903>

- [40] Porta, M., House, B., Buckley, L., & Blitz, A. (2008). Value 2.0: eight new rules for creating and capturing value from innovative technologies. *Strategy & Leadership*, 36(4), 10-18
- [41] Reichwald R., Piller F., *Interaktive Wertschöpfung. Open Innovation, Individualisierung und neue Formen der Arbeitsteilung* (Gabler, Wiesbaden, 2006).
- [42] Ribiere, V. M., & Tuggle, F. D. (Doug). (2010). Fostering innovation with KM 2.0. *VINE*, 40(1)
- [43] Sloane, P. (2011). The brave new world of open innovation. *Strategic Direction*, 27(5), 3-4.
- [44] Vukovic M., Mariana L. and Laredo J., *PeopleCloud for the Globally Integrated Enterprise*. In: D. Asit et al. (eds) *Service-Oriented Computing*. (Springer-Verlag, Berlin/Heidelberg, 2009)
- [45] Wexler M. N., *Reconfiguring the sociology of the crowd: exploring crowdsourcing*, *International Journal of Sociology and Social Policy* 31(1) (2011) 6 - 20
- [46] Whitla P., *Crowdsourcing and Its Application in Marketing*, *Contemporary Management Research* 5(1) (2009) 15-28
- [47] Yang J., Adamic L.A. and Ackerman M.S., *Crowdsourcing and knowledge sharing: strategic user behaviour on taskcn*. In: *Proceedings of the 9th ACM conference on Electronic commerce* (ACM, New York, 2008) 246-255.
- [48] Geerts S., *Discovering crowdsourcing: theory, classification and directions for use*. (Technische Universiteit Eindhoven, 2009)
- [49] Siddique H. *Mob rule: Iceland crowdsources its next constitution*. *The Guardian*. Available from <http://www.guardian.co.uk/world/2011/jun/09/iceland-crowdsourcing-constitution-facebook> (accessed: 1 December 2011)
- [50] Inc, *Using crowdsourcing to control Inventory* (2010). Available from <http://www.inc.com/magazine/20100201/using-crowdsourcing-to-control-inventory.html>. (accessed 18 of August 2011)
- [51] Brabham D.C., *Moving the crowd at Threadless*, *Information, Communication & Society*, 13(8)(2010), 1122-1145
- [52] Surowiecki J., *The wisdom of crowds* (Anchor Books, New York, 2005).
- [53] Giudice K. D., *Crowdsourcing credibility: The impact of audience feedback on Web page credibility*. In: *Proceedings of the 73rd ASIS&T Annual Meeting on Navigating Streams in an Information Ecosystem*, *ASIS&T '10* (2010). 47(1) 1-9

- [54] Superbowl, Crash the SuperBowl. Available from <http://www.crashthesuperbowl.com/> (accessed 18 August 2011).
- [55] Stewart O., Huerta J.M. and Sader M., Designing crowdsourcing community for the enterprise. In: Proceedings of the ACM SIGKDD Workshop on Human Computation, HCOMP '09 (ACM, New York, 2009) 50-53.
- [56] Lakhani K.R., Jeppesen, L.B., Lohse, P.A. and Panetta J.A., The value of openness in scientific problem solving, Harvard Business School Working Paper No. 07-050.
- [57] Maslow A.H., A Theory of Human Motivation. Psychological review 50 (1943)
- [58] Veal A. J., Leisure and tourism policy and planning (CABI Publishing, 2002)
- [59] FBI - Federal Bureau of Investigation, Cryptanalysts: Help Break the Code (2011). Available from [http://www.fbi.gov/news/stories/2011/march/cryptanalysis\\_032111](http://www.fbi.gov/news/stories/2011/march/cryptanalysis_032111) (accessed: 15 July 2011)
- [60] ECMT - European Commission for Mobility and Transport, Door-to-Door in a click (2011). Available from [http://ec.europa.eu/transport/its/multimodal-planners/index\\_en.htm](http://ec.europa.eu/transport/its/multimodal-planners/index_en.htm) (accessed 15 July 2011)
- [61] Dahlander L. and Gann D.M., How open is innovation? Res. Policy (2010). Article in Press
- [62] Heizer J. and Render B., Operations Management, 9th edition (Pearson/Prentice Hall, 2008)
- [63] OSI, The Open Source Definition. Available from <http://opensource.org/docs/osd> (accessed 25 of November 2011)

**CAPÍTULO 3 -**

**Tipología del crowdsourcing  
basada en la actividad de la  
multitud**

## 3.1 Introducción

Este capítulo se corresponde con el artículo “*Clasificación de iniciativas de crowdsourcing basada en tareas*” publicado en la revista *El profesional de la Información*.

### 3.1.1 Resumen del artículo

El mismo problema identificado con la definición de crowdsourcing (diferentes definiciones con diferentes características) se identifica también en el caso de las tipologías que tratan de clasificar las distintas actividades que pueden ser consideradas crowdsourcing. Mediante una revisión de la literatura se encuentran distintas tipologías, escogiendo aquellas que se basan en la acción que debe realizar la multitud y dejando de lado el resto. Mediante una tabla comparativa de doble entrada se comparan todas estas tipologías, formando con las coincidencias una tipología integradora que es probada con éxito en un conjunto de plataformas de crowdsourcing seleccionadas al azar.

### 3.1.2 Datos de la publicación

El artículo ha sido publicado en la revista *El profesional de la Información*, revista internacional de Información, Documentación, Biblioteconomía y Comunicación. La revista está indexada tanto en *Social Science Citation Index* como en *Scopus*. Se encuentra en distintas bases de datos como *Academic Search Premier*, *Francis*, *Business Source Elite*, *Dialnet*, *Latinindex* o *In-Recs*, donde tiene un índice de 0.945 y ocupa la posición 1/22 de las revistas de documentación.

Esta revista tuvo en 2011 un índice de impacto *JCR Social Science Edition* de 0,326, encontrándose la revista en la posición 62/83 en "Ciencias de la Información y Biblioteconomía". Ocupa el segundo cuartil (Q2).

Los autores del artículos son, en orden de aparición, Enrique Estellés-Arolas y Fernando González Ladrón-de-Guevara.

- Nombre de la revista: El Profesional de la Información
- ISSN: 1386-6710
- Fecha: Mayo-Junio 2012
- Volumen: 21
- Nº: 3

## 3.2 Artículo

### Clasificación de iniciativas de crowdsourcing basada en tareas

Enrique Estellés-Arolas

Department of Management, Technical University of Valencia, Valencia, Spain

Fernando González-Ladrón-de-Guevara

Department of Management, Technical University of Valencia, Valencia, Spain

#### Abstract

Las iniciativas de crowdsourcing planteadas por organizaciones de ámbitos diversos como la música, el diseño o la catalogación son cada vez más frecuentes. A pesar de este auge, la ausencia de un fundamento teórico consistente genera problemas como la existencia de tipos o clasificaciones de crowdsourcing que se superponen y entremezclan o la ausencia de una definición compartida. A partir de una revisión sistemática de la bibliografía se analizan las tipologías considerando la naturaleza de las tareas que debe realizar la ‘multitud’ como criterio, y se propone una nueva tipología integradora.

#### Keywords

Crowdsourcing, Tipología, Clasificación, Multitud, Tarea

## 1. Introduction

El crowdsourcing hace referencia a un conjunto de iniciativas de tipo participativo que se nutren de otros fenómenos como la innovación abierta (Chesbrough, 2003) o la inteligencia colectiva (Schenk; Guittard, 2011).

El periodista americano Jeff Howe lo definió en 2006 como una convocatoria abierta iniciada por una empresa o institución –normalmente realizada por un empleado– dirigida a un grupo de individuos indefinido (“la multitud” o crowd), con frecuencia grande (Howe, 2006) con el fin de externalizar una función. Diversos autores han tratado hasta el momento de elaborar una definición, unos centrándose en el uso del crowdsourcing como un proceso de resolución de problemas (Brabham, 2008b, Vukovic, 2009), otros como una forma de externalizar tareas (Oliveira; Ramos; Santos, 2009) o como una manifestación particular de la “innovación abierta” (Sloane, 2011). Estellés y González (2012) presentan una definición que permite identificar cualquier tipo de iniciativa de crowdsourcing en base a 8 elementos: tarea concreta

a realizar, multitud que participará con sus aportaciones (crowdworkers), beneficio para dicha multitud, iniciador, beneficio para el iniciador, utilización de un proceso participativo, uso de una convocatoria abierta y flexible, y uso de internet como infraestructura fundamental.

Crowdsourcing es una iniciativa participativa de innovación abierta o de inteligencia colectiva

El crowdsourcing se lleva a cabo en internet, con el apoyo de las aplicaciones Web 2.0 que facilitan la conexión de miles de usuarios que comparten información y resuelven problemas de forma colaborativa (Burger-Helmchen; Pénin, 2010; Vukovic; Bartolini, 2010a).

Las tareas que realizan los colaboradores pueden abarcar desde la catalogación de documentos hasta la innovación que mejora un proceso o un bien. Atendiendo a su complejidad, pueden ser de tres tipos:

- simples, normalmente repetitivas, que no requieren de un nivel cognitivo alto, como por ejemplo el etiquetado de una imagen;
- complejas que necesiten de una capacidad intelectual y de inventiva mayor, como la resolución de un problema de una empresa; y
- creativas, donde la singularidad de la aportación del usuario es fundamental, como en el diseño de un logo (Schenk; Guittard, 2009).

En muchos casos se trata de tareas modulables, hecho que posibilita su realización por varios usuarios en paralelo, produciendo un ahorro económico y de tiempo (Mazzola; Distefano, 2010; Kleeman; Voss; Rieder, 2008). Por todo ello, empresas como Doritos (SuperBowl, 2011), organizaciones públicas como la Unión Europea (ECMT, 2011) o incluso individuos aislados, como el músico español Carlos Jean (PlanB, 2011), se interesan por el potencial del crowdsourcing (Howe, 2008; Vukovic; Bartolini, 2010a).

Sin embargo el crowdsourcing no dispone de una base teórica que fundamente su estudio (Denyer; Tranfield; VanAken, 2008), aunque este problema está subsanándose. Ya existen puntos de acuerdo entre los autores, como que todas las iniciativas de este tipo deben tener, como mínimo, dos elementos: una multitud a priori indefinida y heterogénea (Geerts, 2009; Schenk; Guittard, 2009) y la utilización de una llamada abierta a todo el mundo (Pénin, 2008; Geerts, 2009; Burger-Helmchen; Pénin, 2010), coincidiendo con los elementos enumerados



por Estellés y González (2012). Se han generado también diversas clasificaciones basadas en criterios como la perspectiva organizativa (Geiger; Seedorf; Schader, 2011).

El presente artículo pretende participar en la creación de esta base teórica: su objetivo es aportar una tipología del crowdsourcing basada en la tarea a realizar por parte de la multitud. Para ello se analizan tipologías planteadas por diversos autores, identificando divergencias y puntos en común y definiendo una tipología integradora.

## **2. Metodología**

El trabajo se ha llevado a cabo en tres fases: 1) revisión sistemática de la bibliografía existente (Delgado-Rodríguez; Doménech; Llorca, 2010), 2) creación de un repositorio documental con los documentos hallados, y 3) descripción de sus categorías, ilustrándolas con ejemplos, comparándolas y detectando relaciones entre ellas. Para este fin se ha elaborado e interpretado una parrilla de análisis (Codina, 1997; Pinto-Molina et al., 2007).

### **2.1. Revisión sistemática: búsqueda de información**

Se han realizado consultas en siete bases de datos: ACM, Scopus, Emerald, SAGE, Wiley, SpringerLink y ScienceDirect. El criterio de selección ha sido la ocurrencia del término crowdsourcing tanto en el título como en las palabras clave del documento. Se han obtenido 151 documentos. Una búsqueda adicional en Google Scholar por “classification of crowdsourcing” OR “crowdsourcing classification” ha permitido obtener nueve documentos, a partir de las bibliografías de los cuales se han hallado 28 documentos más. La composición del repositorio documental obtenido se describe en la tabla 3.1. La mayoría de los documentos (66%) aparecen en actas de congresos, lo que sugiere el carácter preliminar de la investigación existente sobre este objeto de conocimiento.

### **2.2. Filtrado de documentos**

Se han descartado documentos que no aportan clasificación alguna de las iniciativas de crowdsourcing, en total once documentos. De estos, se han rechazado los que no utilizan el criterio del tipo de tarea sino otros como la naturaleza del crowd (Schenk; Guittard, 2009) o la recompensa (Corney et al., 2009). También se han descartado tipologías centradas en un área o subsector específico, como la de Ooman y Aroyo (2011) focalizada en galerías de arte, librerías, archivos, etc.; la de La-Vecchia y Cisternino (2010), centrada en modelos de negocio; y la de Geiger, Seedorf y Schader, (2011) con una perspectiva organizativa.

Finalizado el proceso de filtrado se han obtenido seis documentos que cumplieran los requerimientos citados.

**Table 3.1. Composición del repositorio documental**

Tipo de documento	Búsqueda en bases de datos	Búsqueda en Google Scholar	Total
Ponencia en congresos	108	11	119
Artículos en revista	43	11	54
Monografías	0	1	1
Otros	0	5	5
TOTAL	151	28	179

### 3. Descripción de las tipologías

A continuación se presentan ordenadas cronológicamente las tipologías encontradas, junto con ejemplos que las ilustran. Cada uno de sus subtipos estará identificado por un código que se utilizará después en la parrilla de análisis de subtipos.

#### 3.1. Reichwald & Piller (2006)

Agrupar las tareas de crowdsourcing con dos enfoques:

1. Innovación abierta (RP1). Tareas de cooperación entre la empresa iniciadora y sus clientes en la elaboración de nuevos productos y que suponen la generación de conocimiento.
2. Actividades operativas de soporte (RP2). Mejora de procesos operativos para la personalización masiva de bienes (Heizer; Render, 2010).

Comprenden desde tareas sencillas que requieren un nivel cognitivo bajo, como la búsqueda de información en internet, hasta tareas complejas que exigen competencias específicas como la búsqueda de una solución para un problema científico, pasando por tareas creativas, como el diseño de un logo.

#### 3.2. Howe (2008)

Propone los siguientes tipos de tareas (Howe, 2006):

1. De inteligencia colectiva (*crowdwisdom*) con 3 subtipos:
  - a. Predicción de mercados (H1.1): una comunidad de inversores particulares votan diversas alternativas a partir de la información descriptiva suministrada, como en el caso de Iowa Electronic Markets.

- b. Competición (crowdcasting) (H1.2): se recompensa a quien resuelva un desafío, como los planteados en la plataforma Innocentive. <http://www.innocentive.com>
  - c. Tormenta de ideas online (crowdstorming) (H1.3): similar a la que se produjo en el proyecto Idea Jam de IBM.
2. Creativas (*crowdproduction*) (H2). El promotor de la iniciativa (crowdsourcer) externaliza actividades que necesitan la energía creativa de los colaboradores para obtener un nuevo producto o servicio (una base de datos o cualquier tipo de contenido generado por los usuarios). Wikipedia o iStock son ejemplos de este tipo de tareas.
  3. Recoger opiniones de los usuarios (*crowdvoting*) (H3). Un ejemplo característico es Threadless, empresa de camisetas que pide a los usuarios que elijan sus diseños preferidos para lanzarlos al mercado.
  4. Obtención de fondos económicos (*crowdfunding*) (H4). Se pide una cantidad determinada de dinero a cambio de una recompensa: MyFootballClub es un club de fútbol en el que, a cambio de una cuota anual, los inversores deciden sobre el fichaje de jugadores o el precio de las entradas.

Algunos autores consideran esta clasificación como no demasiado rigurosa, dado el solapamiento de los tipos planteados (Geerts, 2009).

### **3.3. Kleeman, Voss & Rieder (2008)**

Comprende siete tipos:

1. Participación de consumidores en el desarrollo colaborativo de un producto, como en el proyecto *Fiat Mio* o en *Idea Storm* de Dell (K1).
2. El diseño de un nuevo producto, que depende casi por completo de las aportaciones de los usuarios. Así ocurre en Spreadshirt o Fluevog (K2).
3. Ofertas competitivas sobre ciertas tareas o problemas bien definidos, similares a Innocentive (K3).
4. Llamadas abiertas permanentemente donde los crowdworkers presentan información o documentación a lo largo de un período indeterminado de tiempo, como en el caso de iReport. Esta es una iniciativa en la que la CNN ha puesto un conjunto de herramientas online a disposición de cualquier reportero aficionado para recoger imágenes (K4).

5. Periodismo de comunidad (community reporting): los usuarios informan sobre nuevos productos o tendencias sobre algún tipo de comunidad online, como ocurre con Trendwatching.com (K5).
6. Valoración de productos por consumidores o perfiles de consumidor. Un ejemplo es Amazon Reviews (K6).
7. Soporte de clientes. Los usuarios de un servicio resuelven los problemas y las dudas de otros usuarios, como en la Universidad de Indiana donde el teléfono 24 horas de asistencia técnica ha sido sustituido por un foro que tanto empleados como usuarios utilizan para resolver las dudas planteadas (K7).

### **3.4. Brabham (2008a)**

Propone una clasificación de tareas para resolver problemas que comprende cuatro grupos con distintos objetivos:

1. Descubrimiento y gestión de conocimiento (B1). Su objetivo es encontrar y reunir de manera coherente cierto conocimiento disperso. Un ejemplo es el proyecto Peer to Patent Community Patent Review. En él, una comunidad online se encarga de informar sobre patentes existentes que puedan estar relacionadas con solicitudes a la oficina de patentes de EUA (Ghafele; Gibert; DiGiammarino, 2011).
2. Obtención de una respuesta correcta (B2). Se difunde un problema del que se busca una solución, esperando que la aporte un experto que podría encontrarse en la Red. Estas tareas se dan en plataformas como Innocentive, que permite difundir problemas de I+D con carácter científico a un conjunto de especialistas.
3. Diseño y valoración de productos por parte de los usuarios (B3). Este tipo de tareas son útiles cuando se desea conocer la opinión o las preferencias de los usuarios sobre un producto. Un ejemplo es la citada empresa Threadless.
4. Participación distribuida (B4). Las tareas son realizadas por una comunidad online y suelen implicar el procesamiento de grandes cantidades de datos. Un ejemplo son las propuestas en Amazon Mechanical Turk, plataforma donde cualquier empresa puede contratar una comunidad de usuarios para realizar trabajos, como la traducción de textos o la indización de imágenes.

### **3.5. Geerts (2009)**

Propone cuatro tipos que tienen como punto de partida la clasificación de Howe (2008):

1. Crowdcasting (G1). Un grupo de usuarios compiten para obtener una recompensa al proporcionar la mejor solución a un problema, como en Innocentive.
2. Crowdstorming (G2). Mediante foros, como los de Ideastorm de Dell, los participantes discuten, preguntan o proponen enfoques alternativos para resolver un problema de forma colectiva. Se suelen combinar distintas aportaciones para obtener el resultado final.
3. Crowdproduction (G3). Los colaboradores tienen como objetivo la elaboración conjunta de un bien: una base de datos para investigación, el contenido de una wiki, el etiquetado de recursos en red, etc.
4. Crowdfunding (G4). El objetivo de los colaboradores es invertir. Geerts (2009) distingue entre dar dinero a través de mediadores o por iniciativa individual. En el primer caso, aparecen plataformas como Kiva, que permite financiar a empresarios del tercer mundo. En las iniciativas particulares la multitud suele ser recompensada participando en la toma de decisiones relevantes, como en el equipo de fútbol MyFootballClub.

### **3.6. Burger-Helmchen & Pénin (2010)**

Se proponen tres tipos de tareas, difíciles de distinguir en algunos casos:

1. Tareas innovadoras (BH1). La multitud constituye tan sólo un revestimiento adicional que no resuelve los problemas. Para la empresa iniciadora es más importante recoger el conocimiento de un pequeño número de especialistas en diferentes campos que la participación de un número elevado de profanos (Pisano; Verganti, 2008).
2. Tareas rutinarias (BH2). Se trata de tareas modulares que no requieren competencias específicas, únicamente implican el uso de tiempo, como por ejemplo, la búsqueda de direcciones de correo electrónico de un determinado segmento de clientes para una actividad de marketing electrónico. Los colaboradores aportando su tiempo, información y capacidad de cómputo contribuyen a disminuir costes y aumentar la velocidad de ejecución de la tarea. Aquí el tamaño de la multitud sí importa: cuanto más grande, más tareas podrán ser ejecutadas de forma paralela y en menos tiempo.

3. Tareas de contenido (BH3). Los colaboradores aportan su tiempo, información y capacidad de cómputo para generar servicios basados en información (Wikipedia). Además del tamaño de la crowd, importa su heterogeneidad y diversidad.

## 4. Comparación de tipologías y desarrollo de una nueva

En la tabla 3.2 se presenta una parrilla de análisis que compara cada uno de los componentes de las tipologías con el resto. Se trata de una matriz no simétrica que leída por filas, destaca, para cada caso, los tipos (columnas) que no están representados. Es decir, la celda [2,3] (fila 2, columna 3) indica los elementos H3 y H4 de la tipología Howe que no están incluidos o representados en la tipología Reichwald & Piller.

**Table 3.2. Comparación de tipologías: elementos distintivos faltantes**

	<b>Reichwald &amp; Piller</b>	<b>Howe</b>	<b>Kleeman</b>	<b>Brabham</b>	<b>Geerts</b>	<b>Burger-Helmchen &amp; Pénin</b>
<b>Reichwald &amp; Piller</b>		H3 y H4	K4, K5, K6 y K7	B1 y B4	G4	BH2 y BH3
<b>Howe</b>	RP2		K7	B4		BH1 y BH2
<b>Kleeman</b>	RP2	H4		B4	G4	
<b>Brabham</b>		H4	K7		G2 y G4	
<b>Geerts</b>		H3	K7	B1		
<b>Burger-Helmchen &amp; Pénin</b>	RP2	H4			G2 y G4	

Lo primero que se observa es que la tipología de Reichwald & Piller (2006), al ser tan genérica –y la más antigua– no abarca muchos de los elementos del resto y, además, su tipo RP2 no aparece en tres de las tipologías consideradas. Por todo esto, no va a ser tenida en cuenta en el análisis comparativo (fondo azul).

Con respecto al resto, se pueden observar algunos elementos que, aun siendo actividades de crowdsourcing, no son tenidos en cuenta por todos los autores. Destacan fundamentalmente dos: el crowdfunding (H4, G4) (no presentes en Kleeman, Brabham y Burger-Helmchen y Pénin) y el soporte entre clientes (K7), diferenciándolo del crowdstorming (Howe, Brabham y Geerts). El resto de los elementos analizados coinciden en mayor o menor grado, siendo las tipologías de Brabham y Geerts las que presentan un mayor carácter integrador.

En base a los resultados anteriores y a la revisión bibliográfica realizada se propone la siguiente clasificación que intenta recoger las aportaciones y subsanar las carencias de las anteriores. Para cada tipo planteado se mencionan los elementos previos relacionados:

1. *Crowdcasting* (EG1). En este tipo de iniciativas, un individuo, empresa u organización plantea a la multitud un problema o tarea, siendo recompensado quien lo resuelva antes o mejor. Innocentive es un ejemplo paradigmático: en esta plataforma se permite la propuesta de tareas como la elaboración de un tratamiento que permita reducir el coeficiente de fricción en las piezas de metal hechas de acero inoxidable, premiando dicha propuesta con 10.000 dólares (Doan; Ramakrishnan; Halevy, 2011). Este nuevo tipo engloba a H1.2, K3, B2, G1 y BH1.
2. *Crowdcollaboration* (EG2). Considera las iniciativas en las que se produce una comunicación entre los individuos de la multitud, mientras la empresa iniciadora del proceso queda relativamente al margen. Los individuos aportarán su conocimiento para resolver problemas o plantear ideas de forma colaborativa (Doan; Ramakrishnan; Halevy, 2011) y normalmente no existe una recompensa económica. Podemos encontrar dos subtipos que se diferencian en el objetivo final:
  - a. *Crowdstorming* (EG2.1). Sesiones de tormenta de ideas online, en las que se plantean soluciones y la multitud participa con sus comentarios y votos, como en el caso de la plataforma Ideajam (<http://www.idealjam.net> ). Este subtipo está relacionado con H1.3, K1, K2, G2 y BH1.
  - b. *Crowdsupport* (EG2.2). Los propios clientes son los que solucionan las dudas o problemas de otros, sin necesidad de acudir al servicio técnico o posventa de atención al cliente. La diferencia con EG2.1 es que el crowdsupport busca ayudar, como es el caso de Getsatisfaction (<http://www.getsatisfaction.com>), una plataforma que permite a compañías como Microsoft realizar este tipo de tareas. Este subtipo incorpora el tipo K7.
3. *Crowdcontent* (EG3). La gente aporta su mano de obra y su conocimiento para crear o encontrar contenido de diversa naturaleza (Doan; Ramakrishnan; Halevy, 2011). Se diferencia del crowdcasting en que no es una competición, sino que cada individuo trabaja de manera individual y al final se reúne el resultado de todos. Se pueden encontrar tres subtipos que se diferencian en su relación con los contenidos:

- a. *Crowdproduction* (EG3.1). La multitud debe crear contenido colaborando con otros, como en el caso de la Wikipedia, o de manera individual, realizando tareas de dificultad variable como la traducción de fragmentos cortos de texto o el etiquetado de imágenes, como en algunas tareas de Amazon Mechanical Turk. Este subtipo incorpora a H2, K4, K5, B3, B4, G3, BH2 y BH3.
  - b. *Crowdsearching* (EG3.2). Los colaboradores buscan contenidos disponibles en internet con algún fin. Aunque existen proyectos que se basan en este tipo de tareas, como el Peer to Patent Review, también existen otras de menor tamaño como algunas planteadas en Amazon Mechanical Turk. Este subtipo contempla los tipos B1 y BH2. <http://www.peertopatent.org>
  - c. *Crowdanalyzing* (EG3.3). Es parecido al crowdsearching (EG3.2), con la diferencia de que la búsqueda no se realiza en internet, sino en documentos multimedia como imágenes o vídeos. Un ejemplo sería el proyecto stardust@home, en el que cualquier persona puede buscar muestras de polvo interestelar, analizando imágenes en 3 dimensiones de la sonda espacial Stardust. Este subtipo surge de los mismos tipos que el crowdsearching, refinado tras la consulta de los artículos recogidos en la revisión sistemática de la bibliografía.
4. *Crowdfunding* (EG4). Un individuo o una organización buscan la financiación por parte de la multitud a cambio de alguna recompensa. En el mundo del cine, por ejemplo, la película española “El cosmonauta” está siendo financiada de esta manera: los productores ofrecen a los que les financian promoción comercial o aparecer en los títulos de crédito. En el mundo del deporte destaca el caso del equipo inglés de fútbol Myfootballclub. En este caso, la multitud participa aportando su dinero. Este tipo abarca H4 y G4. <http://www.elcosmonauta.es> <http://www.myfootballclub.co.uk>
  5. *Crowdopinion* (EG5). Se intenta conocer la opinión de los usuarios sobre un tema o producto. Es el caso de Modcloth (<http://www.modcloth.com>), tienda de ropa inglesa donde cualquier usuario registrado puede opinar sobre productos que todavía no han salido a la venta, obteniéndose así información sobre su potencial aceptación en el mercado. La gente aporta su opinión o criterio para realizar valoraciones (Doan; Ramakrishnan; Halevy, 2011). Este tipo se corresponde con H3, K6 y B3. También se correspondería con H1.1, subtipo que Howe (2008) denomina market research. En este caso se trata de iniciativas de crowdvoting donde la opinión del usuario no se



manifiesta mediante un voto, sino mediante la compra y venta de acciones vinculadas a algún resultado próximo como una elección presidencial. Para este tipo de iniciativas de crowdvoting se utilizan plataformas especializadas llamadas “mercados de predicción online”, como Intrade (<http://www.intrade.com>) o inkling markets (<http://inklingmarkets.com>).

La tabla 3.3 proporciona información sobre la cobertura y nivel de integración que presenta la tipología propuesta respecto a los elementos componentes de las anteriores, mencionados por filas. Cada celda asocia cada uno de los tipos “previo” con el correspondiente de la nueva tipología. Puede observarse que el elemento que se detecta con más frecuencia es el tipo EG3.1 crowdproduction y se ha elaborado el nuevo componente EG3.3 crowdanalyzing para incorporar la tarea de búsqueda e interpretación de información en documentos multimedia. Finalmente, todos los elementos previos están reflejados por al menos un componente.

**Table 3.3. Encaje de la nueva tipología con las tipologías estudiadas**

	Elemento previo/propuesto						
<b>Howe</b>	H1.1 / EG5	H1.2/EG1	H1.3/EG2.1	H2/EG3.1	H3/ EG5	H4/ EG4	
<b>Kleman</b>	K1/EG2.1	K2/EG2.1	K3/EG1	K4/EG3.1	K5/ EG3.1	K6/ EG5	K7/ EG2.2
<b>Brabham</b>	B1/EG3.2	B2/EG1	B3/EG3.1/ EG5	B4/EG3.1			
<b>Geerts</b>	G1/EG1	G2/EG2.1		G4/EG4			
<b>Burger-Helmchen &amp; Pénin</b>	BH1/EG1/ EG2.1	BH2/EG3. 1/EG3.2					

## 5. Comprobación de la tipología propuesta

Para testear la validez de la propuesta realizada, se han escogido 15 casos al azar, a partir de una lista de 84 iniciativas de crowdsourcing (Wikipedia, 2011).

Los ejemplos seleccionados son:

- 99designs: plataforma web donde las empresas plantean sus necesidades de diseño gráfico para que sean resueltas por la multitud a cambio de una recompensa económica.
- Article One Partners: comunidad de expertos en tecnología que buscan información sobre el estado del arte de un tema relacionado con una patente nueva.
- BlueServo: los colaboradores pueden visualizar las cámaras de la frontera de EUA y México para detectar inmigrantes ilegales.

- GoldCorp: empresa minera que puso a disposición de la multitud su información técnica y material cartográfico, premiando a aquellos que encontraron nuevos yacimientos de oro.
- IBM: recogió más de 37.000 ideas a través de sesiones de crowdstorming donde participaron clientes, empleados y familiares de empleados.
- The guardian: investigó el escándalo de los miembros del Parlamento del Reino Unido y permitió el acceso a 700.000 documentos para que fueran examinados.
- Juratis: plataforma web estadounidense que permite que los usuarios puedan preguntar y resolver dudas legales.
- Lánzanos: plataforma española de crowdfunding que facilita que cualquiera presente un proyecto y sea financiado por la multitud.
- Pepsi: lanzó una campaña de publicidad en la que los usuarios podían diseñar una lata de refresco. El ganador recibía una recompensa económica.
- reCaptcha: utiliza el sistema Captcha para ayudar a digitalizar libros de texto, a la vez que protege los sitios web de accesos inadecuados (anti-bot).
- setiQuest: analizar señales recibidas del espacio para buscar signos de civilizaciones avanzadas.
- SocialAttire: votar diseños de ropa. – StarMind: se plantean problemas de consultoría de empresas cuya resolución implicará una recompensa económica.
- TopCoder: se plantean retos sobre desarrollo de software y creaciones digitales.
- Userfarm: primera plataforma internacional de video elaborado mediante crowdsourcing.

Como se puede comprobar en la tabla 3.4, todos los casos seleccionados se ajustan a alguno de los tipos propuestos en el presente artículo.

## 6. Conclusión

El crowdsourcing es un fenómeno reciente que ha surgido con fuerza y es susceptible de ser utilizado en cualquier ámbito: empresarial, institucional, educativo, etc. Las iniciativas de este tipo proliferan de forma notable y jugarán un papel cada vez más importante en la web del futuro. Con todo, adolece de un adecuado soporte de investigación relativo a su propia

definición y a la descripción y clasificación de sus manifestaciones (Denyer; Tranfield; Van-Aken, 2008).

**Table 3.4. Contraste de la tipología planteada con los casos seleccionados**

	EG1	EG2.1	EG2.2	EG3.1	EG3.2	EG3.3	EG4	EG5
<b>99designs</b>	X							
<b>Article One Partners</b>					X			
<b>BlueServo</b>						X		
<b>GoldCorp</b>	X							
<b>IBM</b>		X						
<b>Guardian (The)</b>					X			
<b>Juratis</b>			X					
<b>Lánzanos</b>							X	
<b>Pepsi</b>	X							
<b>reCaptcha</b>				X				
<b>setiQuest</b>						X		
<b>SocialAttire</b>								X
<b>Starmind</b>	X							
<b>TopCoder</b>	X							
<b>UserFarm</b>				X				

Como cualquier objeto que es sometido a un análisis para ser clasificado, el crowdsourcing presenta una serie de características que pueden ser empleadas como criterios de clasificación (Doan; Ramakrishnan; Halevy, 2011). En este artículo se ha escogido “la tarea a realizar” como criterio fundamental ya que es el elemento que genera más diferencias: producirá lo que el iniciador de la actividad de crowdsourcing necesita y condiciona el resto de características.

Mediante una revisión sistemática de la bibliografía se han obtenido tipologías de crowdsourcing con el criterio de las tareas a realizar. Tras su análisis se ha propuesto una tipología que recoge e integra las anteriores siendo coherente con ellas. Se propone el tipo crowdanalyzing para atender a nuevas realidades colaborativas de generación de contenido dado que el tratamiento de los documentos en internet (numerosos y de calidad dispar) puede beneficiarse de la “inteligencia colectiva”. Además, esta tipología ha sido contrastada mediante quince casos de iniciativas de crowdsourcing.

Existen algunas limitaciones en el estudio: por un lado, la revisión sistemática de la literatura no ha cubierto obviamente todos los documentos que hablan de este tema, por lo que es posible que alguna tipología no haya sido tomada en cuenta; por otro lado, al ser un concepto dinámico (Schenk; Guittard, 2009), esta tipología tiene una validez temporal limitada por la aparición de nuevos modelos de negocio que hagan uso de la inteligencia colectiva.

Además, existen todavía aspectos del crowdsourcing en los que no existe un acuerdo explícito y que convendría que fueran tratados, como su relación con la co-creación o la innovación abierta. Incluso, hay ámbitos donde no ha sido ampliamente utilizado, como la educación superior, donde podría aportar importantes beneficios para todos los agentes implicados, dando lugar a nuevas posibilidades de estudio. Con todo, se ha aportado una chispa en el debate sobre este fenómeno.

## 7. Bibliografía

- Brabham, Daren C. “Crowdsourcing as a model for problem solving: an introduction and cases”. *Convergence: the intl journal of research into new media technologies*, 2008a, Febr., v. 14, n. 1, pp. 75-90. [http://www.clickadvisor.com/downloads/Brabham\\_Crowdsourcing\\_Problem\\_Solving.pdf](http://www.clickadvisor.com/downloads/Brabham_Crowdsourcing_Problem_Solving.pdf)  
<http://dx.doi.org/10.1177/135485650708442>
- Brabham, Daren C. “Moving the crowd at iStockphoto: the composition of the crowd and motivations for participation in a crowdsourcing application”. *First monday*, 2008b, v. 13, n. 6. <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2159/1969>
- Burger-Helmchen, Thierry; Pénin, Julien. “The limits of crowdsourcing inventive activities: what do transaction cost theory and the evolutionary theories of the firm teach us?” En: *Workshop on open source innovation*, 2010, Estrasburgo, Francia. [http://cournot.u-strasbg.fr/users/osi/program/TBH\\_JP\\_crowdsourcing%202010%20ENG.pdf](http://cournot.u-strasbg.fr/users/osi/program/TBH_JP_crowdsourcing%202010%20ENG.pdf)
- Codina, Lluís. “Una propuesta de metodología para el diseño de bases de datos documentales (Parte II).” *El profesional de la información*, 1997, dic., v. 6, n. 12, pp. 20-26.  
[http://www.elprofesionalde lainformacion.com/contenidos/1997/diciembre/una\\_propuesta\\_de\\_metodologia\\_para\\_el\\_diseo\\_de\\_bases\\_de\\_datos\\_documentales\\_parte\\_ii.html](http://www.elprofesionalde lainformacion.com/contenidos/1997/diciembre/una_propuesta_de_metodologia_para_el_diseo_de_bases_de_datos_documentales_parte_ii.html)
- Corney, Jonathan R.; Torres-Sánchez, Carmen; Jagadeesan, Prasanna; Lynn, A.; Regli, William. “Outsourcing labour to the cloud”. *Intl journal of innovation and sustainable development*, 2010, v. 4, n. 4, pp. 294-313. <http://dx.doi.org/10.1504/IJISD.2009.033083>
- Chesbrough, Henry W. *Open innovation: the new imperative for creating and profiting from technology*. Harvard Business Press, 2003, ISBN: 978 15 785 1837 7

- Delgado-Rodríguez, Miguel; Sillero-Arenas, María; GálvezVargas, Ramón. “Metaanálisis en epidemiología (Primera parte): características generales”. *Gaceta sanitaria*, 1991, v. 5, n. 27, pp. 265-272.  
<http://www.elsevier.es/sites/default/files/elsevier/pdf/138/138v05n27a13140889pdf001.pdf>
- Denyer, David; Tranfield, David; Van Aken, Joan-Ernst. “Developing design propositions through research synthesis”. *Organization studies*, 2008, marzo, v. 29, n. 3, pp. 393-413.
- Doan, Anhai; Ramakrishnan, Raghu; Halevy, Alon. “Crowdsourcing systems on the world wide web”. *Communications of the ACM*, 2011, v. 54, n. 4, pp. 86-96.  
[http://cacm.acm.org/magazines/2011/4/106563-crowd\\_sourcing-systems-on-the-world-wide-web/fulltext](http://cacm.acm.org/magazines/2011/4/106563-crowd_sourcing-systems-on-the-world-wide-web/fulltext) <http://dx.doi.org/10.1145/1924421.1924442>
- ECMT. European Commission for Mobility and Transport. Door-to-door in a click.  
[http://ec.europa.eu/transport/its/multimodal-planners/index\\_en.htm](http://ec.europa.eu/transport/its/multimodal-planners/index_en.htm)
- Estellés-Arolas, Enrique; González-Ladrón-de-Guevara, Fernando. “Towards an integrated crowdsourcing definition”. *Journal of information science*, 2012, April, v. 38, n. 2, pp. 189-200. <http://dx.doi.org/10.1177/0165551512437638>
- FBI. Federal Bureau of Investigation. Cryptanalysts: help break the code.  
<http://forms.fbi.gov/code>
- Geerts, Simone. Discovering crowdsourcing: theory, classification and directions for use. Tesis de máster. Technische Universiteit Eindhoven, Netherlands, 2009.  
<http://alexandria.tue.nl/extra2/afstversl/tm/Geerts%202009.pdf>
- Geiger, David; Seedorf, Stefan; Schader, Martin. “Managing the crowd: towards a taxonomy of crowdsourcing processes”. En: *Procs of the 7th Americas conf on information systems*, Detroit, Michigan, August 4th-7th, 2011. [http://schader.bwl.uni-mannheim.de/fileadmin/files/publikationen/Geiger\\_et\\_al.\\_-2011\\_-\\_Managing\\_the\\_Crowd\\_Towards\\_a\\_Taxonomy\\_of\\_Crowdsourcing\\_Processes.pdf](http://schader.bwl.uni-mannheim.de/fileadmin/files/publikationen/Geiger_et_al._-2011_-_Managing_the_Crowd_Towards_a_Taxonomy_of_Crowdsourcing_Processes.pdf)
- Ghafele, Roya; Gibert, Benjamin; DiGiammarino, Paul. “How to improve patent quality by using crowdsourcing”. *Innovation management*, Sept. 2011.  
<http://www.innovationmanagement.se/2011/09/29/howto-improve-patent-quality-by-using-crowd-sourcing/>

- Grier, David-Alan. “Not for all markets”. *Computer*, 2011, mayo, v. 44, n 5, pp. 6-8. <http://www.computer.org/portal/web/The-Known-World/home/-/blogs/not-for-all-markets>
- Heizer, Jay; Render, Barry. *Principles of operations management*. Prentice Hall, 2010. ISBN: 978 01 361 1446 8
- Howe, Jeff. “Crowdsourcing: a definition”. *Crowdsourcing: why the power of the crowd is driving the future of business*. June 2006. [http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing\\_a.html](http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html)
- Howe, Jeff. *Crowdsourcing: why the power of the crowd is driving the future of business*. Great Britain: Business Books, 2008, ISBN: 978 03 073 9620 7
- Kleeman, Frank; Voss, G. Günter; Rieder, Kerstin. “Un(der)paid innovators: the commercial utilization of consumer work through crowdsourcing”. *Science, technology & innovation studies*, 2008, v. 4, n. 1, p. 5-26. <http://www.sti-studies.de/ojs/index.php/sti/article/view/81/62>
- La-Vecchia, Gioacchino; Cisternino, Antonio. “Collaborative workforce, business process crowdsourcing as an alternative of BPO”. En: Floarian, Daniel; Facca, Federico. *Current trends in web engineering, ICWE 2010 workshops*. Viena, July 5-6, 2010, pp. 425-430. ISBN: 978 3 642 16984 7
- Mazzola, Daniele; Distefano, Alessandra. “Crowdsourcing and the participation process for problem solving: the case of BP”. En: VII Conf of the Italian Chapter of AIS. *Information technology and innovation trends in organizations*, 2010. <http://www.cersi.it/itais2010/pdf/041.pdf>
- Oliveira, Fabio; Ramos, Isabel; Santos, Leonel. “Definition of a crowdsourcing innovation service for the European SMEs”. En: Daniel F. et al. (eds.) *Current trends in web engineering. Lecture notes in computer science*, 2010, v. 6385, pp. 412-416. [http://dx.doi.org/10.1007/978-3-642-16985-4\\_37](http://dx.doi.org/10.1007/978-3-642-16985-4_37)
- Oomen, Johan; Aroyo, Lora. “Crowdsourcing in the cultural heritage domain: opportunities and challenges”. En: *Procs of 5th intl conf on communities & technologies – C&T*. Queensland Univ. of Technology, Brisbane, Australia, 29 June-2 July 2011. <http://www.cs.vu.nl/~marieke/OomenAroyoCT2011.pdf>

- Pénin, Julien. “More open than open innovation? Rethinking the concept of openness in innovation studies”. Working papers of BETA, Bureau d’Economie Théorique et Appliquée, 2008, UDS, Estrasburgo. <http://www.beta-umr7522.fr/productions/publications/2008/2008-18.pdf>
- Pinto-Molina, María; Alonso-Berrocal, José-Luis; CordónGarcía, José-Antonio; Fernández-Marcial, Viviana; GarcíaFiguerola, Carlos; García-Marco, Javier; Gómez-Camarero, Carmen; Francisco-Zazo, Ángel; Doucer, Anne-Vinciane. “Análisis cualitativo de la visibilidad de la investigación de las universidades españolas a través de sus páginas web”. Revista española de documentación científica, 2007, v. 27, n. 3, pp. 345-370. <http://redc.revistas.csic.es/index.php/redc/article/view/157/211><http://dx.doi.org/10.3989/redc.2004.v27.i3.157>
- Pisano, Gary P.; Verganti, Roberto. “Which kind of collaboration is right for you”. Harvard business review, 2008, v. 86, n. 12, pp. 78-86.
- PlanB. El Plan Ballantine’s de Carlos Jean, marzo 2011. <http://prensa.elplanb.tv/>
- Reichwald, Ralf; Piller, Frank T. Interaktive wertschöpfung. Open innovation, individualisierung und neue formen der arbeitsteilung. Wiesbaden: Gabler Verlag, 2006. ISBN: 978 38 349 0972 5
- Schenk, Eric; Guittard, Claude. Crowdsourcing: what can be outsourced to the crowd, and why?, 2009. Technical report. [http://halshs.archives-ouvertes.fr/docs/00/43/92/56/PDF/Crowdsourcing\\_eng.pdf](http://halshs.archives-ouvertes.fr/docs/00/43/92/56/PDF/Crowdsourcing_eng.pdf)
- Schenk, Eric; Guittard, Claude. Le crowdsourcing: modalités et raisons d’un recours à la foule. <http://marsouin.infini.fr/ocs2/index.php/frontieres-numeriques-brest2009/frontieres-numeriques-brest2009/paper/viewFile/60/8>
- Schenk, Eric; Guittard, Claude. “Towards a characterization of crowdsourcing practices”. Journal of innovation economics, 2011, v. 1, n. 7, pp. 93-107. [http://www.cairn.info/article.php?ID\\_ARTICLE=JIE\\_007\\_0093](http://www.cairn.info/article.php?ID_ARTICLE=JIE_007_0093)
- Sloane, Paul. “The brave new world of open innovation”. Strategic direction, 2011, v. 27, n. 5, pp. 3-4. <http://dx.doi.org/10.1108/02580541111125725>
- SuperBowl. Crash the SuperBowl, Nov. 2011. <http://www.crashthesuperbowl.com>

- Vukovic, Maja. “Crowdsourcing for enterprises”. En: Procs of the 2009 congress on services I. IEEE Computer Society, Washington DC, 2009, pp. 686-692. <http://dx.doi.org/10.1109/services-i.2009.56>
- Vukovic, Maja; Bartolini, Claudio. “Towards a research agenda for enterprise crowdsourcing”. En: Margaria, Tiziana; Steffen, Bernhard (eds). Leveraging applications of formal methods, verification, and validation. Lecture notes in computer science, 2010a, v. 6415, pp. 425-434. [http://dx.doi.org/10.1007/978-3-642-16558-0\\_36](http://dx.doi.org/10.1007/978-3-642-16558-0_36)
- Vukovic, Maja; Bartolini, Claudio. “Crowd-driven processes: state of the art and research challenges”. En: Maglio, Paul; Weske, Mathias; Yang, Jian; Fantinato, Marcelo. Service-oriented computing. Lecture notes in computer science, 2010b, v. 6470, p. 733. [http://dx.doi.org/10.1007/978-3-642-17358-5\\_79](http://dx.doi.org/10.1007/978-3-642-17358-5_79)
- Vukovic, Maja; Kumara, Soundar; Greenshpan, Ohad. “Ubiquitous crowdsourcing”. En: Procs of the 12th ACM intl conf, 2010, pp. 523-526. <http://dx.doi.org/10.1145/1864431.1864504>
- Wexler, Mark N. “Reconfiguring the sociology of the crowd: exploring crowdsourcing”. Intl journal of sociology and social policy, 2011, v. 31, n. 1, pp. 6-20. <http://dx.doi.org/10.1108/01443331111104779>
- Wikipedia. List of crowdsourcing projects, Febr. 2011. [http://en.wikipedia.org/wiki/List\\_of\\_crowdsourcing\\_projects](http://en.wikipedia.org/wiki/List_of_crowdsourcing_projects)



**CAPÍTULO 4 -**

**Los sistemas de etiquetado social:  
el caso de Diigo**

## 4.1 Introducción

Este capítulo se corresponde con el artículo “*Social bookmarking tools as facilitators of learning and research collaborative processes: The Diigo case*” publicado en la revista *Interdisciplinary Journal of E-Learning and Learning Objects*.

### 4.1.1 Resumen del artículo

La Web 2.0 ha permitido el desarrollo y la aparición de multitud de aplicaciones con una fuerte orientación social. Entre estas aplicaciones, destacan los sistemas de marcado social: aplicaciones que permiten a los usuarios marcar y compartir recursos web, etiquetándolos, convirtiéndose de esta manera también en sistemas de etiquetado social. En este artículo se realiza una descripción general de los sistemas de marcado social, indicando funcionalidades, ventajas e inconvenientes. Se procede posteriormente a describir uno de los sistemas de marcado social más utilizados a nivel académico en Estados Unidos, Diigo, y a compararlo con otro sistema de marcado social muy utilizado a nivel general, Delicious. Mediante un análisis DAFO, quedan plasmadas las características principales, tanto positivas como negativas, del sistema Diigo.

### 4.1.2 Datos de la publicación

El artículo ha sido publicado en la revista *Interdisciplinary Journal of E-Learning and Learning Objects*. Esta revista se centra en la teoría, práctica, innovación e investigación que cubra cualquier aspecto relacionado con el E-learning y con los objetos de aprendizaje, definiendo estos objetos de aprendizaje en sentido amplio incluyendo de esta manera objetos multimedia (audio, video, animaciones, etc.) utilizados para el aprendizaje.

Esta revista está indexada en distintas bases de datos extranjeras: *Cabell's Directory of Publishing Opportunities in Educational, Technology & Library Science*, *Cabell's Directory of Publishing Opportunities in Educational Curriculum & Methods*, *Directory of Open Access Journals (DOAJ)*, *EBSCO*, *EdITLib* (Education and Information Technology - Digital Library) e *Index of Information Systems Journals*.

Los autores del artículos son, en orden de aparición, Enrique Estellés, M<sup>a</sup> Esther del Moral y Fernando González.

- Nombre de la revista: *Interdisciplinary Journal of E-Learning and Learning Objects*
- ISSN: 1552-2210
- Fecha: 2010
- Volumen: 6
- Páginas: 175 - 191

Con respecto a este artículo, es importante destacar que ha sido citado en once ocasiones:

1. del Moral, M. E., Cernea, A., & Villalustre, L. (2013) Connectivist Learning Objects and Learning Styles. *Interdisciplinary Journal of E-Learning and Learning Objects*. 9, 107-124
2. Siemens, R., Timney, M., Leitch, C., Koolen, C., & Garnett, A. (2012). Toward modeling the social edition: An approach to understanding the electronic scholarly edition in the context of new and emerging social media. *Literary and Linguistic Computing*, 27(4), 445-461.
3. Santos, C. Pedro, L. & Almeida, S. (2012). Promover a comunicação e partilha em ambientes pessoais de aprendizagem: O caso do Sapo Campus. *Indagatio Didactica*, 4(3), 64-91.
4. Ovadia, S. (2012). A Brief Introduction to Web-Based Note Capture. *Behavioral & Social Sciences Librarian*, 31(2), 128-132.
5. Pedro, L., Santos, C., Almeida, S., & Koch-Grünberg, T. (2012). Building a Shared Personal Learning Environment with SAPO Campus. In *PLE Conference Proceedings (Vol. 1, No. 1)*.
6. Del Moral, M.E., Cernea, D.A. & Villalustre, L. (2011) Scenari Connettivisti e Progetto di Learning Object Adattabili alla Diversita Cognitiva. *TD-Tecnologie Didattiche*, 19 (2), 102-111
7. Ruffini, M. F. (2011). Classroom Collaboration Using Social Bookmarking Service Diigo. *Educate Review On line*, septiembre, Recuperado de: <http://www.educause.edu/ero/article/classroom-collaboration-using-social-bookmarking-service-diigo>
8. Mu, C. (2011). Impact of New Technologies on Current Awareness Tools in Academic Libraries. *Reference & User Services Quarterly*, Volume 51 (2), 92-97. Recuperado de <http://rusa.metapress.com/content/T86783J565145063>
9. Koch-Grünberg, T.T. (2011). Gameful connectivism: social bookmarking no SAPO Campus. Ponencia presentada en la Facultad de Comunicação e Arte Universidade de Aveiro (Portugal). Recuperado de <http://ria.ua.pt/bitstream/10773/7506/1/245076.pdf>
10. Roig Arderiu, J. (2011). La Web 2.0 en l'ensenyament de les matemàtiques. Memoria de Master. Universitat Politècnica de Catalunya. [http://upcommons.upc.edu/pfc/bitstream/2099.1/15174/1/72409\\_Memoria.pdf](http://upcommons.upc.edu/pfc/bitstream/2099.1/15174/1/72409_Memoria.pdf)
11. Mehlenbacher, B., Holstein, K., Gordon, B., & Khammar, K. (2010). Reviewing the research on distance education and e-learning. In *Proceedings of the 28th ACM International Conference on Design of Communication* (pp. 237-242). ACM

## 4.2 Artículo

### Social bookmarking tools as facilitators of learning and research collaborative processes: The Diigo case

Enrique Estellés-Arolas

Department of Management, Technical University of Valencia, Valencia, Spain

Esther del Moral

Department of Education Science, University of Oviedo, Spain

Fernando González-Ladrón-de-Guevara

Department of Management, Technical University of Valencia, Valencia, Spain

#### **Abstract**

Web 2.0 has created new applications with remarkable socializing nuance, such as the SBS (Social Bookmarking Systems). Rather than focusing on the relationship between users, the SBS provide users with the necessary tools to manage and use information that can later be shared.

This article presents a description, analysis and comparison of different SBS, which are categorized as web applications that help to store, classify, organize, describe, and share multi-format information through links to interesting web sites, blogs, pictures, wikis, videos and podcasts. Also emphasized are the advantages for learning and collaborative research that SBS produce.

In this case, Diigo will be specifically studied for its contribution as a metacognitive tool. Diigo shows the way each user learns, thinks and develops the knowledge that was obtained from the information previously selected, organized and categorized. Thus, the information becomes highly valuable, and knowledge is cooperatively built. This knowledge induces collaborative learning and research, since the tags that describe marked resources are shared between users. Consequently, they become meaningful learning resources that provide a social dimension to both learning and online research processes.

#### **Keywords**

Social Bookmarking Systems, Folksonomies, Collaborative Research, Collaborative Learning.

## 1. Introduction

The rise of Web 2.0 tools has led to the rapid development of a number of applications that enhance collaborative work. These include social bookmarking systems (SBS) that provide users with the reference (marked), description, classification and the possibility to share resources with other users.

In this paper, all this apps (SBS) are addressed: functional features, nature and restrictions. Next, one of these tools, Diigo, is analyzed, taking into account its possible uses and benefits for researching and education. In the fourth section, a comparison between different methods of storing bookmarks (Diigo, Delicious (both SBS) and the old fashion way) is done in order to highlight the advantages of Diigo and its differentiating features. This comparison is completed through the SWOT study done about this kind of tools in a 30 user community. To close, some conclusions and possible future lines of investigation are enumerated.

## 2. Social Bookmarking Systems

The Social Bookmarking Systems are Web 2.0 tools that allow users to store, classify, organize, describe and share links to interesting web sites, blogs, pictures, wikis, videos and podcasts. They also guarantee access from any site to the conventional container of "favorite" links, as well as the possibility to share them with other like-minded users through blogs or RSS technology, for instance.

Depending on the web resources bookmarked, we can talk about different types of SBS. There are SBS focused on collecting web sites (Diigo, del.icio.es, Mister Wong, Blinklist), some focused on collecting news (digg.com), and others on pictures (Flickr) or even on bibliographical references (CiteU).

### 2.1 Characteristics of every SBS

Regardless of the type of content tagged, all the above-mentioned SBS have some common characteristics. The two main common characteristics are the basic unit of referenced information and the use of tags.

To begin with, the basic unit of referenced information used by any SBS is a set of three elements called 'triple' that is represented this way: (user, resource, {tags}) (Cattuto, 2006). This unit, that defines the way the SBS work, indicates that a user has marked a specific resource with a set of concrete tags.

As for the use of tags, it clearly implies the use of folksonomies. A Folksonomy, a term coined by Thomas Vander Wal and which is a combination between "folk" and "taxonomy" (Smith, 2004), is an organic system of organization and a way of social classification using tags. Due to this, any SBS can also be seen as a Social Tagging System.

The folksonomy enables users to organize their bookmarks in a meaningful way and search for resources associated to specific tags. Resources can also be classified according to the amount of users that have tagged them.

Unlike taxonomies (or classifications), where there are multiple types of hierarchical relationships, folksonomies are not based on hierarchies: there are no explicitly indicated relationships between the terms included. They are just the keywords that a group of users have used to describe a specific content (Mathes, 2004; Hamond et al., 2005). The social usage of tags is one of the simplest ways of adding high-semantic-valued metadata to the content.

When a web resource is tagged, SBS enable users to describe its content by adding a set of data known as metadata (data about data). Depending on the SBS, this set of data or metadata contains the following elements (Zubiaga et al., 2009):

- Tags or terms that define and feature the resource. These can be names, acronyms, numbers or any chain of text with no format or meaning restriction.
- Notes or comments: a short text freely describing the content of the resource.
- Highlights: parts of the resource marked as relevant.
- Reviews: texts freely assessing the content of a resource.
- Ratings: personal marks or punctuation indicating whether users liked a specific resource or not in a scale from 1 to 5, for instance.

In this way, folksonomies add high-semantic-valued metadata, which is specially relevant. In academic or research contexts, folksonomies help a specific scientific community of experts to add value to specific learning objects that are significant for collaborative projects. Thus, they help to enrich the learning community by creating and sharing sources of document resources.

According to Millen et al. (2005), other common characteristics are:

- They enable users to create collections of bookmarks individually, classifying them as private (available only for the owner and for those users or groups he wants to invite) or public (available for everybody). Thanks to this characteristic, like-minded users can recover those collections by consulting categorized or tagged links.
- They help to create networks or groups of users interested in similar issues that share links through tag clouds, links to blogs and the possibility to subscribe through RSS to a specific user's account or to tags of interesting contents.
- Users can easily access them from any computer with Internet connection.
- They provide web browser complements that help to store and describe links.
- They use tags: keywords associated with a specific resource that are assigned by users.
- They include pivot browsing. This is a way of exploring, or re-orienting the selection of bookmarks and discovering information by navigating the collections of bookmarks filtered by users and tags (Millen, Whittaker & Yang, 2007; Bateman et al., 2009.)

Within the new functionalities of the different SBS, it must be also considered the storage of a 'snapshot' of the resource in the server or the suggestion of tags depending on the textual analysis of the content of the resource.

## **2.2 Functions and Restrictions**

According to all that has been explained, SBS are useful tools for:

- Managing research groups focused on a specific topic. Researchers navigate the information that has been tagged by the 'collective intelligence' of those users that tagged and stored it previously.
- Organizing and managing relevant information for professors and researchers and also for university students. Therefore, folksonomies become a powerful tool generating knowledge.
- Organizing, communicating and updating bibliographical lists or recommended readings, adding value to the shared information.
- Managing the information collected in any stage of a research process through the use of complements as Zotero. Their collaborative nature make them perfect tools for the cohesion of research groups.

- Searching for information directly related to the issue of interest for the group and the ability access it. It has been proven that when looking for information via the links in SBS, like del.icio.us, newer or more updated contents of better quality can be found, compared to those that can be found through other search engines (Yahoo!) or directories (ODP) (Kolay & Dasdan, 2009). In spite of this, according to Heyman (2008), 25% of the content collected by del.icio.us was not indexed by Yahoo!

Another interesting aspect about the usage of this tool is the fact that each member of a learning community can contribute to improve it. This is significantly relevant in the academic and research field, where collective intelligence undoubtedly favours the advance and development of knowledge. By adding each user's contribution, the value of the knowledge increases. In this way, it is possible to learn from others simply by following the itineraries others have marked. Nevertheless, it must be taken into consideration that bookmarks have some technical restrictions. For example, a lack of homogeneity and agreement on how to define tags gives rise to ambiguities as Mathes (2004) points out: the use of subjective keywords (excessively personal ones that do not have the same meaning for the rest of users); the use of singular and plural words; the inconsistent usage of capital letters in different languages; the use of simple or complex words to define similar things, etc. In an attempt to solve these problems, in certain SBS there has been a common agreement on vocabulary. However, this solution has also its drawbacks, because sometimes the same tag is used with different meanings and the use of synonyms and acronyms leads to a greater confusion, etc.

Despite the above-mentioned difficulties, SBS are useful for collaborative work because links are shared and metadata are cooperatively built.

Currently, experts are working to make SBS more powerful, enabling combined search techniques that integrate conventional search engines functions with those of SBS. An example of this is the plug-in (bookmarklet or browser extension) of the browser incorporated in the search toolbar that enables to associate Google's and Diigo's search.

### **3. Diigo case**

#### **3.1 Description**

Diigo, an acronym for 'Digest of Internet Information, Groups and Other stuff', was launched in 2006 (24/07/2006). After three years of working, Diigo company acquired Furl



(03/09/2009), which enabled it to grow in the field of SBS. Thanks to this growing, it has been recognized by the American Association of School librarians (AASL) as one of the Best Websites for Teaching and Learning (2009), and referred to as a list of 'tools and resources of exceptional value to inquiry-based teaching and learning.'

Tags that define Diigo in Crunchbase give an idea of what Diigo exactly is: "ad-supported-software", "social-bookmarking", "social-annotation", "social-information-network", "web-mark-up", "web-highlighter" and "web-sticky-notes".

Diigo is an application that allows the use of what is known as 'social annotation' through social bookmarking (SB), text annotations in-situ (in the web itself), tags describing the website, clipping (which allows videos to be marked), pictures or Flash animations- and a search in the whole text of the annotated pages (Diigo-1, 2006). All this information is stored in an Internet server allowing users to work with it from any computer with Internet connection, so that it is possible to share that material with other users. It is similar to Delicious, whose bookmarks can be imported by Diigo, but it has additional features that allow users to organize and show their presentations of bookmarks online through interactive slides that are open to public comments and annotations.

Diigo is also a social net, but with some peculiarities. It is an information social net, whose main objective is not socializing the user, but providing him with high-quality tools to recover, highlight, organize and find information mainly for research tasks and for sharing it with other users. It allows a close relationship between its two main components: users and information. The result of the possible relationships (user-user, user-information, information-information) achieves an improvement of the knowledge users share and an increase of the amount of available content. On the other hand, it creates social connections based on preferences about specific type of information, allowing high-quality intellectual exchanges.

The meaning of Diigo suggests the different ways individuals can use it. Depending on that use, Diigo can be defined as a group research tool, as a sharing-knowledge community or as a site with social contents (Diigo, 2009.)

Diigo allows effective and collaborative research because results can be shared by adding notes to the marked webs (electronic sticky notes) or highlighting. By doing this, a research team, a class, a club or any other type of groups can constitute a group in Diigo so that the

users can share resources, relevant outcomes about an issue, or comments. As a site with social content, Diigo is based on the use of tags and online annotations about pages in order to make a repository of quality content, filtered and commented by the community (Heymman et al., 2008.) Thus, a user can access a web site and see who else has marked it and which other sites with similar content have been found. This way of navigating (from link to link) is called ‘social browsing.’

Finally, Diigo can be also understood as a community of users that share information. In this sense, inside Diigo ‘you are what you highlight’: the links you mark, the tags you use to describe them and the annotations you make. With all this information, Diigo enables people to connect very differently. Especially interesting is the fact that a user can be connected to ‘people like me’ (matching based on recent bookmarks), so that he or she can meet like-minded users that are connected or interested in the same areas.

Regardless of the Diigo's various uses, Diigo provides users with a set of tools to manage bookmarks in order to work individually or collaboratively.

### **3.2 Functions for individual work**

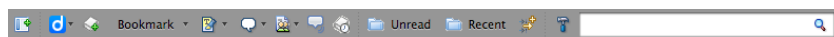
Clearly, the main feature for individual work is, the capacity for managing bookmarks. In regards to this feature, Diigo offers 3 functionalities:

1. Importing bookmarks. Diigo imports the favorite sites of the browser as well as those of several SBS (for example -Delicious, Simpy, Blinklist or Connotea). It also allows the importing of links that might have been stored in Google NoteBook.
2. Exporting bookmarks. Diigo allows the download of a file with marked resources in Internet Explorer, Netscape, RSS, CSV format or the format used by Delicious.
3. ‘Save to del.icio.us’. Apart from exporting bookmarks with delicious format, as has already been stated, Diigo makes it possible for any web resource marked with Diigo Toolbar or Diigolet to also be stored automatically in del.icio.us. At the same time as a resource is marked in Diigo, it can also be marked in other SBS and even in the same browser simultaneously.

As for individual work, Diigo offers a series of complements for browsers that allow the marking of resources:

1. Diigo Toolbar. This toolbar can be installed in different browsers (Explorer, Firefox, Flock and Chrome). It has the following functionalities: it marks new resources

(including describing tags), highlights parts of the web, marks the resources marked as ‘non-read’ and allows quick access to the resources stored in Diigo. This can be done in two different ways. One way is by using the ‘smart folders’, which are icons that display non-read resources, when the user clicks on them. The other way is by using the ‘sidebar’, which opens a small window embedded in the left of the page from where the user can start navigating through the resources stored in Diigo.



**Figure 4.1. Diigo toolbar**

2. Diigolet is a complement similar to the Diigo Toolbar, and it can be applied to any browser (it is especially useful for those browsers incompatible with the Diigo Toolbar). It is a small script developed in javascript that creates a virtual toolbar associated to the web in which it is executed (when the user exists that page, Diigolet disappears). From this virtual toolbar the user can mark resources, highlight webs, add notes and comments and access Diigo web.



**Figure 4.2. Diigolet virtual toolbar**

3. ‘Post to Diigo’. This complement must be put in the “favorites” toolbar that all browsers have. By doing this, when a resource that the user wants to be marked is found, the user only has to click on the button ‘Post to Diigo’ to add that resource to his bookmarks. It is actually a script of javascript that opens the Diigo page ‘Add new Bookmark’ and fills in the data of the title, etc. based on the available metadata of the resource.
4. Button ‘add to Diigo’. This button must be placed next to publications (blogs, webs, news, etc.) and allows a user to mark that publication directly in Diigo only if is a Diigo user.



**Figure 4.3. Button ‘add to Diigo’**

### 3.3 Functions for team work

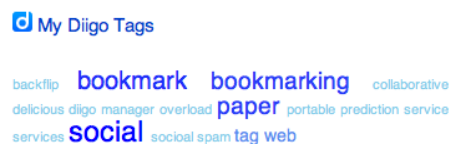
1. Enhanced linkrolls. This is a list of marked web resources that can be shown in a user’s personal web, for example. It can show the last resources or they can be filtered by tags.

This complement enables one to share marked resources so that visiting users, Diigo users or not, can see other users' comments and annotations when accessing a marked web. Through a web form, Diigo allows the user to define the number of resources to be shown, the colors to be used, etc. To summarize, it allows the customization of the list of resources and the creation of the corresponding code to be shown in any other web. As for the collaborative work, this tool keeps the visitors of the web where it is being used, informed about the last findings or interests of the user.



**Figure 4.4. Enhanced linkrolls**

2. Diigo TagRolls. Diigo is able to create tag clouds that may be inserted in a personal web or any other type of web. Like enhanced linkrolls, Diigo provides a form to customize the tag cloud and generate the corresponding code. With this tool, a user can show in a simple and intuitive way the topics in which he is interested or on the ones he is currently working. The visitors can access his marked resources and start to interact.



**Figure 4.5. Diigo tagrolls**

3. Send to blog. Diigo offers a button that complements the characteristics of highlighting text or adding notes. It sends the selected text to a blog that has been previously configured. This way the contents that have been marked, highlighted or annotated

enable visitors to the blog where those content are published to access. Thus, communication between users and collaborative work is enhanced.

4. Auto blog post. This function works as the previous one, but in an automatic and periodic way.

### **3.4 Applications for learning and research**

In the university and research fields it must be pointed out that Diigo:

- Enhances the cohesion of research groups on specific issues by navigating through information that has been tagged by researchers and/or users.
- Enables the organization and management of relevant information for professors as well as researchers, university students, etc. building knowledge cooperatively.
- Makes the organization, communication and updating of bibliographical references or specialized readings of interest more dynamic. It also makes specialized readings and references to which anybody can subscribe and re-tag incorporating new nuances.
- Helps to manage information collected in the different stages of a research work, also using complements such as Zotero.
- Tags in a specific area are more valuable than in general contexts (with more meanings), because a specific context provides additional value: its own specificity and the one given by other tags in its context (Alonso Arevalo, 2009).
- Enables one to visualize the actual interests of a researcher through his tag cloud.
- Favors team work by matching synergies of a specific research group.
- Makes the spreading of ideas between interdisciplinary fields easier.

It offers new opportunities for the learning and building of knowledge. When a user subscribes to watchlists of important researchers or scientists he or she can learn by following their bookmarking system. This is achieved not only through contents since the visitor can catch the researcher's meta-learning process (or the ways he has learnt)- turning them into itineraries of efficient thought that can be extrapolated. These activities include the careful consideration to develop or interpret the meaning of each bookmark and build knowledge cooperatively (Singh et al. 2007). Diigo provides students with a valuable opportunity to learn about their own learning process and identify the aspects of the information they find

relevant. It also make students become aware of their own criteria when they have to tag or categorize them. The very act of refining and defining the tags they use is a valuable feedback in itself.

The professor has to provide the students with proper structure to guide them towards the discovery of their own processes. This is the reason why the use of significant personal tags should be considered as metadata, so the cognitive abilities students apply when they learn become visible. The simple act of tagging a resource or a learning object indicating its objective can help students to think about their own cognitive style, or the way they learn.

Diigo allows students and/or researchers to learn from other members of the learning community when they adopt as their own the more efficient bookmarking structures and strategies used by other colleagues or professors with whom they share resources. It helps them to think about their learning process or metacognitive development through the analysis of how each individual use them according to their particular learning styles.

Diigo also enhances the development of the following wide-ranging capabilities:

- a. Information search and management. Due to the great amount of information that can be found on the web, it is necessary to make a selection in order to detect and distinguish true, reliable and rigorous information. Using Diigo can be a time-saving strategy. It involves identifying what is considered important for a specific community, taking as a starting point the opinions of each member of the group. Using it enhances the development of the very much demanded digital competencies, such as information search and management. Furthermore, the very act of marking a page means that the user is categorizing, summarizing and assessing the information it contains. When students are taught how to mark resources, they are being given a powerful strategy to know how to distinguish valid information by applying criteria to filter it.
- b. Information analysis. Diigo tagging is based on a particular way of understanding information starting from mental maps. An interesting didactic application to be implemented in learning contexts could be an exercise that involves collecting items in order to analyze the value of a web page or a web resource such as: authorship, reliability, scientific rigor, educational potentiality, etc., and ask students to justify their decision to select it as interesting and share it with others, or not.

- c. Categorizing information. Its ability to categorize, organize, establish relationships, describe resources, etc., can be helpful to learn specific contents of any discipline, because it favors the understanding of key concepts and their categorization.

Finally, on the topic of social building of knowledge, by using Diigo as a social bookmarking tool, the contributions of each member of a learning community are enhanced. Building knowledge collectively makes its advance and development possible, especially in a learning and scientific context. Users take part by establishing and sharing what they know from different approaches. By using Diigo, you can learn from others. By following others' bookmark itineraries, both professor and student can share information in a two-way exchange, or a whole learning community (researchers, teaching centre, etc.) can take part in a collaborative project, which can be developed virtually, breaking space-time barriers. Also, a collaborative online database is a cognitive tool that enhances the knowledge building process (Rosen & Rimor; 2009, 189.)

### **3.5 Examples of academic use**

Besides the benefits and advantages that Diigo offers, there are real academic situations where this Web 2.0 tool is used successfully. In table 4.1, some case studies have been collected and described.

The efforts of many teachers and researchers in order to improve in their jobs also results in the creation of user communities where they can comment and share experiences. For example, in Diigo Groups exists many focused on the possible applications of Diigo to education. Two of these groups are:

- "Technology Enabled Learning & Teaching @ UNSW", which is focused on applications, examples, case studies and papers discussing the use of contemporary educational technologies in university learning and teaching practice. 109 users belong to this group, and it has 1955 resources bookmarked.
- "Diigo in Education", where their members share their classroom usecase, ideas, reviews, features, and wishlists for making Diigo a great resource and platform in teaching and learning. In this case, 4889 users belong to it sharing 4063 bookmarked resources.

**Table 4.1. Examples of academic use of Diigo**

University/ College	Knowledge area/ Research group	Application/ Students
Kansas State U.	Cultural Anthropology. Digital Ethnography Working Group	In a class of 200 students, they use Diigo in order to keep track of teaching resources.
Northeast Lakeview College	Introduction Sociology	Online collaborative research replaces the traditional manual. Students research about the concepts and add comments on them.
Master	Photojournalism and documentary	It is used as a dynamic way of sharing links and resources for developing collaborative group research projects. Each project has its own tag and all the students together add links to it. After that, the material is distributed in the classroom and each one has to read selected papers and then develop a summary that will be shared among the other students. In this way, all the class will have a global idea of the research issue.
	Teacher Education	It is used for sharing and comment links and resources of specific issues.
Concordia U.	NEA. National Education Association	Researching about interesting issues. Resource and bibliographic references searching.
Technology Information Center for Administrative Leadership (TICAL)	Teacher Education Program	To share resources specific to course work using the List tool
University of Sheffield	History	Two classes of 10 students per class. They developed an online resources list for their weekly seminars over a semester.



## 4. Comparative analysis of SBS

As previously stated, the main goal of SBS is resource bookmarking and description. This involves storing a link and describing it using metadata. Storage first started in September 1993 as something new inside the ‘Mosaic’ browser. In this case, link storage was called ‘Hotlists’. Then, with Netscape browser (version 1.0, December 1994) this storage of links was called ‘Bookmarks’ and was also called ‘Favourites’ in Internet Explorer (July, 1995). There were some proposals such as SyncIT (1998) to synchronize the favourites of a browser with a web storage system.

The first collaborative attempts regarding links were links directories, where taxonomies were elaborated. Some of these are relevant such as Open Directory Project, Zeal or others for commercial purposes as Yahoo (Hammond et al. 2005.) These bookmarks were improved so much to the degree that javascript was added giving rise to the so-called ‘bookmarklet.’ These are simple links that can be aggregated as favorites, but they incorporate javascript code providing them with extra functions.

After these attempts, social managers of links were created. These non-randomly stored links were found by means of crawlers or robots, and if not registered, were identified with tags and assessed by users, making them available for others. In this context, in April 1996, itList surfaced. It included public and private bookmarks. Then similar services such as Backflip, Blink, BookmarkBox, Bookmarks Plus, Clickmarks, Clip2, Murl, MyPassword.net, Oneview, Hotlink and Quiver appeared (Cf. Extras – itList and Other Bookmark managers | LLRX.com.) Some of these services stopped working after the dot-com boom, but they allowed for the organization of bookmark folders, forwarding those marked by e-mail, along with some additional functions. Finally, a new era began in 2003 with the coming of Delicious and the rest of SBS that are shown below. Here there are various social bookmarks classified by marked resource.

There are a set of tools focused on sharing bibliographical references: Connotea, an open source tool launched in 2004 by Nature Publishing Group; CiteUlike by the University of Manchester - even though presently it is being promoted by the publishing house Springer Verlag; Bibsonomy by the German university of Kassel and 2collab promoted by Elsevier.

**Table 4.2. Comparing SBS**

Type of resource	Bookmarks	Description
General web links	Blackflip (1999), Balatarin (2006), Blinklist (2005)	Web sites references through bookmarks (the object of study in this article.) Most of them allow for synchronization with del.icio.us and import browser favorites. They enhance bookmarks export to be used in other SBS.
News	Digg.com (2004), Meneame.com (2005), Reddit (2005), SpicyBookmark (n.d.), Propeller (2006), Newswine (2005)	Focused on the social bookmarking of specialized literature, news and blogs contributions.
Bibliographical references	2collab (2007), Mekentosj Paper (2001), Mendeley (2007), My NCBI (n.a.) and Zotero (2006)	In these social net of references, folders can be shared, users can create groups, start discussions, include the researchers' CV and profile. If a reference of a specialized area has been aggregated to the manager by many authors, it becomes more and more 'popular' because many experts in that area have found it interesting. It can also help to find out what other researchers interested in that consulted resource are reading and make digital libraries more personal, sociable and integrated places.
Pictures	Flickr (2004), vi.sualize.us (2007), weheartit.com (2008)	
Blogs	Frassle (2003)	

In table 4.2 there are websites such as Digg, Reddit and Propeller that are focused on the social bookmarking of items associated with news (politics, sports, technology, etc.) These services offer headlines of each piece of news and foster users' comments. These are different from general social bookmarks because they are focused on specialized literature and contributions in blogs more than on websites. As a consequence, they can be an

important source of news and they also offer the possibility of taking part in discussions by adding personal comments about interesting news.

#### **4.1 Diigo, Delicious and conventional bookmarking**

Comparing Diigo's tool with conventional bookmarking tools and Del.icio.us will help to understand better Diigo's innovation, which is the most used and widespread bookmarking tool (Diigo-2, 2006.)

One of Diigo's strong points is its highlighting and annotation functionalities (by Sticky notes), which del.icio.us and conventional bookmarking lack. Now Diigo becomes a tool that allows a user to:

- Highlight and annotate while navigating
- Extract and collect automatically highlighted texts of a set of webs associated with a specific topic.
- Interact and cooperate by sharing those highlighted texts and annotations by other users.

Enhance automatically the integration with other blogs and other communication tools such as twitter (in this case, Delicious has some of these functionalities but with some restrictions.)

The table 4.3 shows comparative functions of these tools.

It can be inferred from the table below that Diigo has a set of functions that enhance its versatility and capacity as SBS against other consolidated tools as Delicious and especially conventional bookmarking. The strong point of tools like Delicious is the great amount of people that use it. In fact, most of the SBS (Diigo included) offer the possibility to automatically synchronize with Delicious to prevent the user from abandoning.

Diigo is competing with del.icio.us (SBS reference), especially in countries like India, The United States, China and Germany as shown in figure 6. It shows the distribution of the base of the Diigo users (in these countries), these users' intensity of use, and the position of the service in that country.





**Table 4.3. Comparative between Diigo, Traditional Bookmark and Delicious (Adapted from Diigo help)**

Feature	Diigo	Conventional bookmarking	Delicious
Access from anywhere at anytime	Yes	No	Yes
Simultaneous bookmarking in the application, in other bookmarking tools and in local folders	Yes	No	No
Search: title, tag, highlighted, whole text, users	Yes	No	Yes (except for the search of whole text)
Establishment of all the elements as private. Each user decides what is to be shared	Yes (it can be established by default)	No, there is no possibility to share	Yes (it cannot be established by default)
Organization by tags	Yes	No (organization by folders and sub-folders)	Yes
Storage of a copy of the marked resource in addition to the link	Yes (it can be established a versions record)	No	No
Marking pictures	Yes	Pictures cannot be marked, only webs	No
Easy re-organization and editing of the bookmarks: tags editing and bookmarking editing or deleting	Yes	It can be re-organized, but it is a complex process	Limited
Possibility of marking resources marked as 'non-read' for a later revision	Yes	No	No
Bookmark import indicating	Yes	Import limitation.	Limited. Done by a third party

Relación entre el crowdsourcing y la inteligencia colectiva: el caso de los sistemas de etiquetado social

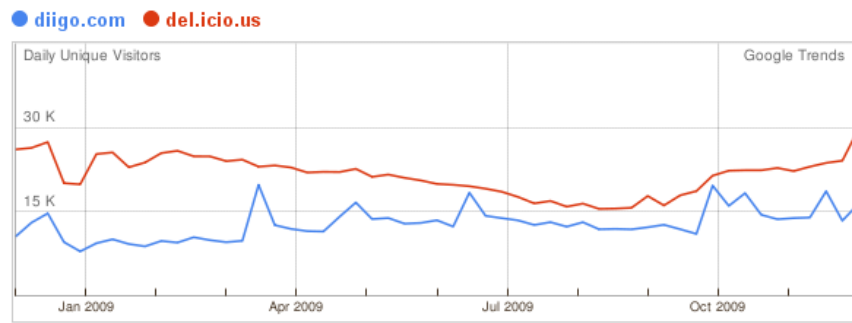
its tags, title and privacy characteristics		Conventional browsers don't use tags or descriptions	extension
Bookmarking status, which allows to mark resources although the connections to Diigo servers is going to break	Yes	No	No
Filtering lists of marked resources by adding or deleting tags and/or users	Yes	No	No

Some of the mentioned functions such as highlighting, aggregation of notes, synchronization with references marked with other SBS (del.icio.us, Blinklist, Connotea, Furl and Simpy) have led to positive assessment. This is proved on figure 7, where a comparison between daily worldwide visits to Diigo vs. del.icio.us is displayed using Google Trend bar.

Country	% User	% Pageviews	Rank in country
 India	34.8%	46.1%	5192
 USA	19.8%	14.6%	38514
 China	7.1%	5.2%	35034
 Germany	5.3%	6.0%	47855
 Pakistan	3.6%	4.6%	11127

**Figure 4.6. Diigo users' database by countries (Adapted from Dataopedia.com)**

The next table displays a synthesis of the most relevant aspects of Diigo tool after applying SWOT Analysis methodology (Strengths, Weaknesses, Opportunities, and Threats), based on the analysis of the content of the assessment of a training community (consisting of 30 members). In the following display, the great contributions of Diigo to favor collaborative research processes, detect technical restrictions that it still has, and enumerate the potential applications that are still to be explored can be identified.



**Figure 4.7. Daily traffic during the year 2009 in Diigo and del.icio.us (Google Trends)**

## 4.2 SWOT

## 5. Conclusions

Virtual environments or communities that foster learning and research from a collaborative approach and introduce new ways of working that highlight the social dimension of knowledge are becoming extremely valuable. By allowing interaction and cooperative problem resolution processes, they become a collaborative social space (Del Moral & Cernea, 2006.) In addition to this, the use of SBS helps to contextualize the learning process and enhances its meaning.

Virtual communities become closely linked groups because each member tries to achieve common objectives, turning the groups into powerful communities with solid internal relationships. Thus, they become important social networks that have great advantages derived from each member's assets.

Collaborative tagging and/or social bookmarking of learning resources foster a context of personalized social learning. From a constructivist point of view, tags shared by users become significant learning resources providing the teaching/learning process and on-line research with a social dimension. The user gets consciously involved in the creation of tags assigning new meanings to the shared resources. This process generates new collaborative learning contexts (Cernea, DelMoral & Labra, 2008.)

**Table 4.4. SWOT analysis**

	<b>Strengths</b>	<b>Weaknesses</b>
<b>Internal analysis</b>	<p>Technical aspects:</p> <ul style="list-style-type: none"> <li>● Intuitive interface, close to human relational thinking.</li> <li>● Synchronization with del.icio.us, blinklist.</li> <li>● Promotion of user interaction with the content based on a pop-up dynamic menu.</li> <li>● Excellent tool to combine notes and bookmarks.</li> <li>● Quick execution with a comprehensive links search engine.</li> </ul> <p>Accessibility:</p> <ul style="list-style-type: none"> <li>● Marked resources on the web that can be found from anywhere.</li> <li>● The bookmarking of resources can be private or public.</li> </ul> <p>Communication:</p> <ul style="list-style-type: none"> <li>● It allows adding comments about the visited webs, identifying who did it and what was highlighted.</li> <li>● Cloud tags show the topics the user is interested in.</li> <li>● It makes efforts dynamic and profitable by allowing to see others' assessment about specific webs .</li> </ul>	<p>Technical aspects:</p> <ul style="list-style-type: none"> <li>● It makes some browsers slower.</li> <li>● Not very dynamic, it takes 20" to finish a task.</li> <li>● It needs to start a session after marking each new resource.</li> <li>● Not all the utilities, colors, etc. can be customized.</li> </ul> <p>User identification:</p> <ul style="list-style-type: none"> <li>● It tracks the user that is bookmarking.</li> <li>● It "compels" to share.</li> <li>● It is necessary to create a new account identifying the user, who cannot be anonymous.</li> </ul> <p>Communication:</p> <ul style="list-style-type: none"> <li>● It doesn't allow an instant feedback between users that add comments.</li> </ul>

Relación entre el crowdsourcing y la inteligencia colectiva: el caso de los sistemas de etiquetado social

	Opportunities	Threats
<b>External analysis</b>	<p>A research tool</p> <ul style="list-style-type: none"> <li>• It is extremely useful for on line research.</li> <li>• It enhances collaborative research projects.</li> <li>• It helps to manage research tasks: selection and categorization of interesting bibliographical sources.</li> <li>• It emphasizes the collaborative dimension based on the shared use of bookmarks.</li> </ul> <p>A learning tool</p> <ul style="list-style-type: none"> <li>• It shows expert bookmarks systems, whose itineraries can be considered as a reference.</li> <li>• It makes cognitive abilities visible for organizing and categorizing information.</li> <li>• It develops competencies – search, management, analysis and categorization of information.</li> </ul> <p>Social building of knowledge</p> <ul style="list-style-type: none"> <li>• The total amount of efficient shared bookmarking strategies enhance knowledge learning and development.</li> </ul>	<p>Social semantics vs. confusion</p> <ul style="list-style-type: none"> <li>• A lack of homogeneity and agreement on the definition of tags. It leads to ambiguities.</li> </ul> <p>Constant updating and change</p> <ul style="list-style-type: none"> <li>• The constant improvement of SBS features make them obsolete and other new systems arise dispersing users, forcing them to constantly migrate.</li> <li>• Incompatibility or lack of entire permeability (import-export) among all the SBS.</li> <li>• Incipient developments to combine its use with conventional browsers, which enhances accessibility.</li> </ul>

Virtual environments where SBS are used are based on constructivist principles, which foster the migration from an intrapersonal learning process to an interpersonal process with a social dimension.

Personal interactions that arise spontaneously through shared annotations make and strengthen the collaborative learning process and make users continuously think about the



relationship between the resource and the tag. As a consequence the conceptual socialization of learning resources is enhanced (Reichel et al., 2006.)

Diigo particularly fosters the cohesion of research groups by monitoring information tagged by different users. It adds more dynamics to the organization, communication and updating of bibliographical references concerning a specific theme. It helps to manage information recovered at the different stages of the research process, along with other tools such as Zotero, in addition to fostering collaborative work by enhancing synergies inside the group while helping to build knowledge cooperatively.

Diigo is a metacognitive tool because it displays different ways to learn, think and build knowledge of each individual by showing the information each member selects, along with his or her preferences and strategies to organize and categorize it. In fact, by sharing with others this specific personal ability, its value is enhanced for the virtual community because it allows other members to opt for more efficient itineraries, maximizing their potentials as a whole.

In virtual learning contexts, Diigo is extremely useful to develop digital competences directly related to information search, management, analysis and categorization.

From a technical point of view, this tool is a step forward compared to other SBS because it has improved functionalities. Among these, it must be taken into consideration the possibility of highlighting contents and adding floating sticky notes on the web pages. Both types of annotation will be available for other users which favors collaborative work by making comments, corrections or explanations. As it has already been said in point 3, apart from the own functionalities of the tool, there are complements that make individual and collective work easier such as toolbars, favorite export and import from and towards other bookmarking toolbars or even a version for an iPhone application. Further development is to be expected so that a more visual version of the tool becomes a reality together with a comprehensive exploitation of its semantic capacities that allows for the suggesting of tags or the finding of users depending on their bookmarking habits.

## 6. References

- Alonso Arevalo, J. (2009). Gestores de referencias sociales | Universo Abierto. <http://www.universoabierto.com/2562/gestores-de-referencias-sociales/> (Accessed December 23, 2009).

- Bateman, S., Muller, M. J., & Freyne, J. (2009). Personalized retrieval in social bookmarking. In Proceedings of the ACM 2009 international Conference on Supporting Group Work, pp. 91-94, Sanibel Island, Florida, USA. ACM.
- Cattuto, C. (2006). Semiotic dynamics in online social communities. *The European Physical Journal C - Particles and Fields*, 46, pp. 33-37.
- Cernea, D. A., Del Moral, M. E., & Labra Gayo, J. E. 2008. SOAF: Semantic Indexing System Based on Collaborative Tagging. *Interdisciplinary Journal of E-Learning and Learning Objects* 4, pp. 137-150.
- Colás Bravo, P. (2003). Internet y aprendizaje en la sociedad del conocimiento. *Comunicar* 20; 31-35.
- Del Moral, M. E. & Cernea, D. A. (2006). “Wikis, Folksonomías y Webquests: trabajo colaborativo a través de Objetos de Aprendizaje”. In Proceedings of III Simposio Pluridisciplinar sobre Diseño, Evaluación y Descripción de Contenidos Educativos Reutilizables (SPDECE06) Oviedo, 2006.
- Diigo 2006. Diigo is about Social Annotation. <http://www.diigo.com/help/about> (Accessed December 31, 2009).
- Dye, J. (2006). Folksonomy: A game of high-tech(and high-stakes) tag. *EContent (Wilton, Conn.)* 29, 2006.
- Golder, S. A., & Huberman, B. A. 2006. Usage patterns of collaborative tagging systems. *Journal of Information Science* 32, 198-208. <http://jis.sagepub.com/cgi/content/abstract/32/2/198> (Accessed December 27, 2009).
- González Navarro, M. (2009). Los nuevos entornos educativos: desafíos cognitivos para una inteligencia colectiva. *Comunicar* 33; 141-148.
- Hammond, T.; Hannay, T.; Lund, B. & Scott, J. (2005). Social Bookmarking Tools (I). *D-Lib Magazine* 11. <http://dlib.org/dlib/april05/hammond/04hammond.html#3> (Accessed December 26, 2009).
- Heymann, P.; Koutrika, G. & Molina, H. G. (2008). Can social bookmarking improve web search? In WSDM '08: Proceedings of the international conference on Web search and web data mining, pp. 195-206, New York, NY, USA. ACM.
- Kolay, S. & Dasdan, A. (2009). The value of socially tagged urls for a search engine. In WWW '09: Proceedings of the 18th international conference on World wide web, pp. 1203-1204, New York, NY, USA. ACM

- Mathes, A. (2004). Folksonomies - Cooperative Classification and Communication Through Shared Metadata. Retrieved December 22, 2009, from <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>
- Millen, D. R.; Whittaker, S. & Yang, M. (2007). Social bookmarking and exploratory search. ESI, 5.
- Millen, D.; Feinberg, J. & Kerr, B. (2005). Social bookmarking in the enterprise. Queue, 3(9), pp. 28-35.
- Monge, S.; Ovelar, R. & Azpeitia, I. (2008). Repository 2.0: Social Dynamics to Support Community Building in Learning Object Repositories. Interdisciplinary Journal of E-Learning and Learning Objects 4, 2008, pp. 191-204.
- Moral Toranzo, F. (2009). Internet como marco de comunicación e interacción social. Comunicar 32; 231-237.
- Nations, D. (n.d.). Social Bookmarking – What is Social Bookmarking? <http://webtrends.about.com/od/socialbookmarking101/p/aboutsocialtags.htm> (Accessed December 21, 2009).
- Reichel, M. & others (2006). Embodied, Constructionist Learning: Social Tagging and Folksonomies in E-Learning Environments. mICTE 2006. Conference Proceedings
- Rosen, Y. & Rimor, R. (2009). Using a Collaborative Database to Enhance Students' Knowledge Construction. Interdisciplinary Journal of E-Learning and Learning Objects 5, 2009, 187-196.
- Singh, G.; Hawkins, L. & Whymark, G. (2007). An integral model of collaborative knowledge building. Interdisciplinary Journal of E-Learning and Learning Objects 3, 85-104. <http://ijello.org/Volume3/IJKLOv3p085-105Singh385.pdf>
- Smith, Gene. Atomiq: Folksonomy: social classification. Aug 3, 2004 [cited 7 April 2010]. Available from world wide web: [http://atomiq.org/archives/2004/08/folksonomy\\_social\\_classification.html](http://atomiq.org/archives/2004/08/folksonomy_social_classification.html)
- Social bookmarking - Wikipedia, the free encyclopedia. [cited 28 December 2009]. Available from world wide web: [http://en.wikipedia.org/wiki/Social\\_Bookmarking](http://en.wikipedia.org/wiki/Social_Bookmarking)
- Zubiaga, A.; Martínez, R. & Fresno, V. (2009) Getting the most out of social annotations for web page classification. In Proceedings of the 9th ACM Symposium on Document Engineering, pp. 74-83, Munich, Germany. AC

## **CAPÍTULO 5 -**

# **Estudio y análisis de los diferentes tipos de etiquetas que se pueden utilizar en los sistemas de etiquetado social**

## 5.1 Introducción

Este capítulo se corresponde con el artículo “*Uses of explicit and implicit tags in social bookmarking*” publicado en la revista *Journal of the American Society for Information Science and Technology*.

### 5.1.1 Resumen del artículo

Las etiquetas que se utilizan para describir los documentos marcados en distintos sistemas de marcado o etiquetado social pueden ser, en base a si aparecen en el contenido marcado o no, implícitas o explícitas. Las etiquetas implícitas son aquellas que se utilizan para marcar un recurso textual y que no aparecen dentro del recurso. Las explícitas son aquellas utilizadas para marcar un recurso y que además, aparecen dentro del mismo.

Este artículo realiza una descripción en profundidad del uso que hacen los usuarios de cuatro sistemas de marcado y etiquetado social de estos dos tipos de etiqueta: Diigo, Delicious, Connotea y Mister Wong. Destaca el resultado que indica que los usuarios hacen un uso similar tanto de las etiquetas implícitas como de las explícitas.

### 5.1.2 Datos de la publicación

El artículo ha sido publicado en la revista *Journal of the American Society for Information Science and Technology*, revista internacional centrada en la producción, descubrimiento, almacenamiento, representación, manipulación, diseminación, uso y evaluación de información, y en las técnicas y herramientas asociadas a estos procesos. La revista está indexada tanto en *Social Science Citation Index*, como en *Science Citation Index*, así como en *Scopus*. Se encuentra en distintas bases de datos como *Academic Search Premier*, *Francis*, *Business Source Elite*, *Information Science and Technology Abstracts*, *Library and Information Abstracts* y *Library Literature and Information Science*.

Esta revista tuvo en 2011 un índice de impacto JCR de 2.081, encontrándose la revista, según el ISI Journal Citation Reports en la posición 10/83 en la categoría “Ciencias de la Información y Biblioteconomía” y 21/135 en la sección “Informática y Sistemas de Información”. Ocupa el primer cuartil (Q1).

Los autores son Enrique Estellés-Arolas y Fernando González Ladrón-de-Guevara.

- Nombre de la revista: Journal of the American Society for Information Science and Technology
- Editorial: John Wiley & Sons, Inc.
- ISSN: 1532-2882
- Fecha: Febrero 2012
- Volumen: 63
- Nº: 2
- Páginas: 313–322

## 5.2 Artículo

### Study about the different use of explicit and implicit tags in social bookmarking

Enrique Estellés-Arolas

Department of Management, Technical University of Valencia, Valencia, Spain

Fernando González-Ladrón-de-Guevara

Department of Management, Technical University of Valencia, Valencia, Spain

#### Abstract

Although Web 2.0 contains many tools with different functionalities, they all share a common social nature. One tool in particular, social bookmarking systems, allows users to store and share links to different types of resources i.e. websites, videos, images, etc. In order to identify and classify these resources so that they can be retrieved and shared, fragments of text are used. These fragments of text, usually words, are called tags. If a tag is found on the inside of a resource text, it is referred to as an obvious or explicit tag. There are also non-obvious or implicit tags, which don't appear in the resource text. The purpose of this paper is to describe the present situation of the social bookmarking systems tool and then to also determine the principal features of and how to use explicit tags. It will be taken into special consideration which HTML tags with explicit tags are used more frequently.

#### Keywords

social bookmarking systems, tagging, explicit tags, resources, social tagging.

## 1. Introduction

Web 2.0 has enabled the proliferation of applications such as blogs, social networks, wikis, social bookmarking systems, etc. These allow users to communicate and share resources in a collaborative way in a professional field as well as in an academic or research sphere. These web applications have 3 common features: there are user profiles, it is possible to follow other users or add them as friends or contacts, and it is possible to add comments to the generated content (Mason & Rennie, 2008).

Another feature that most of these systems share is the possibility of labeling the contents through the use of keywords called tags. The content can be a blog entry (e.g.

technorati.com), a resource marked in a social bookmarking system (e.g. delicious.com), books (e.g. librarything.com), objects in a museum (e.g. www.steve.museum) user-generated videos (e.g. youtube.com), or images (e.g. flickr.com) (Bar-Ilan et al., 2010). Tags are very important in these types of systems because they make the search of these resources as well as their organizations and description easier (Oliveira et al, 2008) and they also enable the user to find similar resources (Millen et al., 2005). Social bookmarking systems are web applications that allow users to store and manage their markers or favorites not in the browser but instead in a central server, so that they can be consulted from different locations and shared by other users (Illig et al., 2009).

Regarding text resources (i.e. text found in a website or in a blog entry), two types of tags can be found: obvious, also called explicit, or non-obvious, also called implicit. Implicit tags are those that do not appear within the textual content of the resource. Explicit tags are those appearing at least once within the textual content visible for users. For example, they can appear within a web title, a paragraph, or a link of the website itself (Farooq et al., 2007, Liu et al., 2008).

Usually, more attention has been given to implicit tags than explicit tags (Farooq et al., 2007), but explicit tags can also be very useful. This paper shows the results regarding the use of explicit tags by analyzing data collected in four different social bookmarking systems: Delicious, Diigo, Connotea, and Mister Wong. It is important to point out that Delicious, which belongs to Yahoo!, is working at full capacity. In spite of the news that arose in December 2010 about the end of SBS, Yahoo! explained that it would not be closed, but instead sold to another company (Delicious' Blog, 2010), so it has been included as a valid source of data for this paper.

Throughout this paper some questions will be answered, such as – in general, do users use the same quantity of explicit tags and implicit tags?, what about on a resource level?, furthermore, on a resource level, are explicit tags stored in a series of specific resources, or are they distributed equally among them all?, is there a difference between the lengths of the two types of tags?, and are the terms most frequently used to tag resources implicit or explicit?

This paper is divided into four sections. The first section consists of a theoretical introduction about the tags and a detailed description of some features. In the second section, the methodology that has been implemented is described and in the third section, the analysis that



has been carried out. Then, the paper takes into consideration the results that have been obtained and answers the questions previously asked. Finally, conclusions are shown along with a series of suggestions about the applications and for future research.

## 2. Tags

### 2.1. Definition

Tags are generated and freely chosen by the user to form descriptive strings, which are assigned or associated with a resource (Millen, Yang, Whittaker, & Feinberg, 2007; Koutrica et al., 2008; Farooq, Zhang, & Carroll, 2009; Lipczak & Milios, 2010). Depending on the tag system design, these descriptive chains can be words, phrases, or a combination of symbols and alphanumeric characters (Yeung, Gibbins, & Shadbolt, 2009).

Tags can also be considered as metadata (Subramanya & Liu, 2008), i.e., data about data. The three types of metadata are administrative, structural, and descriptive (Taylor, 2003) and can be developed by dedicated professionals, authors, or general users (Mathes, 2004).

These tags are used in Collaborative Tagging Systems, enabling users to assign freely chosen tags to web resources (Yeung et al., 2009). When users assign tags to web resources, creating a collaborative classification system, it is called a folksonomy (Illig et al., 2009, Marinho et al., 2011). Coined by Thomas Vander Wal in 2004, the word “folksonomy” comes from the words “folk” and “taxonomy” (Smith, 2004). Folksonomies are considered a set of evolving categorization schemes or, as explained by Mathes (2004), a set of terms with which a group of users tagged content.

A folksonomy can be defined as a tuple  $F := (U, T, R, Y)$ , where  $U$ ,  $T$  and  $R$  are finite sets, whose elements are called users, tags, and resources, respectively.  $Y$  is a ternary relation between them, i.e.,  $Y \subseteq U \times T \times R$ . The elements  $y \in Y$  are called tag assignments (TAS). A post is a triple  $(U, T_{UR}, r)$  with  $u \in U$ ,  $r \in R$  and a nonempty set  $T_{UR} := \{t \in T | (u, t, r) \in Y\}$  (Schmitz, Hotho, Jäschke, & Stumme, 2006).

This article will focus only on the relationship between resources and tags used to mark them, particularly on explicit tags, which will be explained later.

### 2.2. Functions and motivation

According to Golder and Huberman (2005), there are seven basic nonexclusive functions that a tag can carry out: identify what or whom the resource deals with, identify what it is,

identify who owns it, refine categories, identify qualities or features, aid in self-reference (e.g., “myStuff”), and organize tasks (e.g., “toRead”). Körner, Benz, Hotho, Strohmaier, and Stumme (2010) and Millen et al. (2007) are more specific and they put these functions into only two groups, namely, categorize and describe. Users using tags to categorize are called categorizers and they use a more complex set of tags with the main purpose of creating taxonomies for group resources. This system enables users to use multiple tags so that a given resource can belong to more than one category. On the other hand, there are users that use tags with a descriptive purpose. These are called descriptors and they consider the tag as a way of accurately and precisely describing saved resources. The main goal of these users is to use the tagging for a subsequent search and retrieval. The difference between these two functions is minimal in practice and users are capable of tagging with dual intent: categorizing and describing. Other authors, like Ding et al. (2010), argue that the principal functions of the tags are to navigate, browse, and retrieve resources. They highlight the social nature of this type of application by stating that taggers enjoy being embedded in a social environment, being watched by others, and receiving feedback on their actions.

As a consequence of the combination of all the abovementioned functions with the social nature of the applications where tagging is used, secondary functions arise (Koutrika, Effendi, Gyöngyi, Heymann, & Molina, 2008; Jäschke, Marinho, Hotho, Schmidt-Thieme, & Stumme, 2008; Oliveira et al., 2008; Fu, Kannampallil, Kang, & He, 2010; Ding et al., 2010):

- Facilitate sharing between users.
- Facilitate collaborative indexing of information.
- Guide users to interesting and new resources.
- Help users build communities that share their expertise and resources.
- Navigate.
- Browse serendipitously.
- Receive feedback on their actions.

All these functions can be completed through the technique known as *pivot browsing* (Millen et al., 2007; Bateman, Muller, & Freyne, 2009). This technique enables the user to reorient the navigation view by clicking on different elements of the user interface, e.g., the name of the users or the tags. By clicking on a user’s name, all the resources stored by the user will be

displayed. By clicking on a tag, resources marked with the same tag will be shown (Millen et al., 2005).

In regards to the motivation that compels users to mark resources through this technique, Marlow, Naaman, Boyd, and Davis (2008) highlight the following:

- Future retrieval: Users mark resources to remember pending tasks (e.g., “toRead”) or define clusters of objects that will be used later, for example, marking web resources to write a research paper with the tag “research\_paper\_1.”
- Contribution and sharing: Create clusters of resources for oneself and other users, whether or not they are known. An example of this would be marking photos of a group trip with the tag “trip\_Rome\_2010” so that all the members of that group can see those photos.
- Attract attention: By using commonly used tags, as those shown by clouds of tags, the rest of users can be attracted to the resources.
- Play and competition: tagging according to specific rules established by games as the ESP Game.
- Self-presentation: Mark a resource in a particular way, for example, tagging a concert with the tag “SeenInLive.”
- Opinion, expression: Express the opinion about the marked resource by pointing out a subjective category, for example, tagging a link to a blog as “elitist.”

### **2.3. Types of tags and their meaning**

Depending on their meaning, tags elaborated by users can be put into three categories which determine the tag function. These categories are as follows: content tags, which describe the content; attitude tags, which enable opinion expression; or self-reference tags, which are self-reminders (Melenhorst & Van Setten, 2007).

Regardless of the type of tag that is being used, marking resources that are interesting for whatever reason reveals the users’ interests in a specific and explicit way (Li et al., 2008). In other words, the tags posted by a user will be relevant not only to the content of the bookmark but might also be specific to that user (Zhang, Zhang, & Tang, 2009). Essentially, a single resource can be marked by different users with different tags, which will represent a varied set of topics of interest.

## 2.4. Content

The content tag, as it has already been said, consists of a term or a set of terms freely chosen by the user. In this regard, two types of tags can be found (Farooq et al. 2007, Liu et al., 2008):

- Explicit or obvious tags, which can be found within the text content of the marked resource. These types of tags, as this article tries to show, are used very frequently by users.
- Implicit or nonobvious tags, which cannot be found within the text content. According to Farooq et al. (2007), these types of tags have a higher intellectual value because they provide insights into the content of the article.

Various reasons may impel users to use explicit tags. According to Lipczak and Millos (2010), users want to minimize efforts and tend to use tags that are easily available. Farooq et al. (2007) point out that the explicit tag can be just a good descriptor in spite of the fact that it does not add any extra intellectual value.

On the other hand, there are parts of web resources that are frequently used when explicit tags are chosen. Recent studies (Eisterlehner et al., 2009) show that there is a relatively high overlap between the tags marked by users and the words extracted from the title of the resource. The high overlap reveals a combination of an attempt to minimize efforts (because the user can see the title during the tagging process) with the dense resource description that it involves. On the other hand, Liu et al. (2008) show that tags and visible, clickable text in hyperlink (anchor text) tend to overlap.

The results of this article show that there are other parts in web documents that also have a great impact on the selection of explicit tags, thereby verifying the results of Eisterlehner et al. (2009) and Yimming et al. (2008), which show the high percentage of explicit tags found in the title and the anchor text.

Regarding implicit tags, it is important to point out that they do not always have a higher intellectual value as Farooq et al. (2007) suggest. As has already been stated, tags can be used for different functions, including self-reference and the organizing task. In such cases, the information may be valuable for the users using them, but not necessarily for the rest of users. For example, tagging a resource referring to a book as “owned” means that that title can be found in the user’s personal library, which does not add any extra value and it is, in fact, a

handicap for those users looking for books that cannot be found in their libraries (Fu et al., 2010). Other examples would be tags like “must,” “toRead,” or “pendent.”

## **2.5. Disadvantages of tagging**

As it has already been stated, one of the advantages to tagging is the possibility to create tags by combining all types of characters and signs, thereby forming a kind of open vocabulary. Other terms can also be added, which describe specific content even though it is only personally relevant for an individual user. However, that advantage involves two basic problems with regard to social tagging, namely, informational redundancy (Robu, Halpin, & Shepherd, 2009) and the loss of general significance.

The informational redundancy problem refers to the creation of many different tags that describe the same resource, so that different users use synonyms, homonyms, and polysemes (Furnas, Landauer, Gomez, & Dumais, 1987). According to Fu et al. (2010), the increasing number of vocabularies will cause the connections between tags and documents to become less direct and more confusing, making Information retrieval more difficult.

On the other hand, using specific tags excessively will imply a certain level of ambiguity (i.e., “!fic”, “#cm10conf” o “#mn1010”). This is because these can be incomprehensible for other users, thereby limiting the effectiveness of collaborative tagging systems in document description and retrieval (Yeung et al., 2009).

## **3. Methods**

This article is based on the data obtained from the analysis of four SBSs. To select them (see Table 5.1), some of the best-known SBSs were analyzed. Those that did not meet the following standards were dismissed: the marked resources must be a website with text and not other types of files or documents (pdf, doc, etc.), they must be marked with tags, and they must enable access to the web resource. Thus, those resources requiring a subscription or a registration were rejected, as well as those not using tags or those using fragments of texts like comments or descriptions as resource metadata. Furthermore, backFlip was also rejected because it was out of order, as was Gnolia because it offered very few links due to its closure on November 30, 2010.

After this analysis, the four resources that better fit our needs were selected: Delicious, Diigo, Mister Wong, and Connotea (see Table 5.2). The four of them use tags to mark resources, they are free and enable direct access to the marked resource, and do not require registration

to be able to consult available resources. The first three (Delicious, Diigo, and Mister Wong) are general SBSs, which means that they are not specialized in specific types of content. As for Connotea, it defines itself as a “free online reference management for all researches, clinicians and scientists,” which is why it deals with scientific content.

**Table 5.1. List of rejected SBSs (source: by the authors)**

	No tags	Comments/ description	No doc	Pay	Register
Bibsonomy			X		
Bookmarkstyle	X	X			
Buddymarks	X				
Buzz	X	X			
CiteUlike			X		
Digg	X	X			
euri.com	X				
Identi.ca	X				
IndianPad	X				
Knowledge plaza				X	
LinkWad.com					X
MyLinkVault	X	X			
Propeller	X				
Reddit	X				
StumbleUpon	X				
Tweetmeme	X	X	X		

Concerning the feature of suggesting tags to the users that bookmark resources, Connotea does not suggest any, whereas Delicious and Mister Wong suggest tags previously utilized by other users to bookmark the same resource. In addition, Delicious and Diigo also suggest the last tags employed by the user who bookmarks the resource. Finally, Diigo also suggests tags extracted from the content of the resource. Except for Diigo, the nature of these tags, whether they are implicit or explicit, is not taken into account when the different SBSs suggest tags.

To analyze the different SBSs, four crawlers written in Java were created purposefully for each one. These crawlers were run through those sections where the most popular and latest added resources are shown (i.e., those marked for most of the users). In each of these sections, shown in the second column in Table 5.2, the crawlers obtained the different resources available by storing the URL of the resource and the related tags.

Each of the stored resources in each SBS was examined to check whether the resource was active, a website, and another type of web resource (image, text document, spreadsheet, etc.), or had text content (it can be a website made with flash, in which case the language used to write the site is also relevant).

**Table 5.2. A summary chart of those SBSs that were accepted.**

SBS	Section
Delicious	HotList ( <a href="http://www.delicious.com/?view=hotlist">http://www.delicious.com/?view=hotlist</a> )
Diigo	Hot Bookmarks ( <a href="http://www.diigo.com/buzz/hot">http://www.diigo.com/buzz/hot</a> )
Mister Wong	Fresh Bookmarks ( <a href="http://www.mister-wong.com/?more=fresh">http://www.mister-wong.com/?more=fresh</a> )
Connotea	Popular links ( <a href="http://www.connotea.org/popular?">http://www.connotea.org/popular?</a> )

To identify the language of the resource, NGramJ was applied (<http://ngramj.sourceforge.net/index.html>). This is a Java-based library containing two types of NGram based applications, where ngrams are classical instruments in natural language processing (NLP) applications. Its main function is language guessing or language recognition, providing a language identifier (es-spanish, en-english, de-denmark, etc.) starting from a piece of text.

Finally, each of the resources was checked to determine if it was marked with any tag. In this case, apart from storing tags, the text of the web resource was extracted and the quantity of explicit and implicit tags was calculated. To consider a tag as explicit, there must be at least one exact overlap within the text of the resource. In the case where explicit tags did appear, an accurate analysis was carried out to determine in which HTML tags the explicit tags were found and how frequently they occurred.

To manage web resources, Jericho HTML Parser was applied. This is a Java library, which allows analysis and manipulation of parts of an HTML document, including server-side tags, while reproducing verbatim any unrecognized or invalid HTML (<http://jericho.htmlparser.net/docs/index.html>). However, this library did not avoid those

problems arising as a consequence of working with Cyrillic-like alphabets. In some of these cases, characters were written as HTML entities. For example, the character “□” is represented in the source code in its hexadecimal HTML representation: “&#x41F.” On such occasions, the Commons Lang library (<http://commons.apache.org/lang/>) was used, in particular, the `StringEscapeUtils` function, which enables the extraction of characters as such, thereby turning HTML entities into characters.

All this information has been stored in a MySQL database that comprises three tables. The first one, `webs`, deals with the storage of URLs and some of their features (e.g., the language, the availability of tags, whether it is an HTML file, whether it is working properly, whether it has content, from which SBS it was extracted). The second table, `tags`, deals with the storage of the different tags that have been collected, showing whether they are explicit, in which case it shows how many times they appear in the resource text content. The third and final table, `html_tags`, stores the HTML tags where explicit tags have appeared as well as the number of explicit tags found within the HTML tags in each corresponding resource.

Links were collected on working days, from September 1, 2010, to October 15, 2010, each crawler running individually every day. 151,699 URLs were collected and analyzed through the statistics program SPSS, starting with the data stored in a MySQL 5.1.37 database.

## 4. Results

The results obtained are described from a double point of view. First, all of the related data are analyzed to achieve a general view of the SBSs. Secondly, the collected data are filtered to analyze the features and structure of explicit tags properly.

### 4.1. Data about SBS

The collected data can be divided into two groups, webs or resources and tags, as seen in Table 5.3.

It is important to point out that Connotea as well as Mister Wong do not have nonmarked resources because, in both cases, the user is required to introduce at least one tag to be able to mark a resource. Also, Connotea has fewer resources because of time delays in the process of connection to different pages of the website.



**Table 5.3. Summary of the webs collected and number of tags related to each one. By the authors.**

SBS	Total	Resources with no tags	Total amount of Tags
Connotea	21,060 (13,89%)	0	74,378
Delicious	41,347 (27,26%)	3,949	119,726
Diigo	46,171 (30,43 %)	10,011	153,241
Mister Wong	43,121 (28,42%)	0	225,874
TOTAL	151,699	13,960	573,219

## 4.2. Languages of the resources

Regarding the language used in the different resources collected, English is most commonly used (77.9%), followed by Russian, Spanish, German, and French (see Table 4). These languages represent 88.35% of the resources, even though 28 different languages were analyzed altogether. In this respect, it is important to note that Mister Wong's resources were ignored in the language analyses, because it has web portals for different languages including Spanish, French, German, Romanian and Chinese. The resources in these languages are available from those portals. Also, it should be made apparent that Russian is the second most commonly used language because 25% of the resources marked in Connotea were written in this language. In the rest of the SBSs, the resources in Russian do not exceed 2.19%.

Finally in 7.74% of the resources, the language has not been identified properly due to the lack of text in the resource itself or the impossibility of entering the page because it was not possible to connect to the server or because a 404 error message bounced back showing that the requested page was not available.

## 4.3. Number of tags per resource

From among all the chosen resources, 90.79% (137,739) were marked with tags. The distribution of these tags is described in Table 5 below, where it can be observed that 94% of the URLs are marked with 10 or fewer tags.

Generally speaking, SBS resources are marked with a mean of 4.16 tags, with a mode of 1, a median of 4, and a standard deviation of 3.34. Only 0.73% of the resources is marked by more than 14 tags.

**Table 5.4. Use of languages in the analyzed web pages.**

Language	Webs	%	Cumulative %
English	118,180	77.90	77.90
Russian	6,065	3.99	81.89
Spanish	4,260	2.80	84.70
German	2,981	1.96	86.67
French	2,559	1.68	88.35
Romanian	1,092	0.71	89.07
Italian	1,021	0.67	89.75
Others (pt, ua, sv, hu...)	15,547	10.24	100

Depending on the different SBSs, the number of tags used per resource changes, but not significantly (see Table 5.6 and Figure 5.1): the mode changes in Diigo and Mister Wong and the average frequency of use of each tag per resource is  $4 \geq 1$ . In contrast, Connotea has a significantly greater number of maximum tags used compared with Diigo, Delicious, and Mister Wong, with one resource marked with 157 tags. This SBS has 0.75% of its resources (32) marked with more than 39 tags, which is the highest value in Diigo.

It must be pointed out that in Figure 1, the dispersion of SBS Diigo is the lowest and also the behavior of users of Delicious and Connotea are rather similar, even though, unlike Delicious, Connotea does have extreme values.

#### 4.4. Other features

Some specific features of the collected tags are going to be described below (how long they are, how many unique tags exist, and which are mostly used). Then those features can be compared with the same features in explicit and implicit tags, which will permit easier differentiation.

In the first place, the total number of tags (573,219) has an average length of 8.53 characters with a standard deviation of 5.73. The mode value is four characters, which means that most of the tags are that long. On the other hand, finding tags with many characters is not strange. This is because users do not always introduce individual terms, but instead introduce a set of linked words or words separated by different punctuation marks like “-”, “,”, or “#.” A few examples include “registrationsingapore,” “link-building-service,” or “ufc-120-live-stream-

fee-online.” In other cases, bookmarking systems allow the addition of tags that comprise various terms, such as “bisping vs akiyama live stream” or “selling antique rings.”

**Table 5.5. Quantity of webs according to the tags with which they have been marked.**

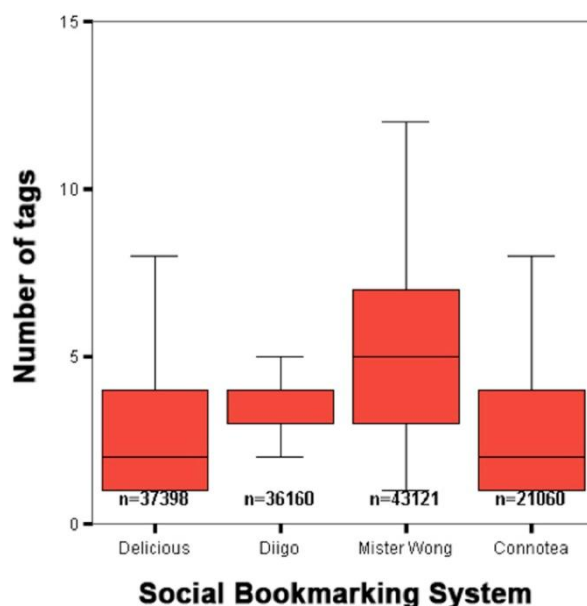
No. tags	No. webs	%	% accumulated
0	13,966	9.18	9.18
1	29,326	19.33	28.51
2	17,595	11.59	40.11
3	21,169	13.95	54.07
4	25,505	16.81	70.88
5	11,420	7.47	78.36
6	7,923	5.12	83.48
7	6,188	2.16	85.65
8	4,566	3.48	89.13
9	3,982	2.92	92.06
10	2,777	2.08	94.14
>10	8,875	5.85	100

**Table 5.6. Data about the use of tags per web according to each SBS.**

SBS	Total amount of tags	Mean	Standard deviation	Max
ALL OF THEM	573,219	4.16	3.34	157
Delicious	119,726	3.20	2.59	20
Diigo	153,241	4.24	2.89	39
Mister Wong	225,874	5.24	3.23	12
Connotea	74,378	3.53	4.57	157

A total of 110,617 unique tags are obtained from these tags, from which 68% are used just once, 11.9% twice, and 5.3% three times. On the whole, 90% of tags are used five times or less. On the other hand, the most commonly used tags reveal which topics are typically discussed in the SBSs and allows the analysis of terms frequently used as tags. Table 5.7

shows the 10 explicit and implicit tags most commonly used and that most of them deal with topics related to the Internet (e.g., blog, technology, computers, online, software).



**Figure 5.1.** A box and whisker diagram showing the number of tags per marked resources. Outliers and extreme values are hidden in order to appreciate the graphic.

**Table 5.7.** Most frequently used tags.

Imp tags.	% marked resources	Expl tags.	% marked resources
articles	7.65 % (11,604)	blog	5.46 % (8,276)
computers	7.62 % (11,567)	online	1.80 % (2,728)
technology	7.44 % (11,292)	video	1.60 % (2,433)
blog	3.98 % (6,040)	technology	1.10 % (1,662)
clip	2.07 % (3,135)	free	1.07 % (1,621)
article	1.97 % (2,986)	design	0.97 % (1,476)
video	1.47 % (2,223)	watch	0.82 % (1,246)
uploaded	0.57 % (861)	business	0.81 % (1,234)
webdesign	0.53 % (810)	to	0.74 % (1,127)
design	0.53 % (801)	web	0.74 % (1,120)

#### 4.5. Analysis of implicit and explicit tags

To carry out this analysis, a subsample was taken from the original, shown in Table 5.8. The original sample comprises 151,699 URLs stored in four different SBSs: Delicious, Diigo,

Mister Wong, and Connotea. From these, 16.35% (24,808) have been rejected for the analysis because they were not working (they bounced back an error 404 message saying that the page was not available), they were not marked with any tag, they were not html files, it was impossible to extract text, or any combination of these four events. Therefore, from among the SBSs below, 126,891 URLs were taken to be analyzed.

**Table 5.8. Itemization of the collected urls.**

SBS	TOTAL	Accepted	Rejected	No tags	No Html	Out of order	No content
Connotea	21066 (13,89%)	18460	2600	0	1746	2514	1675
Delicious	41341 (27,26%)	36225	5116	3939	1262	971	929
Diigo	46171 (30,43 %)	31790	14381	9992	3516	4769	3157
Mister Wong	43121 (28,42%)	40416	2705	0	1984	2261	1540
TOTAL	151699	126891	24802	13931	8508	10515	7301

Regarding the number of resources per SBS, it depends on the response time of the different SBSs. Because crawlers ran at the same time through each SBS, if the response time proved to be short, more resources could be processed.

Altogether, a total number of 524,930 tags associated to those URLs were collected, from which 45.10% (236,782) are implicit tags. As it has already been stated that for a tag to be considered explicit, there must be at least one overlap within the text of the resource. Through the crawlers, this condition was verified.

The selection of the explicit tags allowed us to consider a total of 91,652 resources, which were marked with at least one of these tags. These resources are going to be used as a basis for the analysis of this type of tag.

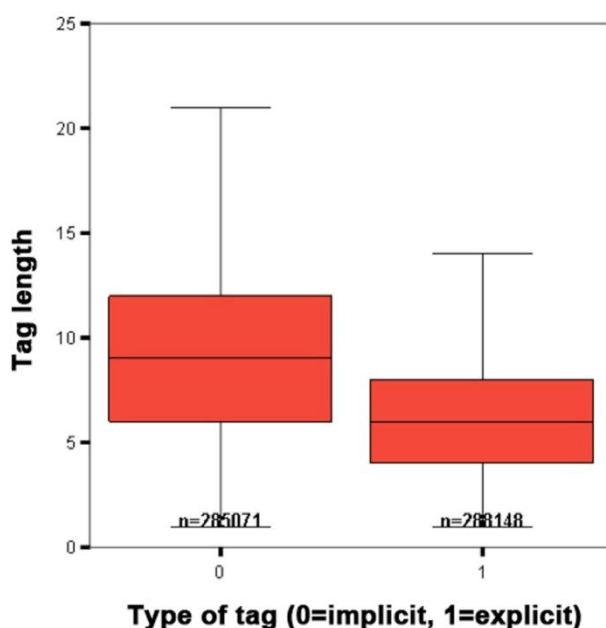
The percentage of explicit and implicit tags that arise in the analysis of the resource is shown in Table 5.9. Diigo is the SBS where there are fewer explicit tags (41%), compared with Mister Wong which has 67% of the explicit tags.

**Table 5.9. Percentages of implicit and explicit tags.**

SBS	TOTAL	Explicit	Implicit
Connotea	62.034	31.672 (51%)	30.362 (49%)
Delicious	116.256	57.118 (49%)	59.138 (51%)
Diigo	134.961	55.421 (41%)	79.540 (59%)
Mister Wong	211.679	143.937 (67%)	67.742 (33%)
TOTAL	587.019	288.148 (100%)	236.782 (100%)

#### 4.6. Length

The average length of the tags was previously calculated. In general, there are 8.53 characters per tag. The length obtained according to the type of tag is different from the general mean (Figure 5.2). In other words, while implicit tags have a mean of 10.23 characters and a mode of 8, explicit tags have a mean of 6.84 characters and a mode of 4.



**Figure 5.2. Distribution of explicit tags and implicit tags.**

#### 4.7. Explicit and implicit tags per resource

Revising general data regardless of the type of tag, it can be seen that resources were marked with a mean of four tags (4.16) and a mode of one. When distinguishing by type of tag, a mean of 2.27 for explicit tags and 2.24 for implicit tags is obtained per resource, but the real distribution is different from these results.

Focusing on this distribution, it can be found that 31.22% (39,618) of the resources of the SBSs are marked only with explicit tags, 27.77% (35,239) are marked only with implicit tags, and the remaining 41.10% (52,034) are marked with implicit and explicit tags. Within this 41.10% of the resources that have both types of tags, explicit tags represent 49.10% and implicit tags 50.90%.

There is a mean of 5.6 tags per resource, with half of the mean being explicit tags and the other half being implicit tags.

#### 4.8. Number of times that tags are used to mark different resources

It was proved that, whether implicit or explicit, most tags are used only once. Thus, in explicit tags (Table 5.10), 85% of them are used five times or less and the same can be said of implicit tags (Table 5.11).

**Table 5.10. Summary of the quantity of times that explicit tags are used.**

Use	No. tags	%	% accumulated
1	30.940	33,8%	33,8%
2	18.088	19,7%	53,5%
3	13.244	14,5%	67,9%
4	9.163	10%	77,9%
5	6.318	6,9%	84,8%
>5	13.899	15,1%	100%

**Table 5.11. Summary of the quantity of times implicit tags are used.**

Use	No. tags	%	% accumulated
1	35.482	36,2%	36,2%
2	17.591	17,9%	54,1%
3	18.346	18,7%	72,8%
4	10.876	11,1%	83,9%
5	5.049	5,1%	89%
>5	10.777	11%	100%

#### 4.9. Explicit and implicit tags mostly used

Tables 5.12 and 5.13, below, show which of the 110,617 unique tags available are most frequently used, making a distinction between explicit and implicit tags.

**Table 5.12. Explicit tags frequently used.**

Tag	Marked resources
blog	8276
online	2728
video	2433
technology	1662
free	1621
design	1476
watch	1246
business	1234
to	1127
web	1120

By observing these data in Tables 5.12 and 5.13, it can be inferred that, in both cases, terms refer to technology and Internet issues (blogs, technology, etc.). From them all, three must be highlighted because they appear in both lists, namely, blog, video (which are used in a similar way), and technology (which is more frequently used in implicit tags).

Within the most commonly used tags, the most utilized tag within implicit tags is used more frequently than the most utilized tag within explicit tags. For example, the implicit tag most commonly used is “articles,” which is used 11,604 times, while in explicit tags, the most commonly used tag is “blog,” which is used only 8,276 times. Even so, at the end of the list, values tend to become equal, for example, the 10th explicit tag (“web”) is used 1,120 times, while the 10th implicit tag (“uploaded”) is used 861 times. This means that as far as implicit tags are concerned, there are some of them that are frequently used and others that are less frequently used, while the use of explicit tags is more consistent.



**Table 5.13. Implicit tags frequently used.**

Tag	Marked resources
articles	11604
computers	11567
technology	11292
blog	6040
clip	3135
article	2986
–	2280
=	2252
video	2223
uploaded	861

#### **4.10. Frequency of appearance of explicit tags within the text of a marked resource**

About explicit tags, it is also important to know how many times these tags appear in the resource. These data are provided in Table 14.

Explicit tags normally appear only once (12.4%) or twice (11.7%) in the text. The frequency of explicit tag appearance in the text decreases gradually. It is important to note that while 24.1% represent the tags appearing once or twice, the quantity of tags appearing more than 15 times is 26.1% of the total.

#### **4.11. Relationship between the frequency of appearance and the length of explicit tags**

According to Lipczak and Millos (2010), users want to minimize their efforts and tend to use more readily available tags. Therefore, it could be stated that in the decision-making process, the length of the potential tags and their frequency of appearance are taken into account. A relationship between the frequency of appearance of explicit tags and their length exists, whereby the shorter the tag length and the higher its frequency of appearance, the easier it will be for the user to choose it as a tag.

**Table 5.14. Frequency of appearance of the different tags in the corresponding text.**

Freq.appear.	No. tags	%	% acc.
1	35,666	12.4	12.4
2	33,628	11.7	24
3	21,653	7.5	31.6
4	18,981	6.6	38.1
5	16,260	5.6	43.8
6	14,028	4.9	48.7
7	12,129	4.2	52.9
8	10,792	3.7	56.6
9	9,617	3.3	60
10	8,895	3.1	63
11	7,670	2.7	65.7
12	6,995	2.4	68.1
13	6,021	2.1	70.2
14	5,499	1.9	72.1
15	4,987	1.7	73.9
>15	75,327	100	100

To investigate if there is a statistically significant association between these two variables (tag length and frequency of appearance), a correlation has been computed. A significance index (Pearson's correlation coefficient) of  $-0.042$  with an alpha value of 0.01 was obtained. The direction of the correlation is irrelevant because, although it's negative, its value is almost 0. This result means that there is no relationship, so these features are not considered relevant in the decision-making process when a tag is to be chosen.

#### **4.12.HTML tags where more explicit tags appear**

Explicit tags appear most often within the HTML tags link and title, as other studies show (Eisterberg et al., 2009; Yimming et al., 2008). Analysis showed that after the HTML tags link and title, p, div, and span are the next HTML tags, where explicit tags are most frequently found. P tag is used to include text in paragraphs, div tag enables the creation of

layers to put inside whatever is wanted (e.g., images, text), and span tag makes the introduction of text fragments possible.

A total of 208 HTML tags containing explicit tags have been identified. Among them, there are obsolete tags (e.g., center, I, font) and HTML tags that do not meet the W3C standard (e.g., figcaption, title1, article\_body). Table 15 shows a summary of the HTML tags containing the 90.21% of the sample.

**Table 5.15. HTML tags frequently used.**

<b>4</b>	<b>Tags it contained</b>	<b>% of the total</b>
a	222,560	20.30
title	159,813	14.58
p	152,361	13.90
div	120,565	10.99
span	57,696	5.26
h1	51,379	4.68
strong	47,474	4.33
h2	43,321	3.95
li	31,864	2.90
img	30,285	2.76
td	26,975	2.46
b	22,925	2.09
h3	20,590	1.87

## **5. Discussion and Conclusion**

From the results, it can be inferred that explicit tags (54.9%) are used just as frequently as implicit tags (45.1%). This suggests that the tags obtained by users from the resource are enough for them to mark it, describe it, or classify it. Or at least, those tags are as useful as the tags not obtained from inside the resource. Explicit tags are shorter (a mean of seven characters) than implicit tags (a mean of 10 characters) and appear in the text 1–15 times in 74% of the cases.

According to these data, the relationship between the frequency of appearance of explicit tags and their length was studied. Because users want to minimize their efforts and tend to use more readily available tags (Lipczak & Millos, 2010), it could be stated that in the decision-making process, the length of the potential tags and their frequency of appearance are taken into account. The obtained results support that these features are not considered relevant in the decision-making process when a tag is to be chosen.

Regarding commonly used tags, it can be observed how implicit tags are used more frequently than explicit tags, especially in global terms such as technology, articles, computers, or clip, which enable a classification of a resource in a general way.

With regard to HTML tags where explicit tags appear, even though “title” and “a” labels have more explicit tags (34.8%), the most important tags are not HTML that somehow highlight the text, but instead are content-tags such as “p,” “div,” and “span,” which represent 30.15% of the remaining HTML tags. This means that when choosing explicit tags, users do not take into account the physical size of the text (such as headlines “H” or those texts highlighted as “strong”) as a reference, but rather they freely choose among the text available. These results can be very useful in tag suggestion systems based on resource content: using only the content inside the most commonly used HTML tags where explicit tags appear can support an improvement, reducing workload and execution time because less content has to be analyzed.

As for the state of SBSs, it must be pointed out that the 9.2% are nonmarked resources and 7% of the resources are offline. Because, in these systems, the pivot browsing is usually performed through the tags, when these are not available in a resource, that resource will rarely be visited because of its little to no visibility. On the other hand, the percentage of offline resources shows that these types of systems need to apply mechanisms that are able to keep them updated. In this case, it is not about removing links to the resources, because they belong to the users, but instead about warning them that they own a link repository containing links to unavailable resources, which are useless.

Several SBSs have been analyzed: With regard to the percentage of use of explicit and implicit tags, Connotea and Delicious have a percentage of use close to 50%, Diigo uses implicit tags a little more (59%), and Mister Wong uses explicit tags even more frequently (67%). Generally speaking, users do not use explicit tags more frequently than implicit tags.

However, several limitations exist because of the selection of the sample and the content analyzed. Regarding the sample, more web resources from additional SBSs could be used to permit generalization of the results. In relation with the content, this study is limited to the text available in the HTML content of web resources, which excludes other resources with content like word processor or PDF documents.

Following the line of investigation of the actual article, it would be interesting to compare the use of implicit and explicit tags in general purpose and specialized SBSs. Finding out the distribution of explicit tags among users would also be of interest to check whether explicit tags are common practice or, on the contrary, if only certain types of users utilize them.

In conclusion, although the use of explicit tags has been generally less valued than the use of implicit tags because their lack of additional intellectual power (Farooq et al., 2007), the results of this study support the idea that explicit tags are as practical and are used as frequently as implicit tags. Therefore, the use of explicit tags is a valid and an important tool for tagging web resources.

## 6. References

- Bateman, S., Muller, M. J., & Freyne, J. (2009). Personalized retrieval in social bookmarking. Proceedings of the ACM 2009 International Conference on Supporting Group Work, pp. 91-94, Sanibel Island, Florida, USA. ACM.
- Bar-Ilan, J., Zhitomirsky-Geffet, M., Miller, Y. and Shoham, S. (2010), The effects of background information and social interaction on image tagging. *Journal of the American Society for Information Science and Technology*, 61: 940–951.
- Delicious' Blog. What's next for Delicious. Available on line: <http://blog.delicious.com/blog/2010/12/whats-next-for-delicious.html>. Retrieved on 16-3-2011.
- Ding, Y., Jacob, E. K., Zhang, Z., Foo, S., Yan, E., George, N. L. and Guo, L. (2009), Perspectives on social tagging. *Journal of the American Society for Information Science and Technology*, 60: 2388–2401. doi: 10.1002/asi.21190
- F. Eisterlehner, A. Hotho, and R. Jäschke, editors. ECML PKDD Discovery Challenge 2009 (DC09), volume 497 of CEUR-WS.org, Sept. 2009.

- Farooq, U., Kannampallil, T. G., Song, Y., Ganoë, C. H., Carroll, J. M. and Giles, L. (2007) Evaluating tagging behavior in social bookmarking systems: metrics and design heuristics. In Proc. GROUP 2007, ACM, 351-360. 6.
- Farooq, U., Zhang, S., M. Carroll, J. (2009) Sensemaking of scholarly literature through tagging. CHI 2009 Sensemaking Workshop, April 4–9, 2009, Boston, MA, USA.
- Fu, W. T., Kannampallil, T., Kang, R., and He, J. (2010). Semantic imitation in social tagging. ACM Trans. Comput.-Hum. Interact., Vol. 17, No.3., pp. 1-37.
- Furnas, G. W., Landauer, T. K., Gomez, L. M., and Dumais, S. T. (1987) The vocabulary problem in human-system communication. Commun. ACM 30, 11.
- Golder, S. A. and Huberman, B. A. (2005) The Structure of Collaborative Tagging Systems. HP Labs technical report, 2005. Available from <http://www.hpl.hp.com/research/idl/papers/tags>
- Illig, J., Hotho, A., Jäschke, R., and Stumme, G. (2009). A Comparison of content-based Tag Recommendations in Folksonomy Systems. In Postproceedings of the International Conference on Knowledge Processing in Practice (KPP 2007).
- Jäschke, R., Marinho, L., Hotho, A., Schmidt-Thieme, L., and Stumme, G. (2007). Tag recommendations in folksonomies. In Knowledge Discovery in Databases: PKDD 2007, pages 506-514. Springer-Verlag, Berlin, Heidelberg.
- Jäschke, R., Marinho, L., Hotho, A., Schmidt-Thieme, L., and Stumme, G. (2008). Tag recommendations in social bookmarking systems. AI Communications, 21(4):231-247.
- Körner, C., Benz, D., Hotho, A., Strohmaier, M., and Stumme, G. (2010). Stop thinking, start tagging: tag semantics emerge from collaborative verbosity. In WWW '10: Proceedings of the 19th international conference on World wide web, pages 521-530, New York, NY, USA. ACM.
- Koutrika, G., Effendi, F.A., Gyöngyi, Z., Heymann, P. & Molina H.G. (2008) Combating spam in tagging systems: An evaluation. ACM Trans. Web, Vol. 2, No. 4., pp. 1-34.
- Lipczak, M. and Milios, E. (2010). The impact of resource title on tags in collaborative tagging systems. In HT '10: Proceedings of the 21st ACM conference on Hypertext and hypermedia, pages 179-188, New York, NY, USA. ACM.
- Liu, Y., Kumar, R., & Lim, K. (2008). Taggers versus Linkers: Comparing Tags and Anchor Text of Web Pages. UC Berkeley: School of Information. Report 2008-020. Retrieved from: <http://escholarship.org/uc/item/8b40q59k>

- Marinho, L. B., Nanopoulos A., Schmidt-Thieme, L., Jäschke R., Hotho, A. Stumme, G. & Symeonidis, P. (2011) Social Tagging Recommender Systems . Recommender Systems Handbook. editor(s) Francesco Ricci and Lior Rokach and Bracha Shapira and Paul B. Kantor. 615-644, Springer.
- Marlow, C., Naaman, M., Boyd, D., and Davis, M. (2006). Ht06, tagging paper, taxonomy, flickr, academic article, toread. In HYPERTEXT '06: Proceedings of the seventeenth conference on Hypertext and hypermedia, pages 31-40, New York, NY, USA. ACM.
- Mason, R. and Rennie, F. (2008). E-Learning and Social Networking Handbook: Resources for Higher Education. Routledge, NY.
- Mathes, A. (2004) Folksonomies Cooperative Classification and Communication Through Shared Metadata. Retrieved from: <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>
- Melenhorst, M., Van Setten, M. (2007). Usefulness of tags in providing access to large information systems. In: Proceedings of the IEEE Professional Communication Conference.
- Millen, D., Feinberg, J. and Kerr, B. (2005) Social Bookmarking in the enterprise. Queue, 3(9):28-35
- Millen, D., Yang, M., Whittaker, S., and Feinberg, J. (2007). Social bookmarking and exploratory search. ECSCW 2007 (2007), pp. 21-40.
- Oliveira, B., Calado, P., Pinto, H.S. (2008) Automatic Tag Suggestion Based on Resource Contents. EKAW '08: Proceedings of the 16th international conference on Knowledge Engineering, 255-264. Springer-Verlag, Berlin / Heidelberg
- Robu, V., Halpin, H., and Shepherd, H. (2009). Emergence of consensus and shared vocabularies in collaborative tagging systems. ACM Trans. Web, Vol. 3, No. 4., pp. 1-34.
- Schmitz, C., Hotho, A., Jäschke, R., Stumme, G. (2006) Mining association rules in folksonomies. Data Science and Classification: proceedings of the 10th IFCS Conference, Studies in classification, Data Analysis and Knowledge Organization
- Smith, G. (2004) Atomiq: Folksonomy: social classification. Aug 3, 2004. Retrieved from: [http://atomiq.org/archives/2004/08/folksonomy\\_social\\_classification.html](http://atomiq.org/archives/2004/08/folksonomy_social_classification.html)

- Subramanya, S. B. and Liu, H. (2008). Socialtagger collaborative tagging for blogs in the long tail. In Proceeding of the 2008 ACM Workshop on Search in Social Media (Napa Valley, California, USA, October 30 30, 2008). SSM '08. ACM, New York, NY, 19-26.
- Taylor, A. G. (2003). The Organization of Information. Library and Information Science Text Series. Libraries Unlimited, 2nd. ed. edition.
- Yeung, C., Gibbins, N., and Shadbolt, N. (2009). Contextualising tags in collaborative tagging systems. In HT ,2009: Proceedings of the 20th ACM conference on Hypertext and hypermedia, pages 251-260, New York, NY, USA. ACM.
- Zhang, N., Zhang, Y., and Tang, J. (2009). A tag recommendation system for folksonomy. In King, I., Li, J. Z., Xue, G. R., Tang, J., King, I., Li, J. Z., Xue, G. R., and Tang, J., editors, CIKM-SWSM, pages 9-16. ACM.



## **CAPÍTULO 6 -**

# **Relación entre el crowdsourcing y la inteligencia colectiva: el caso de los sistemas de etiquetado social**

## 5.1 Introducción

Este capítulo se corresponde con el artículo “*Relationship between Collective Intelligence and Crowdsourcing: the social tagging systems case*”, actualmente en proceso de revisión en la revista *Computer Supported Cooperative Work*.

### 5.1.1 Resumen del artículo

El crowdsourcing es un concepto reciente, que suele asociarse en distinto grado con distintos procesos como la innovación abierta, la co-creación o la inteligencia colectiva. Aunque el crowdsourcing se nutre de todos ellos, no mantiene la misma relación con todos. En este artículo se va a profundizar en la relación que se mantiene con la inteligencia colectiva.

Utilizando los elementos que definen las plataformas de inteligencia colectiva propuestos por Malone et al. (2009, 2010) y utilizando también los elementos propuestos por Estellés-Arolas y González (2012) para identificar qué iniciativas/plataformas son de crowdsourcing, se analizan los sistemas de etiquetado social. De esta manera se comprueba como estos sistemas, aún siendo un claro ejemplo de inteligencia colectiva, no lo son de crowdsourcing.

### 5.1.2 Datos de la publicación

El artículo ha sido enviado a la revista *Computer Supported Cooperative Work (CSCW)*, revista centradas en las bases y características teóricas, prácticas y técnicas del trabajo colaborativo basado en ordenador. Se abarcan desde los estudios etnográficos del trabajo cooperativo hasta informes sobre el desarrollo de los sistemas de CSCW y sus fundamentos tecnológicos.

La revista está indexada tanto en *Science Citation Index* como en *Scopus*. Se encuentra en distintas bases de datos como *Academic Search Premier*, *ACM Computing Reviews*, *EBSCO*, *ACM Digital Library*, *DBLOP* o *OCLC* entre otras.

Esta revista tuvo en 2011 un índice de impacto JCR de 1.071, encontrándose la revista, según el ISI Journal Citation Report, en la posición 60/99 en la categoría de "Ciencias de la Computación, Aplicaciones Interdisciplinarias". Se encuentra en el tercer cuartil (Q3).

Los autores del artículos son, en orden de aparición, Enrique Estellés-Arolas y Fernando González Ladrón-de-Guevara.

- Nombre de la revista: *Computer Supported Cooperative Work*
- Editorial: Springer
- ISSN: 0925-9724

## 5.2 Artículo

# Relationship between Collective Intelligence and Crowdsourcing: the social tagging systems case

Enrique Estellés-Arolas

Department of Management, Technical University of Valencia, Valencia, Spain

Fernando González-Ladrón-de-Guevara

Department of Management, Technical University of Valencia, Valencia, Spain

### Abstract

Crowdsourcing is a term that continues growing in popularity. One of the consequences of this popularity is that it's being used indiscriminately, being identified with similar although not identical processes like open innovation, co-creation or collective intelligence. Another consequence is that, because the Web 2.0 is the technologic basis of crowdsourcing, some authors tend to associate and identify crowdsourcing with different Web 2.0 applications such as social networks or social tagging systems. This situation difficults the study of crowdsourcing because the concept is not completely delimited. To clarify the relationship of crowdsourcing with one the processes mentioned before, collective intelligence, this paper analyses the social tagging systems Web 2.0 application. The objective is to show that social tagging systems like Delicious, Flickr or Bibsonomy should not be considered crowdsourcing platforms. This way, although social tagging systems can be used for crowdsourcing purposes, this paper highlights the fact that their main function is related to collective intelligence, being clear examples of collective intelligence platforms.

### Keywords

collective intelligence; crowdsourcing; social bookmarking; social tagging; Web 2.0

## 1. Introduction

Crowdsourcing is the act of a company or institution taking a task and outsourcing it to an undefined and generally large network of people in the form of an open call (Howe, 2006). It has close ties to a set of processes that feed into it and with which it shares many characteristics, 'collective intelligence' being one of the most significant of these. Malone et

al. (2009; 2010) define collective intelligence as groups of people doing things collectively that seem intelligent, e.g. Linux, Wikipedia, etc.

The term crowdsourcing was coined in 2006 by Jeffrey Howe (2006), at about the same time as the term Web 2.0 was also officially coined (O'Reilly, 2005). In fact, crowdsourcing and Web 2.0 are closely linked (Mazzola & Distefano, 2010). The development of Web 2.0 has facilitated the use of a crowd to carry out archetypal crowdsourcing tasks, namely data collection and problem solving (Vukovic et al., 2009). Furthermore, Web 2.0 is the technological basis upon which crowdsourcing is developed and operates (Vukovic & Bartolini, 2010; Vukovic et al., 2010).

Because of these close ties, certain Web 2.0 applications (Andriole, 2010), for example, social tagging systems (STS) are often mistaken for crowdsourcing platforms (Howe, 2008; Bernstein et al., 2010; Geiger et al., 2011; Hirth, Hoßfeld & Tran-Gia, 2010; Huberman et al., 2009). STS, also known as collaborative tagging systems, are web platforms that allow users to manage online resources, such as web sites (i.e. Delicious) (Trant, 2009), images (i.e. Flickr) (Trant, 2009), scientific documents (i.e. Bibsonomy) (Hotho et al., 2006) or music (i.e. LastFM) (Lamere, 2008) amongst others, whereby metadata can be added in the form of keywords to the shared content (Golder & Huberman, 2006).

This paper tries to show that although STS are platforms that are included inside the collective intelligence paradigm, strictly speaking they are not examples of crowdsourcing platforms. This does not mean that social tagging cannot be used in crowdsourcing initiatives through platforms such as Diigo, Flickr, CiteUlike or Connotea. To illustrate the difference, the fundamental elements of both collective intelligence (Malone et al., 2010) and crowdsourcing (Geerts, 2009; Burger-Helmchen & Penin, 2010; Estellés-Arolas & González, 2012) are identified in three STS.

The paper comprises four main sections: section 2 introduces the fundamental concepts under consideration: STS, crowdsourcing and collective intelligence; section 3 describes the methodology; section 4 presents the outcomes of the research and section 5 presents the conclusions and puts forward potential topics for future research.

## **2. Research background**

This section describes the phenomena being researched in this paper, namely STS, crowdsourcing and Collective Intelligence.

## 2.1. Social tagging

Web 2.0 has facilitated the spread of tools that encourage participation and collaboration. Particularly prevalent are the social or collaborative tagging systems (Shepitsen et al., 2008). These are web applications that allow users to manage online resources, sharing them and allowing other users to add metadata to the shared content (Golder & Huberman, 2006).

Once a resource is stored, the STS allow the users to describe it by adding tags, that is a kind of metadata (Subramanya & Liu, 2008), i.e., data about data (Yeung, Gibbins & Shadbolt, 2009). In different kind of Web 2.0 systems, other metadata can be found: notes or comments, highlights, reviews or ratings. These tags are descriptive strings - words, phrases, or combinations of symbols and alphanumeric characters (Yeung et al, 2009) - generated by users (Millen et al., 2007). Tags are influenced by popular trends and colloquial vocabulary and represent personal knowledge, which imposes a soft organization in the data (Sawant et al., 2011).

When users assign tags to web resources this is described as a folksonomy (Illig et al., 2007), so folksonomies can be defined as sets of evolving categorization schemes or a set of terms with which a group of users tag content (Mathers, 2012).

Although tagging is not the same as social bookmarking (Geerts, 2009), the practices are similar – to the extent that some authors, such as Millen et al. (Millen et al., 2007), describe the use of tags as an essential characteristic of SBS. Furthermore, the use of SBS is intensified when social tagging is integrated within it. However, this does not mean that all SBS have to use tags; some, such as Digg, do not use them.

Although there are different STS that focus on different types of resources, the tags they use have the same kind of functionalities: identify what (or who) is the resource about, identify what the resource is, identify who owns it, refine categories, identify qualities or characteristics, self reference (i.e. mystuff, mycomments) or task organizing. Tags can also be useful for recalling information sources for later use as well as to communicate interesting nuggets of information to other users (Hammond et al., 2005).

The use of tags also implies two important problems: informational redundancy and the loss of general significance. The first one refers to the use of synonyms, homonyms and polysemes that leads to the elaboration of different tags by different users describing the same resource (Golder & Huberman, 2005). The second problem refers to the use of excessive

specific tags that will imply a certain level of ambiguity (i.e., “!fic”, “#cm10conf” o “#mn1010”): these tags will not be comprehensive for other users and will limit the effectiveness of collaborative tagging systems in document description and retrieval (Yeung et al., 2009).

## **2.2. Crowdsourcing**

Web 2.0 has enabled processes, such as crowdsourcing, that take advantage of the participation and collaboration that characterize it (Bonabeau, 2009). In fact, Web 2.0 is the technological basis upon which it is developed and operates thanks to the level of collaboration that can be achieved (Howe, 2008; Vukovic et al., 2009; Vukovic & Bartolini, 2010).

Crowdsourcing is defined as a type of participative online activity where an individual, institution, non-profit organization, or company proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, to voluntarily undertake a task. The task, of varying complexity and modularity, to which the participating crowd should bring their work, money, knowledge and/or experience, always entails mutual benefit. Users will receive the satisfaction of a given type of need, be it economic, social recognition, self-esteem, or the development of individual skills, while the crowdsourcer will obtain and utilize to their advantage whatever the user has brought to the venture, whose form will depend on the type of activity undertaken (Estellés-Arolas & González, 2012).

Some examples of platforms that allow crowdsourcing tasks to be carried out are Amazon and Mechanical Turk, where micro-tasks are proposed to a crowd in exchange for a financial reward (Kittur et al., 2008), or Threadless, an online t-shirt shop that allows users to create their own designs and vote for others' creations, allowing the crowd to decide which models go on sale (Brabham, 2008).

### **2.2.1 Elements of a crowdsourcing initiative**

Different authors identify different sets of elements that define crowdsourcing. Geerts (Illig et al., 2007) identifies three essential characteristics: the task is traditionally performed by a designated agent, the crowd is undefined, and open calls must be used. Burger-Helmchen and Penin (Burger-Helmchen & Penin, 2010) concur that an open call ensures non-discriminatory participation (Pénin, 2008) and a crowd has the characteristics of being large, with heterogeneous members, who do not know each other (Schenk & Guittard, 2009). However,

they point out that these characteristics may vary depending on the company implementing the crowdsourcing initiative and the tasks involved (Wolfson & Lease, 2011).

Estellés-Arolas and González (Estellés-Arolas & González, 2012) analyze more than 40 different definitions of crowdsourcing and identify eight characteristics that define crowdsourcing initiatives:

- Crowd – together with the open call (mentioned above). It is usually considered to be a generic and indeterminate group of individuals who do not necessarily know each other (Howe, 2008; Kleeman et al., 2008; Poetz & Schreier, 2012).
- Open and flexible call - used to contact the crowd (Poetz & Schreier, 2012; Sloane, 2011).
- Task – tasks range from routine to innovation-related tasks with a clear purpose (Reichwald & Piller, 2006; La Vecchia & Cisternino, 2010).
- A clear reward – the members of the crowd require some form of compensation, which may be in the form of social recognition, money, developing creative skills or sharing knowledge (Brabham, 2008).
- A clearly identifiable crowdsourcer – an individual or organization that initiates the crowdsourcing process (Brabham, 2008; Howe, 2008).
- Return to the crowdsourcer who obtains the solution to a problem through the crowd's working on a specific action or task (Kleeman et al., 2008; Vukovic et al., 2009).
- Use of a distributed, online process which enables the resolution to a problem (Ling, 2010).
- Use of the Internet as the medium and technological basis upon which crowdsourcing operates and is developed due to the required level of collaboration (Vukovic et al., 2009).

### **2.3. Collective intelligence**

Wechsler (1971) defines intelligence as the composed or global ability of an individual to act purposefully, think reasonably, and to deal effectively with changing and difficult environment situations. Individuals learn to understand and to adapt to their context drawing on their accumulated knowledge (Leimeister, 2010). Furthermore, collective intelligence is defined as a form of universally distributed intelligence, constantly enhanced, coordinated in

real time, and resulting in the effective mobilization of skills (Lévy, 2001). Surowiecki's (2005) book "The Wisdom of Crowds" describes successful collective decisions being made not by consensus building, but as a result of competition among independent opinions. Lykourantzou et al. (2009) define a collective intelligence system as a "system which hosts an adequately large group of people, who act for their individual benefit, but whose group actions aim at and may result – through technology facilitation – in a higher-level intelligence and benefit of the community".

An important aspect of collective intelligence theory is that the combination of the efforts expended by a crowd of numerous individuals can produce a result that will be better than that provided by an individual expert. The group of individuals is more intelligent than any of its single members. An intelligent and complex behavior emerges from the synergy created by the simple interactions among the members of a group that follows simple rules and competes through diversity (Heylighen, 1999; von Hippel & Katz, 2002). Collective intelligence can be used for social collaboration, crowdsourcing, consensus decision-making (Bonabeau, 2009), mass communications, open innovation (Chesbrough, 2003) and other phenomena (Leimeister, 2010; Gregg, 2010). The concept has become prominent due to the Internet, but it has existed for a long time and has been developed in human cultures either spontaneously or intentionally (Leimeister, 2010; Murty et al., 2010).

### **2.3.1 Elements for a collective intelligence initiative**

Malone, Laubacher and Dellarocas (2009; 2010), from the MIT Center for Collective Intelligence, studied 250 cases of collective intelligence focusing on the diversity reflected in the methods and aims of each case. Malone et al. (2010) identified a small set of elements that combine in each example in different ways. They refer to these components as genes of collective intelligence systems, a biological analogy. They describe the "Collective Intelligence building blocks or genes that can be recombined to create the right kind of system", so the specific combination of genes associated with a specific example of collective intelligence would constitute its "genome" (Georgi & Jung, 2012).

These identifiable elements (genes) in collective intelligence activities are based on four basic questions: "who", "why", "what" and "how".

- "Who" refers to who carries out the task. It can be performed by the Internet crowd, or by a hierarchy. A hierarchy will carry out the task only if the task is assigned to someone



from a higher position considering a more delimited group of people (like the members of a company) (Leimeister, 2010; Georgi & Jung, 2012);

- "Why" refers to why people carry out the task. Its three elements or genes correspond to three incentives: financial benefit (money); love (enjoyment derived from carrying out the task, intrinsic motivation, or satisfaction from contributing to a bigger meaningful task); and glory (the desire to be recognized by peers) (Leimeister, 2010);
- "What" is related to the task that is being performed and has two elements or genes: creating and deciding. In the creating process some types of content are generated, such as source code, text, design, etc.; in the decision process, decisions are made by evaluating and selecting alternatives (Pénin, 2008; Leimeister, 2010);
- "How" refers to how a task is performed and has two associated elements or genes: independent and dependent. Georgi and Jung (2012) point out that these genes/elements depend on what is being carried out. Therefore, if a task consists of a creation process carried out in an independent way, it can be done by elaborating a collection (the task can be split into items that can be solved independently of each other), or a contest (a subtype of a collection involving competing people). If creating is dependent, then collaboration tends to be the choice (there are dependences among the various subtasks of the main task). If the task involves decisions, then in Georgi and Jung's (2012) view, they can be carried out in a dependent way, through voting, consensus, averaging or prediction, or in an independent way based on an individual's decision.

### 3. Methodology

In order to demonstrate that STS are not crowdsourcing platforms, but rather collective intelligence ones, a description of the elements characterizing both paradigms is provided.

In the case of collective intelligence, this paper has used the genes put forward by Malone et al. (2009; 2010) and also used by Leimeister (2010) and Georgi and Jung (2012). With regard to the elements defining crowdsourcing, this paper uses those proposed by Estellés-Arolas and González (2012) because they incorporate elements proposed by other authors as well as adding some of their own.

Then, an analysis trying to identify those elements in examples of social tagging systems is carried out. This analysis consists in the elaboration and interpretation of an analysis grid

(Estellés-Arolas & González, 2012; Vukovic, 2009) that takes into account, for each platform case, both shared and non-shared elements.

So a STS is considered to be an example of collective intelligence or a crowdsourcing platform if it displays all the distinctive elements characterizing these paradigms.

## 4. Results

Three STS cited in the specialist literature on crowdsourcing were chosen: Delicious -a social bookmarking web service (Howe, 2008), Flickr -an online photo management and photo sharing application (Huberman et al., 2009; Bernstein et al., 2010) and Bibsonomy -web-based platform for sharing scientific resources (Yuem et al., 2011).

Delicious is a social bookmarking web service that allows users to save, manage, tag, and share web pages from a centralized source (Maharana et al., 2010). It is possible to view bookmarks added by similar-minded users and to improve the ways in which people discover, share and recall them on the Internet. These public bookmarks are searchable and the results of such searches are considered useful by a number of people because are not currently provided by other sources (Heyman et al., 2008). Delicious allows users to tag content to enable access to it (Boydell & Smith, 2007), and to group links to similar topics to form stacks that include descriptions.

Flickr is an online photo management and photo sharing application that allows users to upload, search, sell and share their personal photos and videos. Users can manage their images through tools that allow the content of their pictures to be tagged and explored and enable the user to comment on others' images (Marlow et al., 2006; Lerman et al., 2007; Angus et al., 2008).

Bibsonomy is a web-based platform of the Knowledge and Data Engineering Group of University of Kassel (Germany), that has features for sharing and tagging bookmarks. It also facilitates the literature exchange and research so users can collect, organize and share bookmarks and publications. About the sharing issue, Bibsonomy allows sharing bookmarks and bibliographic references simultaneously and its lightweight knowledge (folksonomy) evolves from user participation. To store a reference to a resource in BibSonomy, the user has to provide corresponding meta information and tags describing the resource, the user's opinion and the research project for which this resource could be relevant. Publication posts in BibSonomy are stored in the BibTex format and these posts are available to others.

## 4.1. Collective intelligence

According to Malone et al. (2009; 2010), the fundamental questions that generate the genes of collective intelligence are “who”, “why”, “what” and “how” questions. Each gene is identified in the selected STS, and the results can be seen in Table 6.1. This table summarizes the analysis so far and shows the way each STS complies with each collective intelligence gene.

**Table 6.1. Identification of the collective intelligence elements that appear in the selected STS.**

Example	What		Who	Why	How
BibSonomy	Create	Bookmarking a reference	Crowd	Love, Glory	Collection
	Decide	Which references appear in the “Popular” section	Crowd	Love, Glory	Tagging
Delicious	Create	Bookmarking a web resource	Crowd	Love, Glory	Collection
	Decide	Which bookmarks will appear on the front page	Delicious staff	Love, Glory	Hierarchy
Flickr	Create	Upload pictures	Crowd	Love, Glory	Collection
	Decide	Pictures that appear in the “interesting” section	Crowd	Love, Glory	Voting and voting variations (views, comments, etc)

### 4.1.1. Who

An important element used to determine the users of a specific STS, is the type of resource that is stored there. On one hand, there are STS that deal with commonly used resources, such as web pages -Delicious- or images -Flickr. In these particular cases, the potential users of any of these platforms will be the generic Internet crowd (Hammond et al., 2005; Georgi & Jung, 2012). On the other hand, when the resource is very specific, the crowd comprises a specific, and very distinguishable group. This is the case of Bibsonomy, whose users are normally related to academic and research areas due its work with academic publications (Hammond et al., 2005; Golder & Huberman, 2006; Borrego & Fry, 2012).

In both generic Internet crowds and specific groups, we always find two main groups, which often overlap (Boydell & Smith, 2007): users that generate content or creators (whether an image, a bookmark, a reference, etc.) and those that consume that content or consumers, who select only those elements that interest them (EuropaPress, 2012).

### 4.1.2. Why

According to Malone et al. (2009; 2010), there are three main motivations for participating in collective intelligence initiatives: money, love (the enjoyment of performing the task) and glory (or social recognition) (Lykourantzou et al., 2009), all of which apply to open source programming communities (Tapscott & Williams, 2010), which demonstrate the ability of masses to achieve common goals through collaborative effort on the web (Preece & Sneiderman, 2009; Leimeister, 2010).

When talking about tagging, different authors (Hammond et al., 2005; Heckner, Heilemann & Wolff, 2009; Strohmaier et al., 2010; Körner et al., 2010) present two general motivations according to the taggers degree of contribution to emerging semantic structures, and these are represented by two distinct groups: categorizers and describers. Categorizers use a small set of tags instead of hierarchical classification schemes, and describers usually annotate using many freely associated, descriptive keywords.

As far as motivations for using STS are concerned, (Benbunan-Fich & Koufaris, 2008) consider them to be public repositories of information and state that two distinct types can be found: self-oriented reasons and motives related to others. Benbunan-Fich & Koufaris (2008) consider that self-oriented motives are associated with the quantity and quality of the contributions, while other-oriented motives are associated only with the quality of the contributions for others. In the latter, users contribute tagged resources that they believe to be useful to those users.

Wash & Rader (2007) identify two additional incentives for using STS: accessing consolidated sets of bookmarks from different computers, and subsequently organizing them. Tagging, in this case, is used as a way to enhance the organization of information. Other oriented-motives include achievement of a social presence through the sharing of bookmarks and use of tags as a way to express opinion, self-presentation and activism (Bischoff et al., 2008).

Due to the tagged content, each STS is associated with different types of motivation. In fact, it is important to highlight that users' motivations for tagging vary across and within tagging systems (Strohmaier et al., 2010). Wash & Rader (2007) identify three main reasons for using Delicious: to keep track of useful or interesting pages, to access bookmarks from multiple computers and to gain recognition from other users. In other words, the personal benefits of using the tool and the glory or recognition gained, as identified by Malone et al. (2009; 2010). In the case of Flickr, users are motivated primarily by social incentives, including

opportunities to share and play (glory and love) (Marlow et al., 2006). In the case of Bibsonomy, no document studying the users' motivation has been found. Nevertheless, the use of Bibsonomy does not imply any financial gains, so we can disregard this motivation. Along the same lines, glory, in other words acknowledgement of other users (through commentaries, tagging etc) of uploaded resources could be an important factor to take into account (Malone et al., 2010). Moreover, the use of Bibsonomy for its functionality, the ability to organize and share references that are being used, is another clear source of motivation for its use which we could associate with the motivation of love.

#### **4.1.3. What**

The three selected STS allow users to create content and, to a certain extent, to assess it (Bonabeau, 2009); the two tasks that can be done in collective intelligence activities according to Malone et al. (2009; 2010). Delicious allows the addition of links to websites that might be of interest to the user, as well as tag and comment on these links (Hammond et al., 2005; Maharana et al., 2010). Flickr enables uploading, describing and tagging of images, from the owner and from other users (Marlow et al., 2006). Bibsonomy allows users to bookmark, describe and tag researching documents like journal articles, conference proceedings, books, etc. sharing them with any other Internet user and allowing registered users to tag them (Borrego & Fry, 2012).

#### **4.1.4. How**

Malone et al. (2009; 2010) put forward the idea that the performance of a task, whether a creation or an evaluation, can be carried out in a dependent or independent way. An independent creation process can be achieved by a collection or via a contest. A dependent creation process is based on collaboration. An independent process involves an individual decision; a dependent process involves decisions based on assessment, consensus, averages and predictions.

In all three cases studied, creating content (websites bookmarks, images or scientific documents) is considered to be an independent process. Users tag the resources they want or upload the content they consider appropriate.

In some cases, the stored resource can also be evaluated, and this may involve making decisions about its positioning on the front page. Such evaluations are carried out in a dependent way through voting. In the case of Bibsonomy, the most tagged bookmarks appear in the "Popular" section. In the case of Flickr, the section "Explore" includes the photographs

used in most of the user's interactions (the most visited, the most commented on, those that are added to favorites, etc.) (Flickr, 2012). About Delicious, the evaluation of tagged bookmarks is carried out by the Delicious staff.

## 4.2. Crowdsourcing

Next, the selected STS are analyzed based on the criterion of the distinctive elements of crowdsourcing initiatives (Estellés-Arolas & González, 2012). Table 6.2 presents the results of this analysis.

**Table 6.2. Elements of crowdsourcing in the selected SBS. '+' indicates presence of the characteristic; '-' indicates absence of the characteristic**

	E1	E2	E3	E4	E5	E6	E7	E8
<b>Delicious</b>	-	+	-	-	+	+	+	-
<b>Flickr</b>	-	+	-	-	+	+	+	-
<b>Bibsonomy</b>	-	+	-	-	+	+	+	-

### 4.2.1. A task with a clear purpose (E1)

This task is similar to Malone et al.'s (2009; 2010) 'what' question, except that crowdsourcing tasks are aimed at specific goals (Estellés-Arolas & González, 2012): i.e. designing a t-shirt on the creative design web Threadless to sell it in the Threadless online shop, solving a problem on the platform InnoCentive, etc. In these cases, the company that begins the crowdsourcing initiative has an objective that needs the crowd activity to be reached.

In the case of the STS selected for this study, although the task of tagging a resource (an image, a web page or a literature reference) is clearly defined, the task itself can have different objectives according to the needs of the user carrying it out (Benbunan-Fich & Koufaris, 2008). Regardless of whether a user is described as a descriptor or classifier (Strohmaier, Körner & Kern, 2010), he may tag a resource to indicate if he likes it or not or whether it is useful for his work. A user may upload pictures as a means of promoting his skills as a professional photographer, and tag them to make them more Internet searchable. Alternatively, the same user can upload pictures and tag them simply in order to share them with family and friends.

Thus, there is no clear and established purpose for the tagging activity in any of the selected STS.

#### **4.2.2. A reward easily identifiable by the crowd (E2)**

This corresponds to Malone et al.'s (2009; 2010) 'why' question and refers to the crowd's motivation for performing the task. In this case, the same incentives for collective intelligence described in section 4.1.2 can be found: Glory and Love are present in Delicious (Wash & Rader, 2007), Flickr (Marlow et al., 2006) and Bibsonomy (Borrego & Fry, 2012). In addition to the rewards or incentives that encourage people to repeat tasks, personal benefits must also be taken into consideration, such as the ability to organize and describe photos or store a set of web resources in order. This characteristic is present in the selected STS.

#### **4.2.3. A crowdsourcer (E3)**

The crowdsourcer is the individual, institution or organization that proposes the crowdsourcing initiative with the aim of having a specific task performed (Estellés-Arolas & González, 2012). Although the companies behind the STS under study are different - AVOS Systems owns Delicious, the University of Kassel owns BibSonomy and Yahoo! owns Flickr - none of these companies or institutions launched or acquired their platforms for the sole purpose of getting users to perform specific tasks they could directly benefit from. These companies cannot therefore be considered crowdsourcers.

#### **4.2.4. A reward easily identifiable by the crowdsourcer (E4)**

Although these systems use the user-generated content to obtain income (through promotion), the generation of content on its own does not provide any revenue. The company business models envisage different sources of income: advertising and paid-subscription user accounts (in the case of Flickr). So there is no direct reward for the crowdsourcer.

#### **4.2.5. Use of an online participative process (E5)**

This corresponds to Malone et al.'s (2009; 2010) 'how' question. The three STS display this element in crowdsourcing. In all cases, the use of platforms is based on a process whereby users participate by adding and tagging bookmarks of webs, bookmarks of references or images in different ways.

#### **4.2.6. Internet use (E6)**

STS are a clear example of Web 2.0 applications (Farooq et al., 2007) and involve tagging resources from the Internet. Internet use is an essential requirement of STS.

#### **4.2.7. A crowd (E7)**

This crowd is aligned to Malone et al.'s (2009; 2010) 'who' question in relation to collective intelligence. All the SBS analyzed use the general Internet crowd, with the exception perhaps of the specialist STS (Golder & Huberman, 2006) like BibSonomy, which is normally used by academics and scientists (although it is open to anyone in Internet).

#### **4.2.8. An open call (E8)**

An open call refers to the crowdsourcer's request to the crowd for a task to be carried out. In this regard, the STS doesn't use any open call because there is no request for a specific task to be carried out. STS are free services that are always open to new users and encourage existing users to continue to use them.

It is important to point out that the services of the STS, always available, should not be confused with a "permanent open call" (Kleeman et al., 2008). In this type of open call, the crowdsourcer's invitation does not refer to a task to be carried at a specific moment, but rather over a period of time (such as the submission of information or documents in the case of amateur reporters).

## **5. Conclusions and future research**

In this paper, three STS were analyzed in order to determine whether or not they are in fact cases of crowdsourcing platforms. Collective intelligence genes suggested by Malone et al. (2009; 2010) and the elements defining crowdsourcing proposed by Estellés-Arolas & González (2012) were identified. Following analysis, a number of conclusions were able to be drawn.

First, the study of the three social tagging systems has allowed us to identify similar features in all cases, as seen in Table 1. STS are systems where a crowd of people from the Internet community creates content through collections. Some of the content appears in special web sections due to the decision of company staff, some due to the opinion of users, some due to both. In all cases, the motivation behind user decisions tends to be based on Love & Glory. Barring certain exceptions, namely tagging resources with which they work, or if the social tagging systems cater to a more specialized audience (such as the case of Bibsonomy), all follow a similar pattern and structure, and are all clear examples of collective intelligence.

It can however also be concluded that substantial evidence exists to suggest that STS are not examples of crowdsourcing platforms since only four of the eight elements of crowdsourcing



are present (cf. Table 2). Although elements E2, E5, E6 and E7 are present in all the STS studied, it has been shown that other essential elements are absent. In the case of all the social tagging systems analyzed, the achievement of the task (tagging a resource) does not have a clear goal for the crowdsourcer, if this figure exists. Although there are companies that finance these SBS, they do not meet the requirements of being a crowdsourcer nor do they directly profit from the tasks that users perform (tagging a resource or uploading an image). Lastly, none of them employ an open call.

The reason why STS are often seen as the same as crowdsourcing platforms is that the latter initiative occurs within the collective intelligence framework. Similarly, the four elements described by Malone et al. (2009; 2010) can be identified within the crowdsourcing elements described by Estellés-Arolas & González (2012), although admittedly with certain peculiarities.

We can therefore conclude that although crowdsourcing is a particular manifestation of Collective Intelligence (i.e.: the user tasks carried out in InnoCentive), not every Collective Intelligence activity should be classified as crowdsourcing (i.e.: the user tasks carried out in Delicious).

It is also important to note that although STS are not examples of crowdsourcing platforms, these can be used for crowdsourcing tasks that involve bookmarking.

With an eye towards future research, it is essential to continue differentiating crowdsourcing from other similar terms. This paper, in addition to highlighting the fact that STS are not examples of crowdsourcing platforms, helps to differentiate the term crowdsourcing by determining its ties to Collective Intelligence. In this regard, there is still a great deal of terminology that continues being confused with the term crowdsourcing: Open Innovation, Co-Creation, User-Innovation or Outsourcing, for example. It would therefore be of interest to carry out a thorough analysis of the characteristics of said terms in order to identify their similarities and relationships to crowdsourcing.

## 6. Bibliography

- Andriole, S.J. (2010) Business impact of Web 2.0 technologies, *Communications of the ACM*; 53(12): 67-79.
- Angus, E., Thelwall, M. and Stuart, D. (2008) General patterns of tag usage among university groups in Flickr. *Online Information Review* 2008; 32(1):89-101.

- Benbunan-Fich, R. and Koufaris, M. (2008) Motivations and Contribution Behavior in Social Bookmarking Systems: An Empirical Investigation. *Electronic Markets*; 18(2):150-160.
- Bernstein, M.S., Tan, D., Smith, G., Czerwinski, M. and Horvitz, E. (2010) Personalization via friendsourcing. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 17(2):1-28.
- Bischoff, K., Firan, C. S., Nejdil, W. and Paiu, R. (2008) Can all tags be used for search?. In: *Proceedings of the 17th ACM conference on Information and knowledge management, CIKM '08*; New York, USA:ACM; pp. 193–202.
- Bonabeau, E. (2009) Decisions 2.0: The power of collective intelligence. *MIT Sloan Management Review*; 50(2): 45–52.
- Borrego, Á. and Fry, J. (2012) Measuring researchers' use of scholarly information through social bookmarking data: A case study of BibSonomy. *Journal of Information Science*; 38(3):297–308
- Boydell, O. and Smyth, B. (2007) From social bookmarking to social summarization: an experiment in community-based summary generation. In: *Proceedings of the 12th International Conference on Intelligent User Interfaces, IUI '07*. New York, USA:ACM; pp. 42–51
- Brabham, D. C. (2008) Moving the crowd at iStockphoto: The composition of the crowd and motivations for participation in a crowdsourcing application; *First Monday* 13(6)
- Burger-Helmchen, T. and Penin, J. (2010) The limits of crowdsourcing inventive activities: What do transaction cost theory and the evolutionary theories of the firm teach us? In: *Workshop on Open Source Innovation, Strasbourg, France*.
- Chesbrough, H. W. (2003) The Era of Open Innovation. *Sloan Management Review*; 44 (3): 35-41.
- Estellés-Arolas, E. and González Ladrón-de-Guevara, F. (2012) Towards an integrated crowdsourcing definition. *Journal of Information Science*; 38(2): 189-200.
- EuropaPress 'Google explica como se comportan los usuarios de YouTube', <http://www.europapress.es/portaltic/internet/noticia-google-explica-son-usuarios-youtube-20110617153125.html> (2011, accessed december 2012).
- Farooq, U., Kannampallil, T.G., Song, Y., Ganoë, C.H., Carroll, J.M. and Giles, L. (2007) Evaluating tagging behavior in social bookmarking systems: metrics and design heuristics.

- In: GROUP '07: Proceedings of the 2007 international ACM conference on Supporting group work. New York, NY, USA: ACM; 2007. p. 351–360.
- Flickr 'About Interestingness'. <http://www.flickr.com/explore/interesting/> (2012, accessed december 2012)
  - Geerts, S. (2009) Discovering crowdsourcing: theory, classification and directions for use. (Technische Universiteit Eindhoven).
  - Geiger, D., Seedorf, S. and Schader, M. (2011) Managing the Crowd: Towards a Taxonomy of Crowdsourcing Processes. In: Proceedings of the Seventeenth Americas Conference on Information Systems, Detroit, Michigan August 4th-7th 2011.
  - Georgi, S. and Jung, R. (2012) Collective Intelligence Model: How to Describe Collective Intelligence. In: J. Altmann, U. Baumöl, & B. J. Krämer (Eds.), Advances in Collective Intelligence 2011, Advances in Intelligent and Soft Computing. 113: 53–64). Springer Berlin / Heidelberg.
  - Golder S. A. and Huberman B. A. (2006) Usage Patterns of Collaborative Tagging Systems. *Journal of Information Science*; 32(2): 198-208.
  - Golder, S., Huberman, B.A. (2005) The Structure of Collaborative Tagging Systems. *CoRR*, 2005 Aug;
  - Gregg, D. G. (2010) Designing for collective intelligence. *Communications of the ACM*; 53(4): 134–138.
  - Hammond, T, Hannay, T, Lund, B, Scott, J. (2005) Social Bookmarking Tools (I). *D-Lib Magazine* 2005; 11(04).
  - Heylighen, F. (1999) Collective Intelligence and its Implementation on the Web: algorithms to develop a collective mental map. *Computational & Mathematical Organization Theory* 1999; 5(3): 253-280.
  - Heymann, P., Koutrika, G. and Molina, H.G. (2008). Can social bookmarking improve web search? In: Proceedings of the International Conference on Web Search and Web Data Mining. WSDM '08. New York, NY, USA: ACM; 2008. p. 195–206.
  - Hirth, M., Hoßfeld, T., and Tran-Gia, P. (2010) Cheat-detection mechanisms for crowdsourcing. Technical report, University of Würzburg.
  - Hotho, A., Jauchke, R., Schmitz, C. and Stumme, G. (2006) BibSonomy: A Social Bookmark and Publication Sharing System. In: Proceedings of the Conceptual Structures Tool Interoperability Workshop at the 14th International Conference on Conceptual Structures; pp. 87-102.

- Howe J. (2006) The rise of crowdsourcing. *Wired*; 14 (6)
- Howe, J. (2008) *Crowdsourcing: How the Power of the Crowd is Driving the Future of Business*. Great Britain: Business Books.
- Huberman, B.A., Romero, D.M. and Wu, F. (2009) Crowdsourcing, Attention and Productivity. *Journal of Information Science*; 35(6): 758–765.
- Illig, J., Hotho, A., Jäschke R. and Stumme, G. (2007) A comparison of content-based tag recommendations in folksonomy systems. In: *Proceedings of the First International Conference on Knowledge Processing and Data Analysis*, 136–149. Berlin: Springer-Verlag.
- Kittur, A., Chi, E.H., Suh, B. (2008) Crowdsourcing user studies with Mechanical Turk. In: *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*. CHI '08. New York, NY, USA: ACM; p. 453–456.
- Kleeman, F., Voss, G.G. and Rieder, K. (2008) Un(der)paid Innovators: The Commercial Utilization of Consumer Work through Crowdsourcing. *Science, Technology and Innovation Studies*; 4(1): 5-26.
- Körner, C., Benz, D., Hotho, A., Strohmaier, M. and Stumme, G. (2010) Stop thinking, start tagging: tag semantics emerge from collaborative verbosity. In: *Proceedings of the 19th international conference on World wide web*; pp. 521–530.
- La Vecchia, G. and Cisternino, A. (2010) Collaborative workforce, business process crowdsourcing as an alternative of BPO. In: *Proceedings of First Enterprise crowdsourcing Workshop in conjunction with ICWE 2010*. Berlin/Heidelberg:Springer-Verlag; 2010. pp. 425-430
- Lamere, P. (2008) Social Tagging and Music Information Retrieval. *Journal of New Music Research*; 37(2):101–114.
- Leimeister, J. (2010) Collective Intelligence. *Business & Information Systems Engineering*; 2(4): 245–248.
- Lerman, K. (2007) User Participation in Social Media: Digg Study. In: *Proceedings of the 2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology - Workshops, WI-IATW '07*. Washington, DC, USA: IEEE Computer Society, pp. 255–258.
- Lerman, K., Plangprasopchok, A. and Wong, C. (2007) Personalizing Image Search Results on Flickr. In: *Proceedings of AAAI workshop on Intelligent Techniques for Information Personalization*. Vancouver, Canada, AAAI Press.

- Lévy, P. (2001) Collective intelligence. *Reading digital culture*; 4: 253
- Ling, P. (2010) An Empirical Study of Social Capital in Participation in Online crowdsourcing, *Computer*; 7(9): 1-4.
- Lykourantzou, I., Vergados, D. J. and Loumos, V. (2009) Collective intelligence system engineering. In: *Proceedings of the International Conference on Management of Emergent Digital Ecosystems*; New York:ACM; Article N°. 20
- M. Heckner, M. Heilemann, and C. Wolff. (2009) Personal information management vs. resource sharing: Towards a model of information behaviour in social tagging systems. In: *Proceedings of Int'l AAAI Conference on Weblogs and Social Media (ICWSM)*; San Jose, CA, USA; May 2009.
- Maharana, B., Majhi, S. and Bhue, S. (2010) Social Bookmarking: Web2. 0 Tool for Content Sharing and Learning. In: *Proceedings of the 7th Convention PLANNER*, Tezpur University, Assam February 18-2-0.
- Malone, T. W., Laubacher, R. and Dellarocas, C. N. (2009) *Harnessing Crowds: Mapping the Genome of Collective Intelligence*. MIT Sloan; Research Paper No. 4732-09.
- Malone, T. W., Laubacher, R., and Dellarocas, C. N. (2010) The collective intelligence genome. *MIT Sloan Management Review*; 51 (3): 21-31.
- Marlow, C., Naaman, M., Boyd, D. and Davis, M. (2006) HT06, tagging paper, taxonomy, Flickr, academic article, to read. In: *Proceedings of the seventeenth conference on Hypertext and Hypermedia, HYPERTEXT '06*. New York, USA:ACM; pp. 31–40.
- Mathes, A. (2012) 'Folksonomies—Cooperative classification and communication through shared metadata', <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html> (2004, accessed December 2012).
- Mazzola, D. and Distefano, A. (2010) Crowdsourcing and the participation process for problem solving: the case of BP. In: *VII Conference of the Italian Chapter of AIS. Information technology and Innovation trend in Organization*. (Napoles, Italy, 2010)
- Millen, D.R., Yang, M., Whittaker S. and Feinberg J. (2007) Social bookmarking and exploratory search. In: *Proceedings of ECSCW 2007*, Limerick, Ireland (Sept 26-28).
- Murty, P., Paulini, M. and Maher, M.L. (2010) Collective Intelligence and Design Thinking. In: *Proceedings of the Design Thinking Research Symposium, DTRS'10*; Sydney, Australia. 2010.
- O'Reilly, T. (2005) 'What is Web 2.0?', <http://oreilly.com/web2/archive/what-is-web-20.html> (2005, accessed December 2012).

- Pénin, J. (2008) More open than open innovation? Rethinking the concept of openness in innovation studies. Working papers of BETA, Bureau d'Economie Théorique et Appliquée, UDS, Estrasburgo.
- Poetz, M. K. and Schreier, M. (2012) The Value of Crowdsourcing: Can Users Really Compete with Professionals in Generating New Product Ideas? *Journal of Product Innovation Management*; 29(2): 245–256.
- Preece, J. and Shneiderman, B. (2009) The reader-to-leader framework: Motivating technology-mediated social participation. *AIS Transactions on Human-Computer Interaction*; 1(1): 13–32.
- Reichwald, R. and Piller, F.T. (2006) *Interaktive Wertschöpfung. Open Innovation, Individualisierung und neue Formen der Arbeitsteilung*. Wiesbaden:Gabler.
- Sawant, N., Li, J. and Wang, J.Z. (2011) Automatic image semantic interpretation using social action and tagging data. *Multimedia Tools Appl.*; 51(1):213–246.
- Schenk, E. and Guittard, C. (2009) *Crowdsourcing: What can be Outsourced to the Crowd, and Why?* Technical Report. <http://halshs.archives-ouvertes.fr/halshs-00439256/> (2009, accessed December 2012).
- Shepitsen, A., Gemmell, J., Mobasher, B., and Burke, R. (2008) Personalized recommendation in social tagging systems using hierarchical clustering. In: *Proceedings of the 2008 ACM conference on Recommender systems, RecSys '08*, 259-266. New York, NY, USA: ACM.
- Sloane, P. (2011) The brave new world of open innovation. *Strategic Direction*; 27(5):3-4
- Strohmaier, M., Körner, C., Kern, R. (2010) Why do users tag? detecting users' motivation for tagging in social tagging systems. In: *International AAAI Conference on Weblogs and Social Media (ICWSM2010)*, Washington, DC, USA.
- Subramanya, S.B. and Liu, H. (2008) Socialtagger - collaborative tagging for blogs in the long tail. In: *Proceeding of the 2008 ACM workshop on Search in social media. SSM '08*. New York, NY, USA: ACM; p. 19–26.
- Surowiecki, J. (2005) *The wisdom of crowds*. New York: Anchor Books.
- Tapscott, D. and Williams, A. D. (2010) *Wikinomics: How Mass Collaboration Changes Everything*. Penguin Group USA.
- Trant, J. (2009) Studying social tagging and folksonomy: A review and framework. *Journal of Digital Information*; 10 (1): 1-42.

- von Hippel, E. and Katz, R. (2002) Shifting Innovation to Users via Toolkits. *Management Science*; 48(7): 821-833.
- Vukovic M., Crowdsourcing for enterprises. In: *Proceedings of the 2009 Congress on Services – I*, IEEE Computer Society (Washington, DC, USA 2009). 686-692.
- Vukovic M., Mariana L. and Laredo J. (2009) PeopleCloud for the Globally Integrated Enterprise. In: D. Asit et al. (eds) *Service-Oriented Computing*. (Springer-Verlag, Berlin/Heidelberg, 2009)
- Vukovic, M. and Bartolini, C. (2010) Towards a Research Agenda for Enterprise crowdsourcing. In: M. Tiziana and S. Bernhard (eds) *Leveraging Applications of Formal Methods, Verification, and Validation* (Springer, Berlin/Heidelberg, 2010) 425-434 [Lecture Notes in Computer Science 6415].
- Vukovic, M., Laredo, J. and Rajagopal, S. (2010) Challenges and experiences in deploying enterprise crowdsourcing service. In: *Proceedings of the 10th international conference on Web engineering*. Springer-Verlag Berlin, Heidelberg.
- Wash, R. and Rader, E. (2007) Public bookmarks and private benefits: An analysis of incentives in social computing. *American Society for Information Science and Technology (ASIS&T) Annual Meeting*, Milwaukee, WI.
- Wechsler, D. (1971) Intelligence: definition, theory and the IQ. In: Cancro, R. (Ed.) *Intelligence: genetic and environmental influences*. Grune Straton. New York, 50-55.
- Wolfson, S. M. and Lease, M. (2011) Look before you leap: Legal pitfalls of crowdsourcing. In: *Proceedings of the American Society for Information Science and Technology 2011*; 48(1): 1–10.
- Yeung, C., Gibbins, N. and Shadbolt, N. (2009) Contextualising tags in collaborative tagging systems. In: *HT '09: Proceedings of the 20th ACM conference on Hypertext and hypermedia*. New York, NY, USA: ACM; p. 251–260.
- Yuen, M.C., King, I., Leung, K.S. (2011) A Survey of Crowdsourcing Systems. In: *Proceedings of the IEEE Third International Conference on Social Computing (SocialCom)*. IEEE; p. 766–773.

**CAPÍTULO 7 -**

# **Conclusiones y trabajo futuro**



## 7.1. Introducción

En este capítulo, se recogen las conclusiones alcanzadas en los distintos artículos, además de una conclusión general sobre la temática de la tesis. Posteriormente, se finaliza con la enumeración de algunas posibles líneas de investigación.

## 7.2. Conclusiones

La elaboración de la presente tesis y la publicación de los distintos artículos que la forman, ha permitido conocer la situación actual del crowdsourcing y asentar el concepto.

Por un lado, la revisión de la literatura existente ha permitido identificar la falta de consenso y la confusión semántica en determinados campos relacionados con el crowdsourcing.

Uno de estos campos es el de la definición del término, encontrándose más de 30 definiciones realizadas por distintos autores. A partir del estudio de todas estas definiciones se ha podido desarrollar una definición general del crowdsourcing, basada en ocho elementos claramente identificables: la multitud, la tarea a realizar, la recompensa obtenida, el crowdsourcer, el resultado obtenido por el crowdsourcer, el tipo de proceso, la llamada a la participación y el medio utilizado. En cada tipo concreto de iniciativa de crowdsourcing, estos elementos se manifestarán de una manera distinta. Por ejemplo, en el crowdfunding, la tarea a realizar será la donación monetaria, mientras que en el crowdvoting, implicará la manifestación de la opinión de la multitud a través de un voto o comentario sobre un producto.

De esta manera, ante la posible falta de acuerdo sobre qué es o no una iniciativa o plataforma de crowdsourcing, esta definición puede suponer una útil herramienta que permita diferenciar qué es puramente crowdsourcing de lo que no lo es.

Otro de los campos donde existe una falta de consenso, es el de las tipologías para iniciativas de crowdsourcing. Existen distintos criterios que pueden ser utilizados para clasificar dichas iniciativas, aunque el tipo de tarea a realizar por parte de la multitud es uno de los que más reflejan esta falta de consenso.

De esta manera, mediante una revisión de la literatura centrada en la búsqueda de las tipologías existentes, se ha podido identificar hasta 6 tipologías distintas de las iniciativas de crowdsourcing. Comparando los distintos elementos de estas tipologías, se ha podido elaborar una nueva tipología de carácter integrador, que además ha sido testada con éxito en 15 casos de crowdsourcing seleccionados al azar.

Junto con la definición desarrollada, esta tipología integradora ayuda a delimitar más todavía el concepto. De esta manera, una vez se ha identificado que una iniciativa o plataforma es de crowdsourcing, se puede también identificar a qué tipo del mismo pertenece.

Por último, se ha establecido de manera formal la relación exacta entre la inteligencia colectiva y el crowdsourcing. Utilizando varios sistemas de etiquetado social, que es un tipo de plataforma que pertenece claramente al ámbito de la inteligencia colectiva, se ha tratado de identificar los elementos de la inteligencia colectiva propuestos por Malone et al. (2009) y los elementos del crowdsourcing propuestos por Estellés-Arolas y González-Ladrón-de-Guevara (2012). De esta manera se ha podido demostrar que toda iniciativa de crowdsourcing es un caso de inteligencia colectiva, trabaje de forma coordinada la multitud o no, pero que no toda iniciativa o plataforma de inteligencia colectiva, como es el caso de los sistemas de mercado social, es un ejemplo de crowdsourcing.

Es cierto que para demostrar esta relación entre la inteligencia colectiva y el crowdsourcing, solo se ha utilizado una herramienta, que ha sido analizada en profundidad. Esto supone una limitación al presente estudio, lo que no invalida la conclusión alcanzada. A lo largo del desarrollo del capítulo anterior, se pudo ver como en general, los cuatro elementos de la inteligencia colectiva se dan también en el crowdsourcing, con la peculiaridad de que en el crowdsourcing se particularizan estas características.

También es importante destacar la versatilidad de las herramientas Web 2.0 en su relación con el crowdsourcing. Aunque bien es cierto que la mayoría de herramientas Web 2.0 no son herramientas de crowdsourcing *per se*, es completamente válido y correcto afirmar que muchas de esas mismas herramientas pueden utilizarse para una iniciativa de crowdsourcing.

Como ejemplo de herramienta Web 2.0, esta tesis se ha centrado en los sistemas de etiquetado social. En este sentido, se ha profundizado en el uso que los usuarios hacen de las etiquetas que describen el contenido marcado y en las características más relevantes de las mismas. En cuanto al tipo de etiqueta, implícita o explícita, la presente tesis demuestra que ambos tipos son utilizados en una proporción similar (45% y 55%). Aunque las etiquetas implícitas son intelectualmente más valoradas porque añaden nueva información no contenida en el texto etiquetado (Farooq et al., 2007), las explícitas resultan ser igual de útiles.

La recogida de datos para el estudio de las etiquetas ha permitido identificar que un 9,2% de los recursos no estaban etiquetados y que un 7% eran recursos fuera de línea. Estos dos datos representan dos problemas comunes que deberían solventar los sistemas de etiquetado social. Por un lado, el contenido sin etiquetar será contenido menos visitado al no ser accesible mediante el pivot browsing basado en etiquetas, solo en usuarios. Por otro lado, el porcentaje de recursos fuera de línea hace que el contenido de estos sistemas de etiquetado social pierdan valor, por lo tanto sería necesario avisar a los usuarios de que los enlaces que poseen ya no son válidos, para que los usuarios los eliminen si lo estiman oportuno.

En base a las distintas revisiones de la literatura realizadas y al estudio de los sistemas de etiquetado social, se puede afirmar que el crowdsourcing, que tal y como hoy lo conocemos es producto de Internet y más concretamente de la Web 2.0, es un fenómeno que tiene la capacidad de afectar a casi cualquier área: negocios, medicina, investigación, educación, ayuda humanitaria, gestión de catástrofes, etc. De esta manera, su implicación en las vidas de las personas, que viven dentro de un mundo donde se encuentran hiperconectadas, tenderá a ser mayor cada vez. Este hecho se puede ver en que cada vez más empresas y organizaciones hacen uso del mismo.

### **7.3. Líneas de trabajo futuras**

Debido a que el crowdsourcing puede ser aplicado a multitud de campos (a través de las microtarefas, la innovación abierta o cualquiera de sus modalidades) las posibles líneas de trabajo futuras son numerosas.

Un campo importante para trabajar, es la aplicación del crowdsourcing en el ámbito educativo superior. El crowdsourcing, bien a través de la realización de pequeñas tareas, o bien desde los crowdcontests, permite a estudiantes de niveles superiores enfrentarse a problemas de la vida real: diseñar un logo para una empresa, resolver un problema real planteado por una compañía, traducir un fragmento de texto para una persona que lo solicita a través de alguna plataforma como Amazon Mechanical Turk (la plataforma para proponer/realizar microtarefas por excelencia), etc. Es importante averiguar el efecto real que tendría la sustitución de ejercicios académicos cuyo único objetivo es la calificación, por otros ejercicios, que sin dejar de utilizarse con fines académicos, tienen un objetivo distinto y que además se corresponde con la vida real.

En base al trabajo realizado para la elaboración de esta tesis, aparecen dos líneas principales por las que se puede continuar investigando.

### **7.3.1. Relación entre el etiquetado y el crowdsourcing**

Aunque se ha utilizado el ejemplo de los sistemas de etiquetado social para plasmar que no son ejemplos *per se* de crowdsourcing, sí es cierto que pueden ser utilizados con este fin. De esta manera, una de las posibles líneas de investigación es el análisis de cómo se puede utilizar este tipo de sistemas, o incluso únicamente el etiquetado, en iniciativas de crowdsourcing.

En este sentido, existe ya un proyecto de escritura avanzado en el que se ha utilizado una iniciativa de etiquetado basada en el crowdtagging (el etiquetado por parte de la multitud de Internet a cambio de alguna recompensa). En este proyecto, una empresa de muebles ha participado lanzando desde su web una iniciativa en la que los usuarios debían etiquetar con 3 etiquetas dos modelos de muebles. La hipótesis principal que se plantea es si el crowdtagging puede ser válido y suficiente para que las empresas conozcan la opinión que tiene la multitud sobre sus productos.

### **7.3.2. Bases teóricas del crowdsourcing**

Tras los primeros pasos para elaborar una definición y una tipología integradoras, se ha posibilitado la diferenciación clara entre la inteligencia colectiva y el crowdsourcing. Sin embargo, quedan todavía algunos términos afines al crowdsourcing cuya relación debe ser clarificada, ya que todavía suscitan diferentes opiniones entre los investigadores. Algunos de estos términos afines son la co-creación, la innovación de usuario o el modelo de desarrollo del software abierto.

De nuevo, en este área existe un proyecto de escritura en proceso que, basándose en los elementos que definen el crowdsourcing, delimita la relación existente entre éste y algunos de los términos antes mencionados.

## **7.4. Conclusión final**

La aparición de Internet, ha llevado al desarrollo de nuevas aplicaciones que han permitido una nueva forma de comunicación entre las personas. Nueva forma de comunicación que ha propiciado la aparición de distintos procesos y formas de trabajo, de entre las cuales, el crowdsourcing, despunta actualmente.

La razón de este despunte no es otro que la posibilidad de que empresas, individuos, organizaciones de cualquier tipo o incluso organismos institucionales, puedan aprovecharse del poder y capacidad de la inteligencia colectiva. Una inteligencia colectiva que, aunque existía previamente, Internet permite expresar al máximo.

Es cierto que el crowdsourcing plantea en algunas de sus manifestaciones algunos problemas, como los que surgen del uso del crowdsourcing creativo -o crowdcontest-, sin embargo, su potencial es innegable.

**CAPÍTULO 8 -**

# **Bibliografía general**

## 8.1. Bibliografía general

- Aliakbarian, S., Rahimabadi, A.M., Sadeghi, P.H. and Mirsatari, N.S. (2006) Neighbor Definition in P2P Networks. In: Proceedings of 2006 International Conference on Communications, Circuits and Systems. (Guilin, 2007) 1562-1565.
- Alonso Arevalo, J. (2009). Gestores de referencias sociales: la información científica en el entorno 2.0. Universo Abierto. Recuperado el 23 de diciembre de 2012, de <http://www.universoabierto.com/2562/gestores-de-referencias-sociales/>
- Alonso, O. and Lease, M. (2011) Crowdsourcing 101: Putting the WSDM of Crowds to Work for You. In: Proceedings of the fourth ACM international conference on Web search and data mining, WSDM '11 (ACM, New York, 2011) 1-2.
- Andriole, S.J. (2010) Business impact of Web 2.0 technologies. Communications of the ACM, 53(12): 67-79.
- Angus, E., Thelwall, M. and Stuart, D. (2008) General patterns of tag usage among university groups in Flickr. Online Information Review 2008, 32(1), 89-101.
- Bar-Ilan, J., Zhitomirsky-Geffet, M., Miller, Y. and Shoham, S. (2010) The effects of background information and social interaction on image tagging. Journal of the American Society for Information Science and Technology, 61, 940–951.
- Bateman, S., Muller, M. J., and Freyne, J. (2009) Personalized retrieval in social bookmarking. In: Proceedings of the ACM 2009 International Conference on Supporting Group Work, pp. 91-94, Sanibel Island, Florida, USA. ACM.
- Bederson, B. B. and Quinn, A.J. (2011) Web workers Unite! Addressing Challenges of Online Laborers. In: Proceedings of the 2011 annual conference extended abstracts on Human Factors in Computing Systems, CHI '11 (Vancouver, 2011).
- Benbunan-Fich, R. and Koufaris, M. (2008) Motivations and Contribution Behavior in Social Bookmarking Systems: An Empirical Investigation. Electronic Markets, 18(2): 150-160.
- Bernstein, M.S., Tan, D., Smith, G., Czerwinski, M. and Horvitz, E. (2010) Personalization via friendsourcing. ACM Transactions on Computer-Human Interaction, 17(2): 1-28.

- Bischoff, K., Firan, C. S., Nejdl, W. and Paiu, R. (2008) Can all tags be used for search?. In: Proceedings of the 17th ACM conference on Information and knowledge management, CIKM '08; New York, USA:ACM; pp. 193–202.
- Bonabeau, E. (2009) Decisions 2.0: The power of collective intelligence. MIT Sloan Management Review, 50(2): 45–52.
- Borrego, Á. and Fry, J. (2012) Measuring researchers' use of scholarly information through social bookmarking data: A case study of BibSonomy. Journal of Information Science, 38(3), 297–308.
- Boydell, O. and Smyth, B. (2007) From social bookmarking to social summarization: an experiment in community-based summary generation. In: Proceedings of the 12th International Conference on Intelligent User Interfaces, IUI '07. New York, USA:ACM; pp. 42–51.
- Brabham D. C. (2009) Crowdsourcing the public participation process for planning project. Planning Theory, 8(3), 242-262.
- Brabham D. C. (2012) Crowdsourcing: A model for leveraging online communities. In: A. Delwiche & J. Henderson (Eds.), Handbook of Participatory Culture. The Routledge.
- Brabham, D. C. (2008) Crowdsourcing as a Model for Problem Solving: An Introduction and Cases. Convergence: The International Journal of Research into New Media Technologies, 14(1), 75-90.
- Brabham, D. C. (2008) Moving the crowd at iStockphoto: The composition of the crowd and motivations for participation in a crowdsourcing application. First Monday, 13(6).
- Brabham, D.C. (2010) Moving the crowd at Threadless, Information. Communication & Society, 13(8), 1122-1145.
- Buecheler, T., Sieg, J.H., Fuchsli, R.M. and Pfeifer, R. (2010) Crowdsourcing, Open Innovation and Collective Intelligence in the Scientific Method: A Research Agenda and Operational Framework. In: H. Fellerman et al (eds), Artificial Life XII. Proceedings of the Twelfth International Conference on the Synthesis and Simulation of Living Systems, Odense, Denmark, 19-23 August 2010, 679-686.



- Burger-Helmchen, T. and Penin, J. (2010) The limits of crowdsourcing inventive activities: What do transaction cost theory and the evolutionary theories of the firm teach us? In: Workshop on Open Source Innovation, Strasbourg, France.
- Cattuto, C. (2006). Semiotic dynamics in online social communities. *The European Physical Journal C - Particles and Fields*, 46, 33-37.
- Cernea, D. A., Del Moral, M. E., and Labra Gayo, J. E. (2008) SOAF: Semantic Indexing System Based on Collaborative Tagging. *Interdisciplinary Journal of E-Learning and Learning Objects*, 4, 137-150.
- Chanal, V. and Caron-Fasan, M.L. (2008) How to invent a new business model based on crowdsourcing: The crowdspirit ® case. In: EURAM (Ljubjana, Slovenia, 2008).
- Chesbrough, H. (2003). *Open Innovation: The new Imperative for Creating and Profit from Technology*. Boston: Harvard Business School Press.
- Chesbrough, H. W. (2003) The Era of Open Innovation. *Sloan Management Review*, 44 (3): 35-41.
- Codina, L. (1997) Una propuesta de metodología para el diseño de bases de datos documentales (Parte II). *El profesional de la información*, 6(12), 20-26.
- Colás Bravo, P. (2003). Internet y aprendizaje en la sociedad del conocimiento. *Comunicar*, 20, 31-35.
- Cormode, G. and Krishnamurthy, B. (2008) Key differences between Web 1.0 and Web 2.0. *First Monday*, 13 (6).
- Corney, J. R., Torres-Sánchez, C., Jagadeesan, P., Lynn, A. and Regli, W. (2010) Outsourcing labour to the cloud. *Journal of Innovation and Sustainable Development*, 4(4), 294-313.
- Cosma, G. and Joy, M. (2008) Towards a Definition of Source-Code Plagiarism. *IEEE Transactions on Education*, 51(2), 195-200.
- Dahlander, L., and Gann, D. M. (2010). How open is innovation?. *Research Policy*, 39(6), 699-709.
- Del Moral, M. E. and Cernea, D. A. (2006) Wikis, Folksonomías y Webquests: trabajo colaborativo a través de Objetos de Aprendizaje. In: *Proceedings of III Simposio*

Pluridisciplinar sobre Diseño, Evaluación y Descripción de Contenidos Educativos Reutilizables (SPDECE06) Oviedo, 2006.

- Delgado-Rodríguez, M., Sillero-Arenas, M., GálvezVargas, R. (1991) Metaanálisis en epidemiología (Primera parte): características generales. *Gaceta sanitaria*, 5(27), 265-272.
- Delgado, M. (2010) *Revisión sistemática de estudios: Metaanálisis*. Barcelona:Signo.
- Delicious' Blog (2011) What's next for Delicious. Recuperado el 16 de marzo de 2011, de <http://blog.delicious.com/blog/2010/12/whats-next-for-delicious.html>
- Denyer, D., Tranfield, D. and Van Aken, J.E. (2008) Developing design propositions through research synthesis. *Organization Studies*, 29(3), 393-413.
- Diigo Help (2006) Diigo is about Social Annotation. Recuperado el 31 de diciembre de 2009, de <http://www.diigo.com/help/about>
- Ding, Y., Jacob, E. K., Zhang, Z., Foo, S., Yan, E., George, N. L. and Guo, L. (2009) Perspectives on social tagging. *Journal of the American Society for Information Science and Technology*, 60: 2388–2401.
- DiPalantino, D. and Vojnovic, M. (2009) Crowdsourcing and all-pay auctions. In: *Proceedings of the 10th ACM conference on Electronic commerce, EC '09 (2009)*; pp. 119–128.
- Doan, A., Ramakrishnan, R. and Halevy, A.Y. (2011) Crowdsourcing systems on the World-Wide Web. *Communications of the ACM*, 54(4), 86-96.
- Dye, J. (2006). Folksonomy: A game of high-tech (and high-stakes) tag. *EContent*, 29(3).
- ECMT - European Commission for Mobility and Transport (2011) Door-to-Door in a click. Recuperado el 15 de julio de 2011, de [http://ec.europa.eu/transport/its/multimodal-planners/index\\_en.htm](http://ec.europa.eu/transport/its/multimodal-planners/index_en.htm)
- Egger, M., Smith, G.D. and Altman, D. (2001) *Systematic reviews in health care. Meta-analysis in context*. London:BMJ Books.
- Eisterlehner, F., Hotho, A. and Jäschke, R. (2009) editors. *ECML PKDD Discovery Challenge 2009 (DC09)*, volume 497 of CEUR-WS.org, Sept. 2009.
- Emory, M. C., (2007) Changing paradigms: managed learning environments and Web 2.0. *Campus-Wide Information Systems*, 24 (3), 152-161.

- Estellés-Arolas, E. (2012) Situación del crowdsourcing en España. Crowdsourcing Blog. Recuperado el 10 de abril de 2013, de <http://www.crowdsourcing-blog.org>
- Estellés-Arolas, E. and González Ladrón-de-Guevara, F. (2012) Towards an integrated crowdsourcing definition. *Journal of Information Science*, 38(2), 189-200.
- EuropaPress (2011) Google explica como se comportan los usuarios de YouTube. Recuperado el 12 de diciembre de 2012, de <http://www.europapress.es/portaltic/internet/noticia-google-explica-son-usuarios-youtube-20110617153125.html>
- Farooq, U., Kannampallil, T.G., Song, Y., Ganoë, C.H., Carroll, J.M. and Giles, L. (2007) Evaluating tagging behavior in social bookmarking systems: metrics and design heuristics. In: *Proceedings of the 2007 international ACM conference on Supporting group work*. New York, NY, USA: ACM; 2007. p. 351–360.
- Farooq, U., Zhang, S., and Carroll, J. (2009) Sensemaking of scholarly literature through tagging. *CHI 2009 Sensemaking Workshop*, April 4–9, 2009, Boston, MA, USA.
- FBI - Federal Bureau of Investigation (2011) Cryptanalysts: Help Break the Code. Recuperado el 15 de julio de 2011, de [http://www.fbi.gov/news/stories/2011/march/cryptanalysis\\_032111](http://www.fbi.gov/news/stories/2011/march/cryptanalysis_032111)
- Flickr (2012) About Interestingness. Recuperado el 12 de diciembre de 2012, de <http://www.flickr.com/explore/interesting/>
- Fu, W. T., Kannampallil, T., Kang, R., and He, J. (2010). Semantic imitation in social tagging. *ACM Trans. Comput.-Hum. Interact.*, 17 (3), 1-37.
- Furnas, G. W., Landauer, T. K., Gomez, L. M., and Dumais, S. T. (1987) The vocabulary problem in human-system communication. *Commun. ACM*, 30(11).
- Geerts, S. (2009) *Discovering crowdsourcing: theory, classification and directions for use*. Tesis de máster. Technische Universiteit Eindhoven, Netherlands.
- Geiger, D., Seedorf, S. and Schader, M. (2011) Managing the Crowd: Towards a Taxonomy of Crowdsourcing Processes. In: *Proceedings of the Seventeenth Americas Conference on Information Systems*, Detroit, Michigan August 4th-7th 2011.

- Georgi, S. and Jung, R. (2012) Collective Intelligence Model: How to Describe Collective Intelligence. In: J. Altmann, U. Baumöl, & B. J. Krämer (Eds.), *Advances in Collective Intelligence 2011, Advances in Intelligent and Soft Computing*. 113: 53–64. Springer Berlin:Heidelberg.
- Ghafele, R., Gibert, B., DiGiammarino, P. (2011) How to improve patent quality by using crowdsourcing. *Innovation management*. Recuperado el 12 de diciembre de 2012, de <http://www.innovationmanagement.se/2011/09/29/howto-improve-patent-quality-by-using-crowd-sourcing/>
- Giudice, K. D. (2010) Crowdsourcing credibility: The impact of audience feedback on Web page credibility. In: *Proceedings of the 73rd ASIS&T Annual Meeting on Navigating Streams in an Information Ecosystem, ASIS&T '10*, 47(1), 1-9.
- Golder, S. A. and Huberman, B. A. (2005) *The Structure of Collaborative Tagging Systems*. HP Labs technical report, 2005.
- Golder, S. A. and Huberman, B. A. (2006) Usage Patterns of Collaborative Tagging Systems. *Journal of Information Science*, 32(2), 198-208.
- González Navarro, M. (2009). Los nuevos entornos educativos: desafíos cognitivos para una inteligencia colectiva. *Comunicar*, 33; 141-148.
- Gregg, D. G. (2010) Designing for collective intelligence. *Communications of the ACM*, 53(4), 134–138.
- Grier, D. A. (2011) Not for All Markets. *Computer*, 44(5), 6-8.
- Grinnell, R.M., Unrau, Y.A. and Williams, M. (2005) *The Qualitative Research Approach*. In: Grinnell, R.M. & Unrau, Y.A. (eds) *Social work research and evaluation. Quantitative and qualitative approaches (7th ed)*. Oxford: University Press.
- Hammond, T., Hannay, T., Lund, B. and Scott, J. (2005) Social Bookmarking Tools (I). *D-Lib Magazine*, 11(04).
- Heckner, M., Heilemann, M. and Wolff, C. (2009) Personal information management vs. resource sharing: Towards a model of information behaviour in social tagging systems. In: *Proceedings of Int'l AAAI Conference on Weblogs and Social Media (ICWSM)*; San Jose, CA, USA; May 2009.

- Heer, J., and Bostok, M. (2010) Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In: Proceedings of the 28th international conference on Human factors in computing systems, CHI'10 (ACM, New York, 2010) 203-212.
- Heizer, J. and Render, B. (2008) Operations Management, 9th edition. Pearson/Prentice Hall.
- Heizer, J., Render, B. (2010) Principles of operations management. Prentice Hall.
- Hernández Sampieri, R., Fernández Collado, C., and Baptista Lucio, P. (2007). Fundamentos de metodología de la investigación. Mcgraw-Hill.
- Heylighen, F. (1999) Collective Intelligence and its Implementation on the Web: algorithms to develop a collective mental map. Computational & Mathematical Organization Theory, 5(3), 253-280.
- Heymann, P. and Garcia-Molina, H. (2011) Turkalytics: analytics for human computation. In: Proceedings of the 20th international conference on World wide web, WWW '11 (ACM, New York, 2011) 477-486.
- Heymann, P., Koutrika, G. and Molina, H.G. (2008). Can social bookmarking improve web search? In: Proceedings of the International Conference on Web Search and Web Data Mining. WSDM '08. New York, NY, USA: ACM; 2008. p. 195–206.
- Hirth, M., Hoßfeld, T., and Tran-Gia, P. (2010) Cheat-detection mechanisms for crowdsourcing. Technical report, University of Würzburg.
- Hotho, A., Jauchke, R., Schmitz, C. and Stumme, G. (2006) BibSonomy: A Social Bookmark and Publication Sharing System. In: Proceedings of the Conceptual Structures Tool Interoperability Workshop at the 14th International Conference on Conceptual Structures; pp. 87-102.
- Howe, J. (2006) Crowdsourcing: A definition. Recuperado el 27 de julio de 2011, de [http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing\\_a.html](http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html)
- Howe, J. (2006) The rise of crowdsourcing. Wired, 14 (6).
- Howe, J. (2008) Crowdsourcing: How the Power of the Crowd is Driving the Future of Business. Business Books: Great Britain.

- Huberman, B.A., Romero, D.M. and Wu, F. (2009) Crowdsourcing, Attention and Productivity. *Journal of Information Science*, 35(6), 758–765.
- Illig, J., Hotho, A., Jäschke, R., and Stumme, G. (2009) A Comparison of content-based Tag Recommendations in Folksonomy Systems. In: *Postproceedings of the International Conference on Knowledge Processing in Practice (KPP 2007)*.
- Inc. (2010) Using crowdsourcing to control Inventory. Recuperado el 18 de agosto de 2011, de <http://www.inc.com/magazine/20100201/using-crowdsourcing-to-control-inventory.html>
- Jäschke, R., Marinho, L., Hotho, A., Schmidt-Thieme, L., and Stumme, G. (2007). Tag recommendations in folksonomies. In: *Proceedings of Knowledge Discovery in Databases: PKDD 2007*, pages 506-514. Springer-Verlag, Berlin, Heidelberg.
- Jäschke, R., Marinho, L., Hotho, A., Schmidt-Thieme, L., and Stumme, G. (2008). Tag recommendations in social bookmarking systems. *AI Communications*, 21(4), 231-247.
- Kazai, G. (2011) In Search of Quality in crowdsourcing for Search Engine Evaluation. In: *Proceedings of the 33rd European conference on Advances in Information retrieval (Springer-Verlag, Berlin/Heidelberg, 2011)*. [Lecture Notes in Computer Science 6611, 165-176].
- Kickstarter (2012) 2011: The Stats. Recuperado el 10 de abril de 2013, de <http://www.kickstarter.com/blog/2011-the-stats>
- Kittur, A., Chi, E.H. and Suh, B. (2008) Crowdsourcing user studies with Mechanical Turk. In: *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems. CHI '08*. New York, NY, USA: ACM; p. 453–456.
- Kleeman, F., Voss, G.G. and Rieder, K. (2008) Un(der)paid Innovators: The Commercial Utilization of Consumer Work through Crowdsourcing. *Science, Technology and Innovation Studies*, 4(1), 5-26.
- Kolay, S. and Dasdan, A. (2009) The value of socially tagged urls for a search engine. In: *Proceedings of the 18th international conference on World wide web*, pp. 1203-1204, New York, NY, USA. ACM.

- Körner, C., Benz, D., Hotho, A., Strohmaier, M. and Stumme, G. (2010) Stop thinking, start tagging: tag semantics emerge from collaborative verbosity. In: Proceedings of the 19th international conference on World wide web; pp. 521–530.
- Koutrika, G., Effendi, F.A., Gyöngyi, Z., Heymann, P. and Molina H.G. (2008) Combating spam in tagging systems: An evaluation. *ACM Trans. Web*, 2(4), 1-34.
- La Vecchia, G. and Cisternino, A. (2010) Collaborative workforce, business process crowdsourcing as an alternative of BPO. In: Proceedings of First Enterprise crowdsourcing Workshop in conjunction with ICWE 2010. Berlin/Heidelberg:Springer-Verlag; 2010. pp. 425-430.
- Lakhani, K. R., Jeppesen, L. B., Lohse, P. A., & Panetta, J. A. (2007). The Value of Openness in Scientific Problem Solving. Division of Research, Harvard Business School.
- Lamere, P. (2008) Social Tagging and Music Information Retrieval. *Journal of New Music Research*, 37(2), 101–114.
- Lánzos (2012) Lánzos en cifras. Recuperado el 10 de abril de 2013, de <http://www.lanzanos.com/blog/entry/26/Lanzanosencifras/>
- Leimeister, J. (2010) Collective Intelligence. *Business & Information Systems Engineering*, 2(4), 245–248.
- Lerman, K. (2007) User Participation in Social Media: Digg Study. In: Proceedings of the 2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology - Workshops, WI-IATW '07. Washington, DC, USA: IEEE Computer Society, pp. 255–258.
- Lerman, K., Plangprasopchok, A. and Wong, C. (2007) Personalizing Image Search Results on Flickr. In: Proceedings of AAI workshop on Intelligent Techniques for Information Personalization. Vancouver, Canada, AAI Press.
- Lévy, P. (2001) Collective intelligence. *Reading digital culture*, 4: 253.
- Ling, P. (2010) An Empirical Study of Social Capital in Participation in Online crowdsourcing. *Computer*; 7(9), 1-4.

- Lipczak, M. and Milios, E. (2010) The impact of resource title on tags in collaborative tagging systems. In: Proceedings of the 21st ACM conference on Hypertext and hypermedia, pages 179-188, New York, NY, USA. ACM.
- Liu, E., and Porter, T. (2010) Culture and KM in China. VINE, 40(3/4), 326-333.
- Liu, Y., Kumar, R., and Lim, K. (2008) Taggers versus Linkers: Comparing Tags and Anchor Text of Web Pages. UC Berkeley: School of Information. Report 2008-020. Recuperado el 23 de abril de 2013, de <http://escholarship.org/uc/item/8b40q59k>
- Lykourantzou, I., Vergados, D. J. and Loumos, V. (2009) Collective intelligence system engineering. In: Proceedings of the International Conference on Management of Emergent Digital Ecosystems; New York:ACM; Article N°. 20.
- Maharana, B., Majhi, S. and Bhue, S. (2010) Social Bookmarking: Web 2.0 Tool for Content Sharing and Learning. In: Proceedings of the 7th Convention PLANNER, Tezpur University, Assam February.
- Malone, T. W., Laubacher, R. and Dellarocas, C. N. (2009) Harnessing Crowds: Mapping the Genome of Collective Intelligence. MIT Sloan; Research Paper No. 4732-09.
- Malone, T. W., Laubacher, R., and Dellarocas, C. N. (2010) The collective intelligence genome. MIT Sloan Management Review; 51 (3): 21-31.
- Marinho, L. B., Nanopoulos A., Schmidt-Thieme, L., Jäschke R., Hotho, A. Stumme, G. and Symeonidis, P. (2011) Social Tagging Recommender Systems. In Francesco Ricci and Lior Rokach and Bracha Shapira and Paul B. Kantor (Eds.), Recommender systems handbook, pp. 615-644, Springer.
- Marlow, C., Naaman, M., Boyd, D., and Davis, M. (2006) Ht06, tagging paper, taxonomy, flickr, academic article, toread. In: Proceedings of the seventeenth conference on Hypertext and hypermedia, pages 31-40, New York, NY, USA. ACM.
- Maslow, A.H. (1943) A Theory of Human Motivation. Psychological review, 50.
- Mason, R. and Rennie, F. (2008) E-Learning and Social Networking Handbook: Resources for Higher Education. Routledge:NY.



- Mathes, A. (2004) Folksonomies - Cooperative Classification and Communication Through Shared Metadata. Computer Mediated Communication - LIS590CMC. Graduate School of Library and Information Science, University of Illinois Urbana-Champaign.
- Mazzola, D. and Distefano, A. (2010) Crowdsourcing and the participacion process for problem solving: the case of BP. In: VII Conference of the Italian Chapter of AIS. Information technology and Innovation trend in Organization. (Napoles, Italy, 2010).
- McLoughlin, C. and Lee, M. J. (2007). Social software and participatory learning: Pedagogical choices with technology affordances in the Web 2.0 era. In Proceedings ASCILITE, Singapore 2007.
- Melenhorst, M. and Van Setten, M. (2007). Usefulness of tags in providing access to large information systems. In: Proceedings of the IEEE Professional Communication Conference, 2007. IPPC 2007. pp. 1-9.
- Metal 2.0 (2011a) METAL 2.0 CROWDSOURCING - Web 2.0, redes sociales y crowdsourcing aplicados al sector del metal. Recuperado el 1 de marzo de 2013, de <http://www.metal20.org/proyecto>
- Metal 2.0 (2011b) Video "El crowdsourcing desde un punto de vista científico". Recuperado el 1 de marzo de 2013, de [http://youtu.be/khBTMi2\\_4XA](http://youtu.be/khBTMi2_4XA)
- Millen, D. R., Yang, M., Whittaker, S., and Feinberg, J. (2007) Social bookmarking and exploratory search. In: Proceedings of the ECSCW 2007 (pp. 21-40). Springer London.
- Millen, D., Feinberg, J. and Kerr, B. (2005) Social Bookmarking in the enterprise. Queue, 3(9), 28-35.
- Monge, S., Ovelar, R. and Azpeitia, I. (2008) Repository 2.0: Social Dynamics to Support Community Building in Learning Object Repositories. Interdisciplinary Journal of E-Learning and Learning Objects, 4, 191-204.
- Moral Toranzo, F. (2009) Internet como marco de comunicación e interacción social. Comunicar, 32, 231-237.
- Murty, P., Paulini, M. and Maher, M.L. (2010) Collective Intelligence and Design Thinking. In: Proceedings of the Design Thinking Research Symposium, DTRS'10; Sydney, Australia. 2010.

- Murugesan, S. (2007) Understanding Web 2.0. *IT Professional*, 9(4), 34-41.
- Nations, D. (n.d.). Social Bookmarking – What is Social Bookmarking?. Recuperado el 21 de diciembre de 2010, de <http://webtrends.about.com/od/socialbookmarking101/p/aboutsocialtags.htm>
- O'Reilly, T. (2007) What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. *Communications & Strategies*, 1, 17.
- O'Reilly, T. (2005) What is Web 2.0?. Recuperado el 12 de diciembre de 2012, de <http://oreilly.com/web2/archive/what-is-web-20.html>
- Oliveira, B., Calado, P., Pinto, H.S. (2008) Automatic Tag Suggestion Based on Resource Contents. In: Proceedings of the 16th international conference on Knowledge Engineering, 255-264. Springer-Verlag, Berlin/Heidelberg.
- Oliveira, F., Ramos, I., Santos, L. (2010) Definition of a crowdsourcing Innovation Service for the European SMEs. In: Daniel F. et al. (eds.) *Current Trends in Web Engineering* (Springer, Berlin/Heidelberg, 2010), pp. 412-416.
- Oomen, J. and Aroyo, L. (2011). Crowdsourcing in the cultural heritage domain: opportunities and challenges. In: Proceedings of the 5th International Conference on Communities and Technologies (pp. 138–149). New York, NY, USA: ACM. doi:10.1145/2103354.2103373
- OSI (n.d.) The Open Source Definition. Recuperado el 25 de noviembre de 2011, de <http://opensource.org/docs/osd>
- Parvanta, C., Roth, Y. and Keller, H. (2013) Crowdsourcing 101 A Few Basics to Make You the Leader of the Pack. *Health promotion practice*, 14(2), 163-167.
- Pénin, J. (2008) More open than open innovation? Rethinking the concept of openness in innovation studies. Working papers of BETA, Bureau d'Economie Théorique et Appliquée, UDS, Estrasburgo.
- Petitti, D. B. (2000) *Meta-analysis. Decision Analysis and Cost-Effectiveness Analysis*. Oxford University Press:New York.
- Pinto Molina, M., Alonso Berrocal, J. L., Cordon García, J. A., Fernández Marcial, V., García Figuerola, C., García Marco, J., ... and Doucet, A. V. (2004) Análisis cualitativo de

la visibilidad de la investigación de las universidades españolas a través de sus páginas web. *Revista española de documentación científica*, 27(3), 345-370.

- Pisano, G. P. and Verganti, R. (2008) Which kind of collaboration is right for you. *Harvard business review*, 86(12), 78-86.
- PlanB (2011) El Plan Ballantine's de Carlos Jean. Recuperado el 1 de marzo de 2011, de <http://prensa.elplanb.tv/>
- Poetz, M. K. and Schreier, M. (2012) The Value of Crowdsourcing: Can Users Really Compete with Professionals in Generating New Product Ideas? *Journal of Product Innovation Management*, 29(2), 245–256.
- Porta, M., House, B., Buckley, L., and Blitz, A. (2008) Value 2.0: eight new rules for creating and capturing value from innovative technologies. *Strategy & Leadership*, 36(4), 10-18.
- Preece, J. and Shneiderman, B. (2009) The reader-to-leader framework: Motivating technology-mediated social participation. *AIS Transactions on Human-Computer Interaction*, 1(1): 13–32.
- Reichel, M. & others (2006). Embodied, Constructionist Learning: Social Tagging and Folksonomies in E-Learning Environments. In: *Proceedings of mICTE*, 2006.
- Reichwald, R. and Piller, F.T. (2006) Interaktive Wertschöpfung. *Open Innovation, Individualisierung und neue Formen der Arbeitsteilung*. Wiesbaden:Gabler.
- Reinhardt, M., Frieß, R., Groh, G., Wiener, M., and Amberg, M (2010) Web 2.0 driven Open Innovation Networks - A social Network Approach to Support the Innovation Context within Companies. In: Schumann, M., Kolbe, L., Breiner, M., Frerichs, A. (Eds) *Proceedings of the Multikonferenz Wirtschaftsinformatik (MKWI)*, p. 1177-1190, Gottingen.
- Ribiere, V. M., and Tuggle, F. D. (2010) Fostering innovation with KM 2.0. *VINE*, 40(1).
- Robu, V., Halpin, H., and Shepherd, H. (2009). Emergence of consensus and shared vocabularies in collaborative tagging systems. *ACM Trans*, 3(4), 1-34.

- Rosen, Y. and Rimor, R. (2009) Using a Collaborative Database to Enhance Students' Knowledge Construction. *Interdisciplinary Journal of E-Learning and Learning Objects* 5, 187-196.
- Sawant, N., Li, J. and Wang, J.Z. (2011) Automatic image semantic interpretation using social action and tagging data. *Multimedia Tools Appl.*, 51(1), 213–246.
- Schenk, E. and Guittard, C. (2009) Crowdsourcing: What can be Outsourced to the Crowd, and Why? Technical Report (2009) Available from: <http://halshs.archives-ouvertes.fr/halshs-00439256/>
- Schenk, E. and Guittard, C. (2009) Crowdsourcing: What can be Outsourced to the Crowd, and Why? Technical Report. REcuperado el 12 de diciembre de 2012, de <http://halshs.archives-ouvertes.fr/halshs-00439256/>
- Schenk, E. and Guittard, C. (2009) Le crowdsourcing: modalités et raisons d'un recours à la foule. Recuperado el 12 de 11 de noviembre de 2010, de <http://marsouin.infini.fr/ocs2/index.php/frontieres-numeriques-brest2009/frontieres-numeriques-brest2009/paper/viewFile/60/>
- Schenk, E. and Guittard, C. (2011) Towards a characterization of crowdsourcing practices. *Journal of innovation economics*, 1(7), 93-107.
- Schmitz, C., Hotho, A., Jäschke, R., and Stumme, G. (2006). Mining association rules in folksonomies. In: *Proceedings of Data Science and Classification* (pp. 261-270). Springer Berlin Heidelberg.
- Shepitsen, A., Gemmell, J., Mobasher, B., and Burke, R. (2008) Personalized recommendation in social tagging systems using hierarchical clustering. In: *Proceedings of the 2008 ACM conference on Recommender systems, RecSys '08*, 259-266. New York, NY, USA: ACM.
- Siddique H. (2011) Mob rule: Iceland crowdsources its next constitution. *The Guardian*. Recuperado el 1 de diciembre de 2011, de <http://www.guardian.co.uk/world/2011/jun/09/iceland-crowdsourcing-constitution-facebook>

- Singh, G., Hawkins, L. and Whymark, G. (2007) An integral model of collaborative knowledge building. *Interdisciplinary Journal of E-Learning and Learning Objects*, 3, 85-104.
- Sloane, P. (2011) The brave new world of open innovation. *Strategic Direction*, 27(5), 3-4.
- Smith, G. (2004) Atomiq: Folksonomy: social classification. Recuperado el 3 de agosto de 2011, de [http://atomiq.org/archives/2004/08/folksonomy\\_social\\_classification.html](http://atomiq.org/archives/2004/08/folksonomy_social_classification.html)
- Stewart O., Huerta J.M. and Sader M. (2009) Designing crowdsourcing community for the enterprise. In: *Proceedings of the ACM SIGKDD Workshop on Human Computation, HCOMP '09* (ACM, New York, 2009) 50-53.
- Strohmaier, M., Körner, C. and Kern, R. (2010) Why do users tag? detecting users' motivation for tagging in social tagging systems. In: *Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM2010)*, Washington, DC, USA.
- Subramanya, S. B. and Liu, H. (2008) Socialtagger collaborative tagging for blogs in the long tail. In: *Proceedings of the 2008 ACM Workshop on Search in Social Media (Napa Valley, California, USA, October 30 30, 2008)*. SSM '08. ACM, New York, NY, 19-26.
- Superbowl (2011) Crash the SuperBowl. Recuperado el 18 de agosto de 2011, de <http://www.crashthesuperbowl.com/>
- Surowiecki, J. (2005) *The wisdom of crowds*. New York: Anchor Books.
- Sutherlin, G. (2013). A voice in the crowd: Broader implications for crowdsourcing translation during crisis. *Journal of Information Science*.
- Tapscott, D. and Williams, A. D. (2010) *Wikinomics: How Mass Collaboration Changes Everything*. Penguin Group: USA.
- Tatarkiewicz W. (1980) *History of Six Ideas: An Essay in Aesthetics*. Ed. Springer.
- Taylor, A. G. (2003) *The Organization of Information*. Library and Information Science Text Series. Ed. Libraries Unlimited.
- Trant, J. (2009) Studying social tagging and folksonomy: A review and framework. *Journal of Digital Information*, 10 (1): 1-42.
- Veal A. J. (2002) *Leisure and tourism policy and planning*. Ed. CABI Publishing.

- von Hippel, E. and Katz, R. (2002) Shifting Innovation to Users via Toolkits. *Management Science*, 48(7), 821-833.
- Vukovic, M. (2009) Crowdsourcing for enterprises. In: *Proceedings of the 2009 Congress on Services – I*, IEEE Computer Society (Washington, DC, USA 2009). 686-692.
- Vukovic, M. and Bartolini, C. (2010) Towards a Research Agenda for Enterprise crowdsourcing. In: M. Tiziana and S. Bernhard (eds) *Leveraging Applications of Formal Methods, Verification, and Validation* (Springer, Berlin/Heidelberg, 2010) 425-434 [Lecture Notes in Computer Science 6415].
- Vukovic, M., Kumara, S., Greenspan, O. (2010) Ubiquitous crowdsourcing. In: *Proceedings of the 12th ACM intl conf*, pp. 523-526.
- Vukovic, M., Laredo, J. and Rajagopal, S. (2010) Challenges and experiences in deploying enterprise crowdsourcing service. In: *Proceedings of the 10th international conference on Web engineering*. Springer-Verlag Berlin, Heidelberg.
- Vukovic, M., Mariana, L. and Laredo, J. (2009) PeopleCloud for the Globally Integrated Enterprise. In: D. Asit et al. (eds) *Service-Oriented Computing*. (Springer-Verlag, Berlin/Heidelberg, 2009).
- Vukovic, M.; Bartolini, C. (2010) Crowd-driven processes: state of the art and research challenges. In: Maglio, P.; Weske, M.; Yang, J.; Fantinato, M. *Service-oriented computing*. Lecture notes in computer science, 2010b, v. 6470, p. 733.
- Wash, R. and Rader, E. (2007) Public bookmarks and private benefits: An analysis of incentives in social computing. *American Society for Information Science and Technology (ASIS&T) Annual Meeting*, Milwaukee, WI.
- Wechsler, D. (1971) Intelligence: definition, theory and the IQ. In: Cancro, R. (Ed.) *Intelligence: genetic and environmental influences*. Grune Straton. New York, 50-55.
- Wexler, M. N. (2011) Reconfiguring the sociology of the crowd: exploring crowdsourcing. *International Journal of Sociology and Social Policy*, 31(1), 6 - 20.
- Whitla P. (2009) Crowdsourcing and Its Application in Marketing. *Contemporary Management Research*, 5(1), 15-28.

- Wikipedia (n.d.) Social bookmarking. Recuperado el 28 de diciembre de 2009, de [http://en.wikipedia.org/wiki/Social\\_Bookmarking](http://en.wikipedia.org/wiki/Social_Bookmarking)
- Wikipedia (n.d.) Crowdsourcing. Recuperado el 15 de agosto de 2011, de <http://en.wikipedia.org/wiki/Crowdsourcing>
- Wikipedia (n.d.) List of crowdsourcing projects. Recuperado el 11 de febrero de 2011, de [http://en.wikipedia.org/wiki/List\\_of\\_crowdsourcing\\_projects](http://en.wikipedia.org/wiki/List_of_crowdsourcing_projects)
- Wolfson, S. M. and Lease, M. (2011) Look before you leap: Legal pitfalls of crowdsourcing. In: Proceedings of the American Society for Information Science and Technology 2011; 48(1): 1–10.
- Yang J., Adamic L.A. and Ackerman M.S. (2008) Crowdsourcing and knowledge sharing: strategic user behaviour on taskcn. In: Proceedings of the 9th ACM conference on Electronic commerce (ACM, New York, 2008) 246-255.
- Yeung, C., Gibbins, N. and Shadbolt, N. (2009) Contextualising tags in collaborative tagging systems. In: Proceedings of the 20th ACM conference on Hypertext and hypermedia. New York, NY, USA: ACM; p. 251–260.
- Yuen, M.C., King, I. and Leung, K.S. (2011) A Survey of Crowdsourcing Systems. In: Proceedings of the IEEE Third International Conference on Social Computing. IEEE; p. 766–773.
- Zhang, N., Zhang, Y., and Tang, J. (2009). A tag recommendation system for folksonomy. In King, I., Li, J. Z., Xue, G. R., Tang, J., King, I., Li, J. Z., Xue, G. R., and Tang, J., (eds.), CIKM-SWSM, pages 9-16. ACM.
- Zubiaga, A.; Martínez, R. & Fresno, V. (2009) Getting the most out of social annotations for web page classification. In: Proceedings of the 9th ACM Symposium on Document Engineering, pp. 74-83, Munich, Germany. AC.