



Universidad de San Carlos de Guatemala  
Facultad de Ingeniería  
Escuela de Ingeniería en Ciencias y Sistemas

**USO DE *CROWDSOURCING* PARA LA ALIMENTACIÓN DE CATÁLOGOS DE  
BIBLIOTECAS CON BAJO PRESUPUESTO**

**Carlos Alberto Aguilar Sandoval**  
Asesorado por el Ing. José Julio Pineda

Guatemala, noviembre de 2017

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERÍA

**USO DE *CROWDSOURCING* PARA LA ALIMENTACIÓN DE CATÁLOGOS DE  
BIBLIOTECAS CON BAJO PRESUPUESTO**

TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA  
FACULTAD DE INGENIERÍA

POR

**CARLOS ALBERTO AGUILAR SANDOVAL**  
ASESORADO POR EL ING. JOSÉ JULIO PINEDA

AL CONFERÍRSELE EL TÍTULO DE

**INGENIERO EN CIENCIAS Y SISTEMAS**

GUATEMALA, NOVIEMBRE DE 2017

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA  
FACULTAD DE INGENIERÍA



**NÓMINA DE JUNTA DIRECTIVA**

DECANO	Ing. Pedro Antonio Aguilar Polanco
VOCAL I	Ing. Ángel Roberto Sic García
VOCAL II	Ing. Pablo Christian de León Rodríguez
VOCAL III	Ing. José Milton de León Bran
VOCAL IV	Br. Jurgén Andoni Ramírez Ramírez
VOCAL V	Br. Oscar Humberto Galicia Nuñez
SECRETARIA	Inga. Lesbia Magalí Herrera López

**TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO**

DECANO	Ing. Murphy Olympto Paiz Recinos
EXAMINADOR	Ing. Marlon Antonio Pérez Türk
EXAMINADORA	Inga. Floriza Felipa Ávila Pesquera de Medinilla
EXAMINADORA	Inga. Sonia Yolanda Castañeda de De Paz
SECRETARIA	Inga. Marcia Ivónne Véliz Vargas

## **HONORABLE TRIBUNAL EXAMINADOR**

En cumplimiento con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

### **USO DE *CROWDSOURCING* PARA LA ALIMENTACIÓN DE CATÁLOGOS DE BIBLIOTECAS CON BAJO PRESUPUESTO**

Tema que me fuera asignado por la Dirección de la Escuela de Ingeniería en Ciencias y Sistemas, con fecha 20 de abril de 2016.

**Carlos Alberto Aguilar Sandoval**

Guatemala, 20 de julio de 2016

Ingeniero  
Carlos Azurdia  
Revisor de Trabajo de Graduación  
Escuela de Ciencias y Sistemas  
Facultad de Ingeniería  
Universidad de San Carlos de Guatemala

Estimado Ingeniero Azurdia,

Por este medio hago de su conocimiento que he revisado el trabajo de graduación del estudiante CARLOS ALBERTO AGUILAR SANDOVAL, titulado: "USO DE CROWDSOURCING PARA LA ALIMENTACIÓN DE CATÁLOGOS DE BIBLIOTECAS CON BAJO PRESUPUESTO". A mi criterio, dicho trabajo cumple con los objetivos propuestos para su desarrollo, según el protocolo.

Para cualquier comunicación relacionada a este proceso, quedo al pendiente.

Atentamente,



---

José Julio Pineda Chinchilla  
Ingeniero en Ciencias y Sistemas  
Asesor de trabajo de graduación  
Colegiado 10340  
[julio@glifos.com](mailto:julio@glifos.com)

José Julio Pineda Chinchilla  
Ingeniero en Ciencias y Sistemas  
Colegiado No. 10,340



Universidad San Carlos de Guatemala  
Facultad de Ingeniería  
Escuela de Ingeniería en Ciencias y Sistemas

Guatemala, 3 de Agosto de 2016

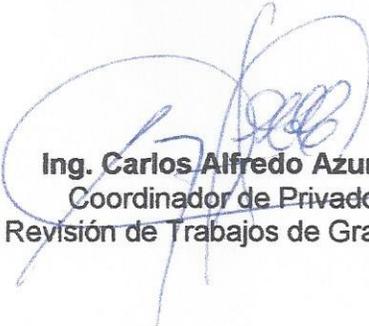
Ingeniero  
**Marlon Antonio Pérez Türk**  
Director de la Escuela de Ingeniería  
En Ciencias y Sistemas

Respetable Ingeniero Pérez:

Por este medio hago de su conocimiento que he revisado el trabajo de graduación del estudiante **CARLOS ALBERTO AGUILAR SANDOVAL** con carné **199516424**, titulado: **“USO DE CROWDSOURCING PARA LA ALIMENTACIÓN DE CATÁLOGOS DE BIBLIOTECAS CON BAJO PRESUPUESTO”**, y a mi criterio el mismo cumple con los objetivos propuestos para su desarrollo, según el protocolo.

Al agradecer su atención a la presente, aprovecho la oportunidad para suscribirme,

Atentamente,

  
**Ing. Carlos Alfredo Azurdia**  
Coordinador de Privados  
y Revisión de Trabajos de Graduación



UNIVERSIDAD DE SAN CARLOS  
DE GUATEMALA



FACULTAD DE INGENIERÍA  
ESCUELA DE INGENIERÍA EN  
CIENCIAS Y SISTEMAS  
TEL: 24767644

*El Director de la Escuela de Ingeniería en Ciencias y Sistemas de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer el dictamen del asesor con el visto bueno del revisor y del Licenciado en Letras, del trabajo de graduación **“USO DE CROWDSOURCING PARA LA ALIMENTACIÓN DE CATÁLOGOS DE BIBLIOTECAS CON BAJO PRESUPUESTO”**, realizado por el estudiante CARLOS ALBERTO AGUILAR SANDOVAL, aprueba el presente trabajo y solicita la autorización del mismo.*

**“ID Y ENSEÑAD A TODOS”**

*Ing. Marlon Antonio Pérez Türk*

**Director**

**Escuela de Ingeniería en Ciencias y Sistemas**



Guatemala, 17 de noviembre de 2017

Universidad de San Carlos  
de Guatemala

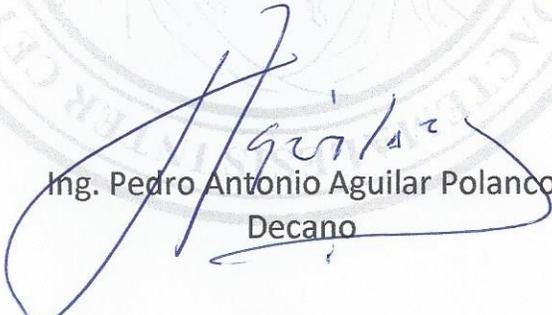


Facultad de Ingeniería  
Decanato

DTG. 582.2017

El Decano de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Ingeniería en Ciencias y Sistemas, al Trabajo de Graduación titulado: **USO DE CROWDSOURCING PARA LA ALIMENTACIÓN DE CATÁLOGOS DE BIBLIOTECAS CON BAJO PRESUPUESTO**, presentado por el estudiante universitario: **Carlos Alberto Aguilar Sandoval**, y después de haber culminado las revisiones previas bajo la responsabilidad de las instancias correspondientes, autoriza la impresión del mismo.

IMPRÍMASE:



Ing. Pedro Antonio Aguilar Polanco  
Decano

Guatemala, noviembre de 2017

/gdech



## **ACTO QUE DEDICO A:**

<b>Mis padres</b>	Por iniciarme en el camino de la educación.
<b>Mis hermanos</b>	Apoyo y mis amigos de toda la vida.
<b>Mi esposa</b>	Por compartir su vida conmigo y por su ayuda.
<b>Mis abuelos</b>	Por haberme inculcado valores.
<b>Mis tíos y primos</b>	Por brindarme siempre su cariño y ayuda.
<b>Mi Padrino Amílcar y mi Tía Hilda</b>	A mi Padrino, por ser la persona que creyó en mí. A mi Tía, por ser incondicional apoyo para todos.
<b>Mis amigos</b>	Por los momentos alegres y por su compañerismo.

## **AGRADECIMIENTOS A:**

**Facultad de Ingeniería**

La entidad que me formó académicamente.

**Ing. Carlos Azurdia**

Por su paciencia y dedicación con los estudiantes que están presentando su tesis.

**Ing. Julio Pineda**

Por toda su asesoría y ayuda en la elaboración de este trabajo.

**A toda mi familia, mis amigos y su persistencia**

Por su empeño en alentarme a terminar mi carrera.

## ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES .....	V
GLOSARIO.....	IX
RESUMEN.....	XIII
OBJETIVOS .....	XV
INTRODUCCIÓN.....	XVII
1. CATALOGACIÓN BIBLIOGRÁFICA.....	1
1.1. Organización de la información.....	1
1.1.1. Componentes .....	2
1.1.2. Funciones .....	4
1.1.3. Usos .....	5
1.2. Catálogos .....	6
1.2.1. Definiciones y funciones .....	6
1.2.2. Componentes .....	10
1.3. Catalogación.....	11
1.3.1. Tipos de catalogación .....	11
1.3.1.1. Catalogación descriptiva .....	11
1.3.1.2. Análisis de temas (catalogación de temas) .....	12
1.3.1.3. Control de autoridades .....	12
1.3.1.4. Catalogación cooperativa y por copia .....	13
1.3.2. Formatos y estándares .....	13
2. <i>CROWDSOURCING</i> .....	15
2.1. Conceptos, teorías e historia .....	15

2.2.	¿Qué no se considera <i>crowdsourcing</i> ? .....	17
2.2.1.	<i>Open source</i> .....	17
2.2.2.	Producción paritaria de bienes comunes .....	18
2.2.3.	Estudios de mercado y compromiso de marca	18
2.3.	Bases del <i>crowdsourcing</i> .....	19
2.3.1.	Internet y la cultura de participación .....	19
2.3.2.	Resolución de problemas e innovación .....	21
2.3.3.	Inteligencia colectiva y la sabiduría de las masas.....	24
2.3.4.	<i>Microtasks</i> de <i>crowdsourcing</i> .....	26
2.3.5.	<i>Crowdfunding</i> .....	26
2.4.	Organización del <i>crowdsourcing</i> .....	27
2.4.1.	Descubrimiento de conocimiento y administración .....	28
2.4.2.	Búsqueda por difusión .....	28
2.4.3.	Producción creativa vetada por los pares .....	29
2.4.4.	Labor de inteligencia humana distribuida.....	30
2.4.5.	Divisiones disciplinarias.....	32
2.5.	Problemática del <i>crowdsourcing</i> .....	35
2.5.1.	Moviendo a la multitud.....	35
2.6.	Problemas legales.....	37
2.6.1.	Libertad de expresión y disensión .....	38
2.6.2.	Propiedad intelectual y derechos de copia .....	40
2.6.3.	Prácticas de negocio injustas .....	41
2.7.	Problemas éticos.....	41
2.7.1.	Derechos laborales.....	42
2.7.2.	Eficiencia.....	42
2.7.3.	Tiranía del acceso .....	43
2.8.	Usos y ejemplos .....	44

2.8.1.	En astronomía.....	44
2.8.2.	En botánica.....	45
2.8.3.	En entomología.....	45
2.8.4.	En monitoreo del clima.....	46
2.8.5.	En geología.....	46
2.8.6.	En procesos de digitalización.....	46
2.8.7.	En aprendizaje.....	46
3.	<i>CAPTCHA</i> .....	47
3.1.	Conceptos.....	47
3.2.	Aplicaciones.....	48
3.2.1.	Registro de correos electrónicos.....	48
3.2.2.	Motores de búsqueda.....	49
3.2.3.	Servicios de compartición de archivos.....	49
3.2.4.	Formularios para compartir links.....	49
3.3.	Tipos de <i>CAPTCHA</i> .....	49
3.3.1.	<i>MATCHA</i> .....	49
3.3.2.	<i>PIX</i> .....	50
3.4.	Ejemplos.....	51
3.4.1.	<i>RECAPTCHA</i> .....	51
3.4.2.	<i>CAPTCHA</i> de reconocimiento de imagen.....	51
3.4.2.1.	<i>CAPTCHA</i> de nombramiento de imagen.....	52
3.4.2.2.	<i>CAPTCHA</i> de distinción de imágenes:.....	53
3.4.2.3.	<i>CAPTCHA</i> de identificación de anomalía.....	53
3.4.3.	<i>CAPTCHA</i> de identificación de rompecabezas.....	54

4.	BREVE RESEÑA DEL ESTADO ACTUAL DE LOS CATÁLOGOS BIBLIOGRÁFICOS EN LAS BIBLIOTECAS DE GUATEMALA. ....	55
5.	DESARROLLO DEL MÓDULO WEB QUE USARÁ <i>CROWDSOURCING</i> PARA LA ALIMENTACIÓN DE CATÁLOGOS DE BIBLIOTECAS. ....	59
5.1.	Descripción .....	59
5.2.	Funcionamiento.....	59
5.3.	Alcances y límites .....	60
5.4.	Diagramas.....	60
5.4.1.	Diagrama de clases .....	60
5.4.1.1.	Método generateCAPTCHA() .....	60
5.4.1.2.	Método checkCAPTCHA() .....	61
5.4.2.	Diagrama Entidad Relación .....	62
5.4.3.	Diagrama de casos de uso .....	63
5.4.3.1.	Descripción de los casos de uso... ..	63
6.	PRUEBAS REALIZADAS .....	65
6.1.	Implementación de prototipo de módulo web .....	65
6.2.	Pruebas de uso del módulo web implementado .....	69
7.	RESULTADOS .....	73
7.1.	Implementación de prototipo de módulo web .....	73
7.2.	Pruebas de uso del módulo web implementado .....	73
	CONCLUSIONES .....	81
	RECOMENDACIONES .....	83
	BIBLIOGRAFÍA .....	85

# ÍNDICE DE ILUSTRACIONES

## FIGURAS

1.	Objetivos de un catálogo, según Charles A. Cutter .....	8
2.	Medios de un catálogo, según Charles A. Cutter .....	9
3.	<i>CAPTCHA</i> de nombramiento de imagen.....	52
4.	<i>CAPTCHA</i> de identificación de anomalía.....	53
5.	<i>CAPTCHA</i> de identificación de rompecabezas .....	54
6.	Diagrama de clases .....	61
7.	Diagrama de Entidad Relación.....	62
8.	Diagrama de casos de uso .....	63
9.	Pantalla inicial del módulo web <i>CAPTCHA</i> .....	66
10.	Pantalla para la resolución del <i>CAPTCHA (PIX)</i> .....	67
11.	Pantalla de procesamiento y despliegue de mensaje.....	68
12.	<i>Pop-up</i> de encuesta de SurveyMonkey.....	70
13.	Encuesta de usabilidad y disposición de colaboración con el proyecto de <i>crowdsourcing</i> .....	71

## TABLAS

I.	Tipos de <i>crowdsourcing</i> .....	30
II.	Estadísticas de tipos de catálogo en bibliotecas del país .....	57
III.	Resultados de generación de <i>CAPTCHAS</i> .....	74
IV.	Totales de <i>CAPTCHAS</i> generados, procesados y resueltos .....	75
V.	Porcentajes de <i>CAPTCHAS</i> procesados y resueltos.....	76
VI.	Respuestas a la pregunta No.1 de la encuesta .....	77
VII.	Respuestas a la pregunta No.2 de la encuesta .....	78
VIII.	Respuestas a la pregunta No.3 de la encuesta .....	79



## LISTA DE SÍMBOLOS

<b>Símbolo</b>	<b>Significado</b>
@	En
%	Porcentaje
BOOL	Tipo de datos: booleano
FA1	Flujo alterno 1
INT	Tipo de datos: entero
VA1	Validación alterna 1
VARCHAR	Tipo de datos: cadena de texto



## **GLOSARIO**

**Administración de  
comunidad en línea**

En el contexto de una organización, proyecto o marca, es el trabajo profesional de construir y mantener comunidades en línea alrededor de una organización, proyecto o marca.

**Computación distribuida**

Es la descomposición de un problema de computación a lo largo de una red de computadoras individuales o nodos para que efectúen el procesamiento necesario para resolver el problema.

**Computación humana**

Es la aplicación de los principios de computación distribuida a una red de agentes o nodos humanos, o nodos para procesar información que requiere inteligencia humana.

**Comunidad en línea**

Es un grupo, conjunto o red de personas que están interconectadas o se relacionan dentro de una red computacional.

**Contenido generado por  
el usuario**

En el contexto de Internet, es el texto o contenido multimedia que ha sido producido y distribuido en línea por los usuarios de Internet.

<b>Crowd</b>	Llamada también una masa o multitud, es una comunidad en línea de individuos involucrados en una actividad de <i>crowdsourcing</i> .
<b>Crowdfunding</b>	Es el uso de una comunidad en línea para traer una idea o producto al mercado, a través del financiamiento colectivo de muchos donadores en la comunidad.
<b>Crowdslapping</b>	Es la resistencia del <i>crowd</i> , dentro de una actividad de <i>crowdsourcing</i> o un espacio comunal, hacia un <i>crowdsourcer</i> o hacia la misma actividad de <i>crowdsourcing</i> .
<b>Crowdsourcer</b>	Es una organización que opera una aplicación de <i>crowdsourcing</i> .  Es el uso de un modelo online, de resolución distribuida de problemas y de producción, para potenciar la inteligencia colectiva de comunidades en línea que puedan contribuir para servir a metas organizacionales específicas.
<b>Innovación abierta</b>	Es la inclusión estratégica de <i>stakeholders</i> en el proceso de innovación de una organización.

<b>Inteligencia colectiva</b>	Es el fenómeno donde un grupo de personas que trabajan conjuntamente, o que tomadas en consideración como un conjunto, se vuelven inteligentes colectivamente, como una entidad.
<b>Marginalidad</b>	Es el efecto de aquellos individuos en los márgenes de un dominio de resolución de problemas, que brinda heurísticas diversas en la resolución de problemas y experiencias para lidiar con un problema.
<b><i>Microtasks</i></b>	Son pequeñas tareas que pueden ser realizadas fácilmente, que usualmente son trabajos que contribuyen a un trabajo más grande que ha sido descompuesto y distribuido entre muchos trabajadores.
<b>Sabiduría de las masas</b>	Es el fenómeno donde grupos de individuos sobresalen sobre individuos que trabajan solos para resolver un problema.
<b><i>Stakeholder</i></b>	Persona con un interés o preocupación con algo relacionado a una actividad, organización o proceso.
<b><i>Solver</i></b>	Es el individuo que trabaja en proponer o propone una solución para la problemática definida por la actividad de <i>crowdsourcing</i> .

**Tasks (HITs)**

Son tareas fáciles de realizar por un ser humano, pero significativamente más difíciles de realizar para una computadora.

## RESUMEN

Este trabajo de graduación inicia por definir las bases de la catalogación bibliográfica, los tipos que existen de ésta y los mecanismos para aprovecharla. Cuáles son los elementos que son importantes para los usuarios al momento de usar un catálogo y los formatos y estándares más usados.

Se presenta la definición de *crowdsourcing*, la forma en que se organiza para la resolución de problemas, los tipos de *crowdsourcing* que existen, la problemática que enfrenta.

Se ahonda un poco más en un tipo de herramienta de *crowdsourcing*, que es el *CAPTCHA*, los tipos y variantes que existen de éste, y como puede ser usado como herramienta de *crowdsourcing*.

Luego se presentan algunas estadísticas relacionadas a la problemática de las bibliotecas en nuestro país y la ausencia de catálogos que las personas puedan usar para efectuar consultas.

Se procede con el desarrollo del prototipo del módulo que integra los conceptos anteriores, con el propósito de generar catálogos para aquellas bibliotecas que no tiene opción de poder generar su propio catálogo.

El prototipo del módulo es puesto a la disposición de un grupo de personas para su uso y evaluación de resultados.



## OBJETIVOS

### General

Aplicar el uso del *crowdsourcing* para evaluar la alimentación de catálogos electrónicos de bibliotecas.

### Específicos

1. Investigar los fundamentos del *crowdsourcing* para determinar de qué forma pueden ser usados para resolver un problema específico del país.
2. Generar un prototipo de módulo web que permita a los usuarios de dicho módulo, el ingreso ordenado de información bibliográfica.
3. Validar que el empleo del prototipo del módulo web desarrollado, genere registros bibliográficos que puedan ser usados en sistemas integrados de bibliotecas.



## INTRODUCCIÓN

En la actualidad, las bibliotecas manejan contenidos de diversos tipos: materiales impresos, visuales, auditivos, digitales en todas sus variedades, los que están a disposición de los usuarios que los quieran consultar, y que sepan solicitarlos.

Para esta tarea de búsqueda, es común el uso de catálogos electrónicos, catálogos provistos por sistemas integrados de bibliotecas, en los cuales los usuarios, a través del ingreso de términos de búsqueda en un formulario, pueden encontrar los materiales deseados.

Es necesario que los materiales físicos como los digitales pasen por un proceso de catalogación que permita la extracción de *metadata* de dichos materiales, para ser ingresados en un sistema integrado de biblioteca, y a su vez, al catálogo electrónico de búsqueda.

La catalogación bibliográfica normalmente es efectuada por gente capacitada para este propósito. Pero es común encontrar situaciones en que existen bibliotecas que poseen colecciones de materiales, pero no poseen los recursos para poder contratar los servicios de los profesionales que los puedan catalogar, esto presenta como resultado, una colección sub utilizada, pues ante la ausencia de *metadata*, la búsqueda a través de un catálogo electrónico es imposible.

En muchos casos, los mecanismos de catalogación que permiten la automatización de estos materiales, no siempre son funcionales, pues se pueden dar los casos en que se poseen materiales antiguos o de producción local, que no tienen identificadores únicos que permitan su búsqueda a través de estos

mecanismos. En otras ocasiones son publicaciones propias de la institución, que no aparecerían en otros catálogos.

Mecanismos alternos para la catalogación de materiales implican el uso de técnicas de inteligencia artificial, las que actualmente no podrían identificar fácilmente los elementos que conforman la *metadata* a capturar de los materiales de una biblioteca.

Ahora bien, existe una base de usuarios de computadoras, que las usan en redes internas o en Internet, volviéndose un recurso valioso del cual se puede sacar provecho, pues poseen una inteligencia común que les permite el análisis de contenido a través de su contexto.

El uso del *crowdsourcing* para la obtención de información en otros ámbitos ha sido aplicable, bajo este mismo enfoque se puede lograr la alimentación de catálogos de bibliotecas, que si bien no tendrían los niveles de catalogación que algunos reglamentos consideran mínimos, si servirían para efectos de búsqueda dentro de un catálogo electrónico.

Dentro de este trabajo, se explorará la viabilidad del uso del *crowdsourcing* para la alimentación de catálogos electrónicos, para bibliotecas con bajo presupuesto.

# 1. CATALOGACIÓN BIBLIOGRÁFICA

## 1.1. Organización de la información

La catalogación bibliográfica es parte de un campo más grande llamado organización de la información (el cual también es llamado control bibliográfico). Es el proceso de describir recursos de información y de proveer nombre, título, y temas (tópicos) a estas descripciones, dando como resultado registros que sirven como representaciones de ítems de información registrada para recursos que están lógicamente ordenados. Adicionalmente hay quienes consideran que la organización de la información también nos permite guardar para la posteridad, copias de todos los tipos de trabajos que resultan de los esfuerzos humanos<sup>1</sup>.

Siempre que alguien trata de mantener una lista o un archivo de referencias, artículos, libros y otros tipos de materiales que contengan información de un tema particular o por un autor particular, ha practicado un control bibliográfico sobre una pequeña parte del universo de la información disponible. Para que dicho proyecto sea funcional, es necesario decidir qué elementos de los datos se deben de registrar acerca de cada libro, revista, artículo, entre otros. Se puede registrar autores, títulos, palabras claves, resúmenes, contenidos, temas, idiomas, de dichos recursos, estos datos eventualmente serán almacenados, manipulados y consultados. Conforme este archivo crece, su almacenamiento, manipulación y consulta se vuelve más compleja, de ahí se vuelve evidente que se necesitan herramientas y habilidades para el mantenimiento de dicho archivo.

Si se considera el universo de todo el conocimiento (literatura, música, arte, etc.), hay cierta cantidad de dicho conocimiento que ha sido registrada de alguna

---

<sup>1</sup> JOUDREY, Daniel N, TAYLOR, Arlene G y MILLER, David P. *Introduction to cataloging and classification*. Santa Barbara, California : Libraries Unlimited, 2015. p.3-4.

manera, ya sea escrita, impresa, pintada. Esta cierta cantidad de información normalmente se le llama universo bibliográfico. Sobre este universo se puede ejercer control, a través de herramientas bibliográficas, en las cuales cada elemento o pieza de conocimiento es representada por un conjunto de metadatos. Un conjunto de metadatos puede ser referido como un registro bibliográfico, un registro de metadatos o un registro sustituto.

Las herramientas bibliográficas incluyen, bibliografías, índices, catálogos, ayudas de búsqueda, registros, bases de datos bibliográficas y motores de búsqueda. Si bien es cierto, que eventualmente, los recursos pueden estar disponibles en línea, requiere cierto tipo de control, y alguna manera de tener información preliminar (título, autor, etc.) antes de tratar de navegar entre millones de estos recursos. La creación de un número de estándares de metadatos es el resultado de esta necesidad por la información elemental relacionada con estos recursos.

Dado que se ha estado hablando de metadatos, sería recomendable acordar una definición respecto a este concepto. La propuesta por Siegel y Madnick (1991), parece bastante concreta, donde definen *metadata* como el significado, contenido, organización o propósito de los datos<sup>2</sup>.

### **1.1.1. Componentes**

Los componentes usados para la organización de la información pueden ser variados, pero algunos de los más utilizados son:

- Bibliografías

---

<sup>2</sup> SIEGEL, Michael D y MADNICK, Stuart E. *A metadata approach to resolving semantic conflicts*. Cambridge, Mass. : Sloan School of Management, Massachusetts Institute of Technology, 1991. p. 3.

- Índices
- Catálogos
- ILS

Daniel Joudrey expresa que “a los usuarios de biblioteca, e incluso a los bibliotecarios, tradicionalmente se les ha informado que las bibliografías, índices, catálogos, y demás, son herramientas diferentes una de otras, en realidad son parte de la misma organización de la información, pero que se han desarrollado separadamente, en distintos formatos, por múltiples razones, por ejemplo uno de los factores ha sido el económico pues no es común que los catálogos de bibliotecas pudieran proveer información en todos los niveles. Por ejemplo, podrían proveer información acerca de los libros pero no de sus capítulos, o de las revistas pero no de sus artículos, proveer una información integral resultaba más caro, por lo cual el desarrollo de otros elementos como índices, bibliografías, y otros se vio como una alternativa por aquellas instituciones que solo podrían brindar información en un solo nivel específico”<sup>3</sup>.

Debido a este tipo de limitantes, anteriormente, el uso de índices era común, debido a que los catálogos estaban circunscritos a mostrar información solo de los materiales que poseían, debido a eso, el enfoque principal de los índices era hacia las publicaciones periódicas, ya que eran publicaciones que podía o no poseer la biblioteca, pero que era de utilidad para el usuario.

Luego, las bibliografías cumplían con la función de describir materiales, de forma similar a los catálogos, pero sin mostrar la información pertinente a ciertos aspectos del material, como datos de publicación, ISBN, clasificación.

---

<sup>3</sup> JOUDREY, Daniel N, TAYLOR, Arlene G y MILLER, David P. *Introduction to cataloging and classification*. Santa Barbara, California : Libraries Unlimited, 2015. p.4-5.

Fredson Bowers describe la función básica de la bibliografía como "[proveer] suficientes datos de manera que el lector pueda identificar el libro descrito, entendiendo su impresión y reconociendo su contenido preciso"<sup>4</sup>.

Por último, los ILS, que engloban otras herramientas adicionales a la organización de la información, aunque relacionadas a ella, poseen catálogos que son más completos que los mencionados al inicio, pues al ser unidos a estándares de catalogación, incluyen recursos electrónicos/en línea, y al estar desarrollados para poder ser utilizados desde la web, incorporan fácilmente dentro de sí, todas las herramientas mencionadas previamente.

### **1.1.2. Funciones**

Las funciones de las herramientas de la organización de la información son principalmente 3:

- Identificación
- Recolección
- Evaluación

La función de identificación permite al usuario, que tiene un ítem bibliográfico particular en mente, asociar lo que sabe de ese ítem con los recursos de información gestionados por las herramientas descritas. De esa manera el usuario puede encontrar el ítem que busca, dentro de los recursos que la biblioteca posee, o en otro lugar del cual la biblioteca posee información.

---

<sup>4</sup> BOWERS, Fredson. *Principles of bibliographical description*. Winchester : ST Paul's Bibliographics and Oak Knoll, 1994. p.3-4.

La función de recolección provee los medios por los cuales se puede ubicar en un solo lugar de una herramienta bibliográfica todos los materiales relacionados (o aquellos estrechamente relacionados) a un tema. De esta forma, el registro de un libro de álgebra estará conceptualmente cerca del registro de un libro de aritmética.

La función de evaluación permite al usuario escoger entre los registros que la herramienta posea, aquellos que parecen representar el conocimiento, información o características físicas deseadas. Por ejemplo, un usuario podría escoger entre las múltiples ediciones de Don Quijote de La Mancha, que una biblioteca pueda poseer o referenciar.

### **1.1.3. Usos**

Según la IFLA (*International Federation of Library Associations*), las funciones de las herramientas bibliográficas, son esenciales para el propósito de permitir a los usuarios llevar a cabo ciertas tareas. Dichas tareas las describe como<sup>5</sup>:

- Encontrar entidades que correspondan al criterio de búsqueda del usuario, (i.e., ubicar una sola entidad o un conjunto de entidades en un archivo o base de datos como resultado de una búsqueda usando un atributo o relación de la entidad)

---

<sup>5</sup> INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS, & INSTITUTIONS. Section on Cataloguing. Standing Committee. *Functional requirements for bibliographic records* [en línea]. [http://www.ifla.org/files/assets/cataloguing/frbr/frbr\\_2008.pdf](http://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf) IFLA Study Group on the Functional Requirements for Bibliographic Records, 2009. p.79. [Consulta: 7 de octubre de 2016].

Identificar una entidad (i.e., confirmar que la entidad descrita corresponda a la entidad buscada, o distinguir entre dos o más entidades con características similares)

- Seleccionar una entidad que es apropiada a las necesidades del usuario (i.e., elegir una entidad que cumpla con los requerimientos del usuario con respecto a su contenido, formato físico, etc., o rechazar una entidad si no es apropiada a las necesidades del usuario)
- Solicitar u obtener acceso a la entidad descrita (i.e., solicitar una entidad a través de la compra, préstamo, etc. u obtener acceso a una entidad electrónica a través de una conexión en línea a una computadora remota)

## **1.2. Catálogos**

### **1.2.1. Definiciones y funciones**

Un catálogo es una compilación organizada de *metadata* bibliográfica que representa el fondo bibliográfico de una institución particular y los recursos accesibles en una ubicación particular. La colección de recursos de una biblioteca puede consistir de varios tipos de materiales (libros, publicaciones periódicas, mapas, monedas, grabaciones de sonido, grabados, pinturas, partituras, etc.). Normalmente, la colección representada por el catálogo se encuentra en una o varias ubicaciones de la institución. Aunque, actualmente los catálogos pueden representar el fondo bibliográfico de más de una biblioteca, ya sea como un consorcio de bibliotecas o interconectando los catálogos de las bibliotecas involucradas. Adicionalmente, los catálogos pueden incluir recursos de internet, a los cuales la biblioteca puede proporcionar acceso.

Los catálogos son necesarios siempre que una colección crece demasiado para que el propietario o el usuario sea capaz de recordar y recuperar ítems

específicos. Una pequeña colección privada o la biblioteca de un aula tienen poca necesidad de un catálogo formal, pues el usuario puede recordar cada libro, mapa, grabado, u otros ítems por autor, título, tema, forma, color, posición o estantería particular.

Cuando una colección se vuelve un poco más grande, un arreglo informal, por ejemplo, por temática, puede ser suficiente para proveer acceso a los ítems. Pero cuando la colección crece demasiado para dichas aproximaciones, un registro formal es necesario. Las principales razones para dicho registro formal son para propósitos de recuperación y de inventario. Pues además que al propietario/usuario se le vuelve difícil recordar que existe dentro de dicha colección, también se le dificulta saber que se adquirió, que se perdió y que ha sido reemplazado. Es ahí donde el catálogo sirve como registro de lo que se ha adquirido.

Las funciones de las herramientas bibliográficas mencionadas previamente son extensivas a los catálogos. Estas funciones fueron descritas primordialmente por Charles A. Cutter.<sup>6</sup>

Cutter, primero declaró lo que él creía que debían de ser los objetivos de un catálogo, y después dio su punto de vista de los medios por los cuales los objetivos pueden ser cumplidos. Sus conceptos sirven como base para el entendimiento actual de las funciones de un catálogo, aun cuando en la práctica moderna no son del todo completos.

---

<sup>6</sup> CUTTER, Charles A. *Rules for a printed dictionary catalogue*. Washington : Govt. Print. Off., 1889. p.10-11.

Figura 1. **Objetivos de un catálogo, según Charles A. Cutter**

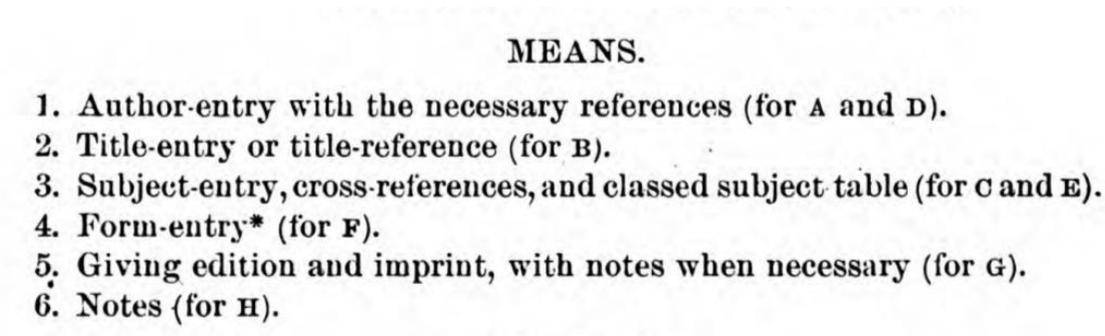
- OBJECTS.**
1. To enable a person to find a book of which either
    - (A) the author
    - (B) the title
    - (C) the subject } is known.
  2. To show what the library has
    - (D) by a given author
    - (E) on a given subject
    - (F) in a given kind of literature.
  3. To assist in the choice of a book
    - (G) as to its edition (bibliographically).
    - (H) as to its character (literary or topical).

Fuente: CUTTER, Charles A. *Rules for a printed dictionary catalogue*. Washington : Govt. Print. Off., 1889. p.10.

En la ilustración anterior, Cutter define que los objetivos de un catálogo son:

- Permitir a la persona encontrar un libro cuando es conocido cualquiera de sus autores, título o temas
- Mostrar lo que la biblioteca posee por autor, tema o tipo de literatura
- Asistir en la elección de un libro por su edición, por su carácter literario o por su carácter temático

Figura 2. **Medios de un catálogo, según Charles A. Cutter**



Fuente: CUTTER, Charles A. *Rules for a printed dictionary catalogue*. Washington : Govt. Print. Off., 1889. p.10.

De igual manera expresó que los medios serían:

- Entradas de autor, con las referencias necesarias
- Entradas de título o referidas a título
- Entradas de temas, referencias cruzadas y tablas clasificadas de temas
- Entradas de forma
- Mención de edición e impresión, con notas (cuando fuera necesario)
- Notas

Las funciones de los catálogos, así como las definió Cutter, permanecieron invariantes hasta 1961, cuando la IFLA las redefinió, y nuevamente en 2009<sup>7</sup>.

---

<sup>7</sup> STATEMENT OF INTERNATIONAL CATALOGUING PRINCIPLES. *IFLA Meetings of Experts on an International Cataloguing Code* [en línea]. [http://www.ifla.org/files/assets/cataloguing/icp/icp\\_2009-en.pdf](http://www.ifla.org/files/assets/cataloguing/icp/icp_2009-en.pdf). IFLA Cataloguing Section and IFLA Meetings of Experts on an International Cataloguing Code, 2009. P. 3-4. [Consulta: 3 de octubre de 2016].

Las funciones del catálogo, como se conceptualizan actualmente, son:

- Encontrar recursos bibliográficos en una colección como el resultado de una búsqueda usando atributos o relaciones de los recursos
- Identificar un recurso bibliográfico
- Seleccionar el recurso bibliográfico que sea apropiado a las necesidades del usuario
- Obtener acceso o acceder a un ítem descrito
- Navegar dentro del catálogo, a través del ordenamiento lógico de datos bibliográficos y de autoridad, y la presentación de las formas claras de hacerlo

### **1.2.2. Componentes**

Básicamente, tres componentes conforman un catálogo completo:

- Un catálogo de acceso público
- Una lista de estantería (o catálogo topográfico)
- Un archivo de autoridades

El catálogo de acceso público es el medio por el cual los usuarios pueden hacer sus consultas sobre el acervo de la biblioteca.

La lista de estantería es la catalogación topográfica (i.e. clasificación) de los materiales.

El archivo de autoridades son los registros de los nombres, títulos, series, encabezamientos de materia que han sido elegidos como las formas estandarizadas a usar como puntos de acceso autorizado.

Los catálogos en línea usualmente contienen estos tres componentes de forma integrada.

### **1.3. Catalogación**

La catalogación es el proceso por medio del cual los catálogos son preparados. Estos procesos usualmente inicia con la catalogación descriptiva, pero continua con el análisis de temas, y ambas fases se interconectan con el proceso de control de autoridades.

La catalogación se hace a través de un proceso de codificación. El resultado de los procesos de catalogación descriptiva, de temas y de autoridad es ingresado en un formato computable, que es compatible con el sistema en línea, en el cual la *metadata* será usada.

#### **1.3.1. Tipos de catalogación**

##### **1.3.1.1. Catalogación descriptiva**

La catalogación descriptiva es la fase del proceso de catalogación que está relacionada con la identificación y descripción de un recurso de información, la grabación de esta información en la forma de un registro de catalogación, la selección de nombres y títulos necesarios para proveer acceso a recurso y el establecimiento de puntos de acceso autorizados para nombres y títulos.

La catalogación descriptiva, describe lo que constituye un recurso de información e identifica las entidades responsables por el contenido artístico e intelectual sin referencia a su clasificación por temas o la asignación de

encabezamientos de materia, ambos de los cuales son relacionados a la catalogación de temas (encabezamientos de materia).

#### **1.3.1.2. Análisis de temas (catalogación de temas)**

Implica determinar qué temas son cubiertos por el contenido intelectual o artístico de una obra. Después que esto ha sido determinado se pueden elegir los encabezamientos de materia y términos de indexamiento apropiados, normalmente de una lista estandarizada. Un archivo de autoridad debe ser consultado en aquellos casos en que existan otros materiales que cubran materias o temas similares/relacionados.

El paso final en el análisis de temas es elegir una clasificación apropiada para el material. La clasificación, normalmente sirve para ubicar un material junto con otros de temática relacionada y también como un mecanismo de ubicación del mismo material.

#### **1.3.1.3. Control de autoridades**

El control de autoridades es el resultado del proceso de mantener la consistencia en la forma verbal usada para representar un punto de acceso y el proceso posterior de mostrar las relaciones entre los nombres, obras y temas. Se logra a través del uso de guías de catalogación, el uso de vocabulario controlado y la referencia a un archivo de autoridad para crear un término autorizado. El archivo de autoridad es la agrupación de registros de las formas de los nombres, títulos y temas elegidos para usar en un catálogo. Cada registro de autoridad en un archivo de autoridad, puede contener una lista de formas variantes que pueden ser usadas como referencias.

#### **1.3.1.4. Catalogación cooperativa y por copia**

No es necesario para cada recurso en cada biblioteca ser catalogado originalmente a través del proceso descrito anteriormente. Debido a que las bibliotecas adquieren copias de muchos de los mismos títulos, o deciden catalogar los mismos recursos de internet, sus catalogadores pueden compartir *metadata*, a través de adaptar a sus propios catálogos, una copia de la catalogación original creada por otra biblioteca, este proceso es llamado Catalogación por copia.

El uso de formatos como MARC (*Machine Readable Cataloging*) permitió que la catalogación por copia se propagara, porque al poseer un estándar, su intercambio a través de medios computacionales se propagó.

#### **1.3.2. Formatos y estándares**

El estándar más usado para el almacenamiento y representación de la información organizada en el proceso de catalogación es el estándar MARC. Si bien, algunos otros estándares son usados en bibliotecas con orientación a contenido en línea, como es el caso de Dublin Core, o estándares usados como una transición entre la complejidad de MARC y la simplicidad de Dublin Core, que sería en este caso MODS.

MARC es el estándar por excelencia, usado y desarrollado por la *Library of Congress* de Estados Unidos de América, alrededor de 1965. Su última versión, y la más usada por bibliotecas, catálogos, ILS es la MARC 21.

Entre los formatos que MARC maneja están:

- MARC 21, que es el formato original del estándar, basado en caracteres, con una estructura de campos de control y campos de datos, que se subdividen a su vez en subcampos, identificados por códigos.
- MARCXML, es una adaptación del formato MARC 21, representado a través de un documento XML

En la última versión de MARC<sup>8</sup>, se pueden contabilizar aproximadamente 249 campos de datos, sin contar los campos de control, que son 7 (incluyendo el Leader).

Dublin Core es un pequeño conjunto de términos usados para describir recursos web ((video, imágenes, páginas web, etc.), así como recursos físicos (libros, cd's, obras de arte, otros). A diferencia de MARC, su conjunto original de elementos de *metadata* es de solo 15 elementos.

MODS por su parte es un estándar intermedio entre la complejidad de MARC y la simplicidad de Dublin Core. Si bien incluye algunos elementos de MARC, no es totalmente compatible con este.

---

<sup>8</sup> OFFICE LIBRARY. *MARC 21 Format for Bibliographic Data - Field List*. [en línea]. <https://www.loc.gov/marc/bibliographic/ecbdlst.html>. [Consulta: 13 de octubre de 2016].

## 2. CROWDSOURCING

### 2.1. Conceptos, teorías e historia

El concepto de *crowdsourcing* es un tanto complicado, pues no existe una definición totalmente consensuada al respecto. Hay diferencias en sus acepciones dependiendo del enfoque y el contexto con el que se evalúe el término.

El término fue acuñado por Jeff Howe en junio del 2006, en una edición de la revista *Wired*<sup>9</sup>, donde describe un nuevo modelo basado en la web, que reúne las soluciones creativas de una red distribuída de individuos en la cual se hace una invitación a participar con propuestas.

Posteriormente, Howe ofrece la siguiente definición: Simplemente definido, *crowdsourcing* representa el acto de una compañía o institución, de trasladar una función antes ejecutada por empleados o por *outsourcing*, a una red (generalmente grande) de gente en la forma de una invitación abierta. Esto puede tomar la forma de producción por pares (*peers*), pero también puede ser ejecutada por individuos. El prerrequisito crucial es el uso de una invitación abierta y la red de trabajadores potenciales<sup>10</sup>

En el 2008, los primeros intentos escolares de definir *crowdsourcing* comenzaron a ser publicados en publicaciones académicas. Estas definiciones

---

<sup>9</sup> HOWE, Jeff. *The Rise of Crowdsourcing*. [en línea]. <http://www.wired.com/2006/06/crowds/> [Consulta: 5 de agosto de 2016].

<sup>10</sup> HOWE, Jeff. *Crowdsourcing: A Definition*. [en línea]. [http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing\\_a.html](http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html). [Consulta: 5 de agosto de 2016].

conflictivas explicaban *crowdsourcing* acorde a los participantes y a las razones para su participación, de acuerdo a las herramientas usadas en los diferentes casos, acorde a las características organizacionales de los casos, o acorde al grado de complejidad o al grado de participación del usuario.

Estos intentos de definir *crowdsourcing* llevaron a una competencia de definiciones: ¿qué podía considerarse como *crowdsourcing*? y ¿qué se excluía?

Una de las aproximaciones más integrantes sucede en un artículo del 2012, en el *Journal of Information Science*, donde Enrique Estellés-Arolas y Fernando González Ladrón de Guevara analizaron la literatura escolar acerca de *crowdsourcing* y encontraron cerca de 40 diferentes interpretaciones acerca de *crowdsourcing*.

Después de un análisis sistemático y validación de estas múltiples definiciones, generaron una definición muy completa acerca de *crowdsourcing*:

*Crowdsourcing* es un tipo de actividad participativa en línea, en la cual un individuo, una institución, una organización no lucrativa o una compañía proponen a un grupo de individuos de conocimiento, heterogeneidad y número variable, vía una invitación abierta, la realización de una tarea. La realización de la tarea, de complejidad y modularidad variable, y en la cual el *crowd* debe participar ofreciendo su trabajo, dinero, conocimiento y/o experiencia, siempre implica un beneficio mutuo. El usuario recibirá la satisfacción de una necesidad dada, sea económica, reconocimiento social, autoestima, o el desarrollo de habilidades individuales, mientras que el convocador (*crowdsourcer*) obtendrá y

utilizará para su beneficio lo que el usuario haya aportado al negocio, cuya forma dependerá del tipo de actividad realizada<sup>11</sup>.

Los puntos claves del *crowdsourcing* son:

- Una organización que tiene una tarea que necesita ser realizada
- Una comunidad que desea hacer la tarea voluntariamente
- Un ambiente en línea que permita que el trabajo tome lugar y que la comunidad interactúe con la organización
- Beneficio mutuo para la organización y la comunidad

## **2.2. ¿Qué no se considera *crowdsourcing*?**

### **2.2.1. *Open source***

No se considera como *crowdsourcing* porque en su diseño no hay una administración *top-down* del proyecto. En principio, los proyectos *open source* están concebidos para ser de administración *bottom-up*, colaboraciones auto organizadas de programadores que trabajan hacia una meta común. Las labores del día a día de la producción *open source* suelen estar regidas por la comunidad y no necesariamente por el proyecto o por el patrocinador del proyecto (aunque en la práctica, grandes proyectos *open source* han empezado a adoptar procesos administrativos más jerárquicos).

---

<sup>11</sup> ESTELLÉS-AROLAS, Enrique y GONZÁLEZ-LADRÓN-DE-GUEVARA, Fernando. *Towards an integrated crowdsourcing definition*. *Journal of Information Science*. SAGE Publications, 2012, Vol. 38, no. 2, p. 189-200.

### **2.2.2. Producción paritaria de bienes comunes**

Bajo este concepto se puede ubicar Wikipedia. Por razones similares al *open source*, este tipo de producción no puede ser considerada como *crowdsourcing*, debido a que no existe una directiva *top-down* de qué artículos necesitan ser escritos, o que contenidos deben de cubrir dichos artículos. El crecimiento de Wikipedia es dirigido solamente por editores que contribuyen con su labor e inteligencia para mejorar el recurso. Y aunque Wikipedia provee de herramientas para ser usadas por la comunidad (*sandbox*, sintaxis simple, etc.), Wikipedia no dirige la producción de conocimiento en el sitio, esto es dirigido por la comunidad, es decir *bottom-up*.

### **2.2.3. Estudios de mercado y compromiso de marca**

El *crowdsourcing* no es lo mismo que una simple votación o campañas de estudios de mercado. Las personas que simplemente están expresando sus opiniones o emitiendo sus votos no están haciendo *crowdsourcing*. Estas actividades son lo mismo que lo realizado por grupos de enfoque, la única diferencia es que estos esfuerzos actualmente toman lugar en escalas más grandes debido al alcance del internet. A diferencia del *open source*, estas actividades incluyen mucho control *top-down* y poca creatividad *bottom-up*. Las organizaciones han reducido las posibles salidas incluso antes que a las personas les sea pedida su colaboración. En simples votaciones o valoraciones, las compañías solo dan una pequeña cantidad de poder de decisión o creatividad, así que no está sucediendo *crowdsourcing*. En este caso, la producción del control creativo reside en la organización, no en los consumidores.

### **2.3. Bases del *crowdsourcing***

Las condiciones necesarias para implementar el *crowdsourcing* son condiciones técnicas y condiciones conceptuales. El internet y otras tecnologías proveen la base sobre la que se fundamentan las aplicaciones del *crowdsourcing*, estas tecnologías también proporcionan lugar a ciertas actitudes y mecanismos para involucramiento en una cultura de participación (condiciones técnicas). Conceptualmente, el *crowdsourcing* puede ser explicado a través de los procesos de resolución de problemas e innovaciones, así como a través del fenómeno de inteligencia colectiva (condiciones conceptuales).

#### **2.3.1. Internet y la cultura de participación**

El internet permite un tipo de pensamiento conectado y creativo. El *crowdsourcing* solo puede existir en línea. Tiziana Terranova escribe que internet es una tecnología ideal para el pensamiento distribuido, porque no es simplemente un medio específico, sino un tipo de implementación activa de una técnica de diseño capaz de lidiar con la apertura que poseen los sistemas.<sup>12</sup>

Otros aspectos del internet, que lo hacen un medio ideal para facilitar la participación creativa, incluyen su velocidad, alcance, flexibilidad temporal, anonimidad, interactividad, bajas barreras de entrada, y la habilidad para llevar cualquier otra forma de contenido. Respecto a su velocidad y alcance, el internet es una plataforma de comunicación instantánea en la cual los mensajes, y por consiguiente el intercambio de ideas, puede viajar tan rápido que el medio virtualmente elimina el problema del tiempo, y de esa forma acelera el desarrollo creativo. Además, tiene un alcance global si las personas tienen acceso a la

---

<sup>12</sup> TERRANOVA, Tiziana. *Network Culture: Politics for the Information Age*. London : Pluto Press, 2004. p. 3, p. 62.

tecnología. Esto quiere decir que la comunicación puede tomar lugar rápidamente entre personas en diferentes lugares. Junto con la virtual eliminación del tiempo, también elimina el espacio.

Debido a que el internet es un medio anónimo, los usuarios pueden desarrollar su propia identidad en línea, bajo sus propios términos, o pueden permanecer anónimos. La anonimidad es importante para la colaboración en línea, especialmente cuando las personas expresan ideas y opiniones a otras personas.

Al ser una comunicación no verbal, todos los aspectos relacionados al lenguaje corporal, posición, aspecto físico, son eliminados, lo que empodera la discusión y el veto de ideas sin importancia de aspectos de género, raza, discapacidades, y otros.

Más que el simple modo de transmisión de información, que es típico con otros medios más viejos como periódicos, radio y televisión, el internet motiva la creación de nuevas ideas. El contenido del internet es una mezcla de contenido de la persona común, contenido de negocios y organizaciones mediáticas. Para algunos, el internet puede aislar a los usuarios de sus vecinos interpersonalmente y permite que los usuarios sean explotados financieramente por algunas compañías. Los usuarios de internet han aprendido como propagar sus propias ideas, descubrir información "enterrada", y replantear ideas y contenido previo en nuevas e innovativas formas. Los usuarios de internet representan solucionadores de problemas.

Un aspecto final del internet que permite que el *crowdsourcing* suceda es que ha bajado las barreras de entrada para una variedad de actividades. El internet ha permitido a las personas conectarse por la velocidad y alcance que

posee, ha roto las barreras que imponen el tiempo y la geografía. Por otro lado, ha bajado las barreras hacia la información e incrementando el acceso a herramientas útiles que antes no eran accesibles.

El internet y otras tecnologías de medios, como por ejemplo las cámaras digitales, han bajado las barreras al conocimiento y han dado acceso a nuevos espacios, para la expresión creativa, el uso compartido de medios, interacción y el hacer negocios. Tomadas juntas, estas características del internet (velocidad, alcance, eliminación de barreras, etc.), permiten una cultura de participación en línea.

### **2.3.2. Resolución de problemas e innovación**

La resolución de problemas a través de *crowdsourcing* se logra definiendo cual es el problema que se posee, y cuál sería el estado deseado. De esta forma, el ambiente de trabajo ya estaría definido, y los *solvers* podrían proponer las soluciones al problema.

El *crowdsourcing* es un modelo de resolución de problemas porque permite a una organización, que posee un problema, escalar el ambiente de la tarea a un nivel global, ampliar la base de soluciones, todo esto a través de plantear el problema a una comunidad en línea a través del internet.

Hay cuatro componentes para la resolución de problemas:

Primero, estado inicial, es el estado del conocimiento de la persona al inicio del problema. Segundo, está el estado deseado, es la meta que la persona desea alcanzar. El tercero son las acciones u operaciones que el *solver* del problema

puede usar para conseguir el estado deseado, y cuarto, es el ambiente de la tarea en la cual el *solver* está trabajando.

El ambiente de la tarea consiste en las características del ambiente físico, que pueden, directa o indirectamente, restringir o sugerir diferentes formas de resolver un problema.

El problema que la organización de *crowdsourcing* necesita resolver puede variar, pero generalmente implica desarrollar un producto, resolver un problema científico difícil, alcanzar un consenso en un problema público o procesar grandes cantidades de datos con inteligencia humana. Más que abordar esos problemas internamente, la organización externaliza el proceso a la multitud (*crowd*) lo cual proporciona un gran y diverso conjunto de herramientas, habilidades e ideas para atacar el problema.

Christian Terwiesch y Yi Xu encontraron que los problemas de generación de ideas son apropiados para una base en línea de solventadores (*solvers*), y concluyen que "la promesa de la innovación abierta es atractiva: se puede incrementar la capacidad para innovar, adhiriéndose a una red de conocimiento que trascienda las fronteras organizacionales"<sup>13</sup>

Esto quiere decir que los problemas que implican la generación de diseños e ideas únicas son buenos candidatos para ser expuestos a una comunidad en línea de individuos que pueden tener algo que aportar al respecto. Visto desde

---

<sup>13</sup> TERWIESCH, Christian y XU, Yi. *Innovation Contests, Open Innovation, and Multiagent Problem Solving*. [en línea]. [https://www.researchgate.net/profile/Christian\\_Terwiesch/publication/220534500\\_Innovation\\_Contests\\_Open\\_Innovation\\_and\\_Multiagent\\_Problem\\_Solving/links/00b495257417332845000000.pdf](https://www.researchgate.net/profile/Christian_Terwiesch/publication/220534500_Innovation_Contests_Open_Innovation_and_Multiagent_Problem_Solving/links/00b495257417332845000000.pdf) Institute for Operations Research and the Management Sciences : 2008, Vol. 54, no. 9. p. 1529-1543. p. 1542 [Consulta: 16 de agosto de 2016]

un enfoque de programas de investigación y desarrollo (R&D), la resolución de problemas bien puede ser un sinónimo con la innovación. La investigación en innovación, tiene dos aspectos bastante relacionados, cuando se refiere a *crowdsourcing*, innovación por el usuario e innovación abierta. En la innovación por el usuario, los individuos fueran de una compañía, por ejemplo los clientes que usan el producto de una compañía, modifican el producto para que cumpla con sus necesidades específicas, y la organización incorpora esas modificaciones en futuras iteraciones que producen al mercado.

Un ejemplo de innovación por el usuario es la que nota, Eric von Hippel, Y está relacionada con la industria de las bicicletas de montaña, donde los usuarios desarrollaron sus propias modificaciones a bicicletas estándar para poder recorrer terrenos difíciles.<sup>14</sup>

La industria del ciclismo se acomodó a este nuevo deporte en buena medida a causa de su inventiva improvisación y experimentación de usuarios que tomaron la iniciativa.

En la innovación abierta, las organizaciones sistemáticamente adoptan la apertura con participantes externos para desarrollar nuevos productos y servicios. El proceso de investigación y desarrollo en la innovación abierta abarca las fronteras entre la firma y el consumidor y el desarrollo puede ir iterativamente en ambas vías. Llevando esta innovación abierta al internet, la amplifica. Llegar a muchos solventadores en sus muchos ambientes de tarea vía internet nos enlaza con un conjunto de heurística cognitiva relacionada a la resolución de problemas y formas de ver el mundo. Una ventaja de la apertura en la resolución de problemas es que las personas externas no expertas tienen una oportunidad

---

<sup>14</sup> VON HIPPEL, Eric. *Democratizing innovation*. London : MIT Press, 2005. p. 72-73.

para proveer soluciones a desafíos organizacionales, a veces brindando mejores soluciones que expertos en el tema.

Lars Bo Jeppesen y Karim R. Lakhani, después de estudiar iniciativas donde se exponían problemas, para que pudieran ser resueltos por científicos, y el éxito que conseguían, concluyeron que “la relación positiva y significativa entre marginalidad, en términos de estar lejos del área de conocimiento asociada al problema y la resolución del problema, sugiere que los solventadores exitosos fueron capaces de brindar perspectivas y heurísticas nuevas al problema en cuestión, en la medida que sus campos de conocimiento estaban lejos del campo de conocimiento del problema”.

Esto quiere decir que un biólogo puede haber tenido mejor desempeño que un químico, en la resolución de un problema de ingeniería química. Encontraron también que las mujeres desempeñaron mejor que los hombres en la resolución de problemas.

Esta marginalidad técnica y social en la resolución de problemas científicos es una ventaja, porque las perspectivas y heurísticas de personas externas permiten ver soluciones novedosas a problemas que personas que están en el centro del dominio del tema, no logran visualizar.

### **2.3.3. Inteligencia colectiva y la sabiduría de las masas**

Un campo interdisciplinario de estudio de la inteligencia colectiva ha surgido, y el *crowdsourcing* está frecuentemente incluido en el discurso de la inteligencia colectiva. La inteligencia colectiva es una forma de inteligencia universalmente distribuida, constantemente mejorada, coordinada en tiempo real y resultante en la movilización efectiva de habilidades. La habilidad para

coordinarse con otros es el corazón de la inteligencia colectiva, y la aparición de internet como una red global que conecta individuos con otros en actividades creativas ha dado marcha a un interés en los estudios de inteligencia colectiva.

Por ejemplo, *Threadless*, es una compañía de ropa en línea. El reto consiste en que los miembros registrados de la comunidad online hagan el diseño y selección de camisetas. Los miembros pueden descargar el diseño de las camisetas y paletas de colores para ser utilizados como herramientas de diseño gráfico, luego ellos suben los diseños a la galería del sitio donde los diseños sometidos permanecen en concurso por una semana. Los miembros votan por los diseños en la galería durante este tiempo, en una escala de 0 a 5 puntos. Al final de la semana, los diseños mejor calificados son los candidatos finalistas para impresión, y el staff de *Threadless*, selecciona cinco diseños para producir cada semana en grandes y limitadas cantidades. *Threadless* proporciona al usuario ganador una participación por cada camiseta vendida.

El *crowdsourcing* ha sido usado para generar contenido periodístico, aunque con resultados mixtos. Por ejemplo, *Assignment Zero*<sup>15</sup>, fue un experimento de periodismo basado en *crowdsourcing* que se enfocaba en los tópicos de *crowdsourcing*, el cual en el momento de ser lanzado este sitio, era un término recién acuñado. Lo que se esperaba era que la comunidad en línea de escritores y editores voluntarios produjeran el compendio más completo acerca de *crowdsourcing* en ese momento. Algunos de los problemas que resultaron fueron que las tareas no eran tan específicas o granulares para los voluntarios, de esta forma el resultado era variable. El producto final fue sometido a muchas críticas y considerado por algunos como un fallo.

---

<sup>15</sup> ASSIGNMENT ZERO. [en línea]. [https://en.wikipedia.org/wiki/Assignment\\_Zero](https://en.wikipedia.org/wiki/Assignment_Zero) [Consulta: 5 de septiembre de 2016]

#### **2.3.4. *Microtasks de crowdsourcing***

Es una simple forma de *crowdsourcing*, es una forma de conectar organizaciones a sus trabajadores potenciales vía internet. Este parece ser un buen método para asignar pequeñas cantidades de trabajo, llamadas microtareas, a lo largo del internet. En *Mechanical Turk*, un sitio de Amazon, Los solicitantes pueden usar el sitio para coordinar una serie de *HIT's*, tareas simples que necesitan ser hechas por humanos. Estas son tareas que las computadoras no pueden hacer fácilmente, Como por ejemplo etiquetar correctamente el contenido de las imágenes de internet para un motor de búsqueda.

Los individuos en la comunidad de *Mechanical Turk*, conocidos como *turkers*, pueden registrarse para cumplir una serie de estas tareas que requieren inteligencia humana llamadas *HITs*, y el solicitante paga pequeñas cantidades de dinero, normalmente centavos, por la realización de dichas tareas. *Mechanical Turk* coordina grandes colecciones de simples tareas que requieran inteligencia humana, así organizaciones que usan este servicio pueden importar análisis de datos de forma rápida y barata.

#### **2.3.5. *Crowdfunding***

Dado que el concepto de *crowdfunding* ha ganado mucha popularidad, vale la pena mencionar como éstos se relacionan con el *crowdsourcing*. El *crowdfunding* describe un modelo de financiamiento donde los individuos usan el internet para contribuir con cantidades relativamente pequeñas de dinero para apoyar la creación de un producto específico y la inversión en una idea de negocios específica. *Kickstarter.com* es un ejemplo de este modelo.

Sin embargo *crowdfunding* no cumple estrictamente con la definición de *crowdsourcing*. *Crowdsourcing* es una mezcla de procesos administrados *top-down* y procesos abiertos *bottom-up*, y el control de la producción reside tanto en la organización y en el *crowd* en una forma compartida. Pero en *crowdfunding* un artista o un emprendedor desarrollan una idea que busca apoyo monetario para brindar esta idea al mercado. No hay compromiso a una participación activa con la comunidad.

*Crowdfunding* es financiamiento distribuido, inversión de grupo, no *crowdsourcing*. Incluso el ejemplo de *Mechanical Turk*, en el cual invita al *crowd* a realizar tareas simples humanas por un pago, los miembros de la comunidad contribuyen su talento y su intelecto al trabajo. En *crowdfunding*, no hay energía creativa o inteligencia humana que sea atraída al producto mismo por parte de la comunidad. El único intelecto requerido por la comunidad es elegir a qué producto apoyar.

#### **2.4. Organización del *crowdsourcing***

Los tipos de *crowdsourcing* pueden ser ordenados de muchas maneras, dependiendo del enfoque que se les asigne. Pero según Daren Brabham, los tipos de *crowdsourcing* pueden ser organizados en una tipología de 4, de acuerdo a la naturaleza de los problemas que están resolviendo.

Estos son:

- Descubrimiento de conocimiento y administración
- Búsqueda por difusión
- Creación y producción vetada por los pares
- Labor de inteligencia humana distribuida

#### **2.4.1. Descubrimiento de conocimiento y administración**

En este enfoque, las comunidades en línea son retadas a descubrir conocimiento existente en la red, amplificando de esta manera las capacidades de descubrimiento de una organización con recursos limitados.

La premisa es que existe una de conocimiento desorganizado, pero que con procesos *top-down* de administración, se puede dispersar eficientemente una comunidad en línea de individuos, para encontrar conocimiento específico y colectarlo de una forma específica en un repositorio común.

Este tipo de *crowdsourcing* se parece la producción basada en pares, como la escritura y edición hecha en Wikipedia, excepto que una organización patrocinadora determina exactamente qué información es buscada, cual es el propósito y como la información es estructurada. En esta aproximación, cuantos más usuarios existen y cuanto más involucrados estén, mejor es el funcionamiento del sistema. Un hecho que puede ser aplicado a la mayoría de fenómenos de cultura de participación.

#### **2.4.2. Búsqueda por difusión**

El *crowdsourcing* basado en búsqueda por difusión está orientado a encontrar un solo especialista, quien probablemente está fuera del campo directo de experticia del problema y quien tiene el tiempo y es capaz de adaptar trabajo previo para producir una solución. En teoría, cuanto más grande sea la red de difusión, más fácil es que la compañía encuentre el especialista que busca, es decir, la persona que sabe la respuesta al problema.

Este tipo de *crowdsourcing* es apropiado para problemas donde una probable y empírica respuesta correcta existe, pero no es conocida aún por la organización. Difundir el problema en una manera en línea, abierta, resulta en soluciones potenciales. Problemas científicos, como desarrollar nuevos químicos y materiales, o ubicar recursos para minería usando datos geofísicos, son más apropiados para este enfoque. En tipo de *crowdsourcing* de búsqueda por difusión, las recompensas monetarias son comunes para individuos en el *crowd* que proveyeron una solución al reto, aunque los incentivos financieros no son la única motivación para estas multitudes para participar en estas situaciones.

### **2.4.3. Producción creativa vetada por los pares**

Con esta modalidad, la fase creativa de un producto diseñado es abierta a una red de usuarios de internet, quienes envían a su vez sus comentarios, y en algunos casos, incluso ideas superiores. Este tipo de *crowdsourcing*, simultáneamente identifica las mejores ideas y transforma el proceso de la investigación de mercado en una instancia de cocreación firma-consumidor. Es un sistema donde una buena solución es también la solución popular que el mercado apoyará. Este tipo de *crowdsourcing* es apropiado para la resolución de problemas relacionados a gustos y preferencias del usuario, tales como problemas de estética o de diseño. Aquí se puede dar el proceso de ideamiento donde la multitud resulta con ideas creativas para productos, contenido o diseños de espacio físico.

Debido a que la multitud es el eventual usuario del producto, del contenido o del espacio, ellos están empoderados para seleccionar las mejores ideas.

#### 2.4.4. Labor de inteligencia humana distribuida

Este tipo de *crowdsourcing* es apropiado cuando la cantidad de datos es conocida, y el problema no es producir diseños, encontrar información o desarrollar soluciones, sino procesar datos. Es similar a proyectos de computación distribuida como *SETI*, excepto que reemplaza los ciclos disponibles (*spare*) de computación con humanos involucrados en ciclos cortos de trabajo. Problemas de gran cantidad de datos son descompuestos en pequeñas tareas, que requieren inteligencia humana, y los individuos en la multitud, son recompensados por el procesamiento de estos pedazos de datos. Debido a que este tipo de *crowdsourcing* es el menos demandante a nivel creativo e intelectual, la compensación monetaria es un motivador común para la participación. Un ejemplo de este tipo de *crowdsourcing* es el ya mencionado *Mechanical Turk* o implementaciones computacionales como *CAPTCHA*.

Tabla I. Tipos de *crowdsourcing*

Tipo	Como funciona	Problemática que resuelve	Ejemplos
Descubrimiento de conocimiento y administración	La organización encomienda al <i>crowd</i> con encontrar y coleccionar información en una ubicación y formato común	Ideal para la recopilación de información y reporte de problemas, tales como la creación de recursos colectivos	peertopatent.org
Búsqueda por difusión	La organización encomienda al	Ideal para problemas de	<i>InnoCentive.org</i>

Continuación de Tabla I

	<i>crowd</i> con resolver problemas empíricos	ideación con soluciones empíricamente probables, tales como problemas científicos	
Producción creativa vetada por los pares	La organización encomienda al <i>crowd</i> con la creación y selección de ideas creativas	Ideal para problemas de ideación, donde las soluciones tienen que ver con el gusto o con el apoyo del mercado, tales como diseños o problemas estéticos	threadless.com
Labor de inteligencia humana distribuida	La organización encomienda al <i>crowd</i> con analizar grandes cantidades de información	Ideal para análisis de datos a gran escala donde la inteligencia humana es más eficiente o efectiva que el análisis de computadoras	Amazon <i>Mechanical Turk</i>

Fuente: elaboración propia, basado en el diagrama de BRABHAM, Daren C. *Crowdsourcing*. Cambridge, Massachusetts : The MIT Press, 2013. p. 45.

#### **2.4.5. Divisiones disciplinarias**

La investigación sobre el *crowdsourcing* se ha incrementado en una variedad de disciplinas académicas, y cada disciplina ha enfocado el tema desde un ángulo distinto.

Por ejemplo, la investigación computacional sobre el *crowdsourcing* se ha enfocado en el diseño y los aspectos técnicos de sistemas de *crowdsourcing*. Algunos estudios evalúan el rendimiento de sistemas de *crowdsourcing* existentes, algunos proponen sistemas hipotéticos o modelos para varias aplicaciones de *crowdsourcing*. Y otros encuentran formas novedosas de aplicación o de reenfocar sistemas para propósitos distintos para los cuales fueron creados.

La disciplina de computación es la que más investigación ha hecho sobre el tema de *crowdsourcing*. Este gran volumen de investigación puede ser explicado por la brevedad de las publicaciones en la disciplina, tendencia a publicar artículos de las conferencias, mientras suceden o hacerlos disponibles en línea lo más rápido posible y el hecho que la conferencia completa, así como sus talleres han sido dedicados al tema del *crowdsourcing*, atrayendo grandes números de académicos en conversación con sus iguales.

Sin embargo, el volumen de investigación sobre el *crowdsourcing* por parte de la disciplina de la computación se debe principalmente a que esta lleva tiempo investigando la computación distribuida.

La ciencia de la computación ya ha tenido un momento presente para este tipo de trabajo y esta forma de pensar, acerca de problemas en una forma distribuida, colectiva y multitudinaria. La computación distribuida es simplemente

la distribución de pequeñas partes de un problema de computación a diferentes computadoras en una red.

Dado un problema computacional grande que una computadora no puede manejar rápidamente, tiene sentido descomponerlo en pequeñas tareas de asignar las diferentes computadoras individuales que se comuniquen a través de una red compartida. Esto esencialmente amplifica el poder de la computación de una manera distribuida y permite a un problema computacional grande ser solucionado en un tiempo razonable.

Un ejemplo relativamente bastante conocido de computación distribuida es el proyecto de búsqueda de inteligencia extraterrestre, *SETI@home project* (*SETI es Search for Extra-Terrestrial Intelligence*). En este caso una gran cantidad de datos de radiotelescopio Arecibo es distribuido en pequeñas cantidades a voluntarios en el internet, los cuales poseen computadoras personales que procesan los datos con un algoritmo que busca por patrones y otras características que puedan sugerir la evidencia de comunicación alienígena en el espacio.

Los descubrimientos de las computadoras individuales sobre los datos que se le han proporcionado, son comunicados de vuelta a las redes de computadoras del proyecto *SETI@home*. El descomponer enormes cantidades de datos de Arecibo en porciones pequeñas para que las computadoras personales puedan manejarlas, ayuda al proyecto en su búsqueda por vida extraterrestre.

En este proyecto así como en otros, las computadoras personales en las redes procesan la data usando un software que el voluntario descarga hacia su

computadora. Sin embargo, no todas las tareas de computación pueden ser manejadas eficientemente por algoritmos de computadoras.

Si por ejemplo la tarea es etiquetar un gran número de fotos por su contenido relevante, los algoritmos existentes pueden eficientemente detectar el tamaño de la foto, el color dominante en la imagen, e incluso el número de personas presente en la foto. Pero probablemente no existe un algoritmo que pueda fácilmente etiquetar una persona sosteniendo una caja de pizza como *Pizza vegetariana* o *Pizza de pepperoni*, y no tendría sentido escribir un algoritmo tan específico para este tipo de etiquetado. Un humano, puede simplemente y de una manera rápida identificar los aderezos en la pizza y proporcionar una etiqueta apropiada para la imagen.

Tomando los principios de descomposición de un problema en la computación distribuida y reemplazando las computadoras en la red con humanos, produce un proceso híbrido computadora-humano que Luis Von Ahn llama computación humana. A través de la investigación que von Ahn realizó, concluye: "creemos que los resultados presentados aquí son parte de una prueba de concepto de una idea más general, el procesamiento humano desperdiciado puede ser utilizado para resolver problemas que las computadoras no pueden. Algunos han referido a esta idea como computación humana" <sup>16</sup>

Grandes cantidades de datos que requieren inteligencia humana pueden ser procesados a través de la computación humana de la misma forma que grandes grupos de datos son procesados en computación distribuida. El tipo de

---

<sup>16</sup> VON AHN, L., MAURER, B., MCMILLEN, C., ABRAHAM, D. y BLUM, M. *reCAPTCHA: Human-Based Character Recognition via Web Security Measures*. [en línea]. [https://www.cs.cmu.edu/~biglou/reCAPTCHA\\_Science.pdf](https://www.cs.cmu.edu/~biglou/reCAPTCHA_Science.pdf) . 2008, Vol. 321, no. 5895, p. 1465-1468. p.1467 [Consulta: 8 de septiembre de 2016]

*crowdsourcing* basado en inteligencia humana distribuida y la computación humana, son los mismos conceptos, y la investigación en *crowdsourcing* desde la perspectiva de computación de una forma ha juntado las terminologías y las teorías de computación distribuida, computación humana y *crowdsourcing*.

Mucho del trabajo del *crowdsourcing* en la disciplina de computación sin embargo ha tomado lugar de una forma desconectada de investigaciones similares en otras disciplinas. La investigación acerca de *crowdsourcing* en la disciplina de computación ha sido robusta, y en buena medida se puede atribuir este alto nivel al hecho que la disciplina de computación incluye diseñar hacer y construir, igual a casi todo científico de computación, que sabe cómo programar.

## **2.5. Problemática del *crowdsourcing***

Como cualquier actividad, el *crowdsourcing* afronta una problemática asociada, Desde situaciones que involucran la actividad del *crowdsourcing* por sí mismo, hasta situaciones que se dan dentro de la sociedad.

### **2.5.1. Moviendo a la multitud**

Todos los individuos que están involucrados en el *crowdsourcing* están de alguna forma motivados, y entender cómo y porque las multitudes participan es necesario para diseñar aplicaciones efectivas de *crowdsourcing*. la motivación para participar en *crowdsourcing* no es diferente de la motivación para participar en *blogging*, crear software *open source*, postear vídeos a Youtube, contribuir a Wikipedia e incluso etiquetar contenido de una red social.

Según Edward L. Deci y Richard M. Ryan, hay 2 tipos de motivadores que hacen que las personas realicen o se involucren en actividades: Los motivadores intrínsecos y los extrínsecos

Ellos escriben que:

"Los comportamientos motivados intrínsecamente son desempeñados sin ningún interés y no requieren una consecuencia separable, producto externo o amenaza. Cuando la gente es motivada intrínsecamente, está motivada simplemente para desarrollar la actividad y para tener experiencias espontáneas de interés, disfrute, excitación y satisfacción que acompañan dicho comportamiento".<sup>17</sup>

El comportamiento motivado intrínsecamente representa el prototipo de la autodeterminación. Cuando es motivada intrínsecamente, las personas se sienten totalmente autónomas y experimentan este comportamiento como una expresión de ellos mismos.

En contraste a la motivación intrínseca, ser extrínsecamente motivado implica desarrollar actividades con la intención de obtener algún tipo de consecuencia separable, como por ejemplo recibir una recompensa, evitar una culpa o ganar aprobación. Los comportamientos que son extrínsecamente motivados generalmente no ocurren de forma espontánea, así que su ocurrencia debe ser típicamente promovida por algún tipo de instrumentalidad."<sup>18</sup>

---

<sup>17</sup>DECI, Edward L., RYAN, Richard M. y WILLIAMS, Geoffrey C. *Need satisfaction and the self-regulation of learning. Learning and Individual Differences*. [en línea]. [http://www.academia.edu/download/44801628/Need\\_satisfaction\\_and\\_the\\_self-regulatio20160416-28775-cexpr1.pdf](http://www.academia.edu/download/44801628/Need_satisfaction_and_the_self-regulatio20160416-28775-cexpr1.pdf) . Elsevier BV, 1996, Vol. 8, no. 3, p. 165-183. p. 167 [Consulta: 13 de septiembre de 2016]

<sup>18</sup> DECI, Edward L., RYAN, Richard M. y WILLIAMS, Geoffrey C. *Need satisfaction and the self-regulation of learning*. [en línea].

Los motivadores más comunes de los individuos que participan el *crowdsourcing* son:

- Ganar dinero
- Desarrollar habilidades creativas
- Establecer un vínculo con otros profesionales creativos
- Construir un portafolio para empleo futuro
- Aceptarse a sí mismo para solventar un problema difícil
- Socializar y hacer amigos
- Pasar el tiempo
- Contribuir a un proyecto más grande de interés común
- Compartir con otros
- Divertirse

## **2.6. Problemas legales**

Debido a que el *crowdsourcing* se mueve entre las fronteras del profesionalismo y el amateurismo, así como entre esquemas de negocios caseros y accionistas externos. Muchos problemas legales rodean este modelo, listando a continuación algunos de ellos.

---

[http://www.academia.edu/download/44801628/Need\\_satisfaction\\_and\\_the\\_self-regulatio20160416-28775-cexrp1.pdf](http://www.academia.edu/download/44801628/Need_satisfaction_and_the_self-regulatio20160416-28775-cexrp1.pdf). Elsevier BV, 1996. Vol. 8, no. 3, p. 165-183. [Consulta: 1 de octubre de 2016] p. 167.

### 2.6.1. Libertad de expresión y disensión

Las sociedades libres defienden el principio de libre expresión. La libre expresión alienta un amplio rango de criticismo. El escepticismo por otro lado tiende a mejorar los sistemas, las personas, las ideas que critican.

La libre expresión es un valor democrático y también un valor importante dentro de las organizaciones que apoyan la innovación y la resolución de problemas. Seguramente la libertad de expresión es crucial para el éxito de cualquier aplicación de *crowdsourcing*, Pero, ¿qué pasa cuando las multitudes protestan y amenazan con destruir la aplicación?

Esta pregunta es especialmente importante para actividades de *crowdsourcing* manejadas por el gobierno. Una aplicación de *crowdsourcing* manejada por el gobierno enfrenta el reto de manejar las protestas de los ciudadanos sin infringir su derecho a expresarse acerca del proyecto.

De aquí se deriva un término llamado *crowdslapping*, esto es:

La resistencia de una multitud dentro de una actividad de *crowdsourcing* el espacio de una comunidad, hacia el *crowdsourcer* o a la actividad misma.

Existen varias formas de *crowdslapping*:

#### a) **Crowdslapping disruptivo:**

El cual se parece a las protestas físicas en espacios públicos. Busca disruptir el buen funcionamiento de un proyecto de *crowdsourcing* a través de quejas posteadas en foros en línea.

Este tipo de *crowdslapping* puede tomar la forma de un argumento razonado que es articulado por un individuo en la multitud y posteado en un sitio del *crowdsourcing*, o puede tomar la forma de una petición virtual "pacífica" contra el gobierno.

**b) *Crowdslapping* destructivo:**

Por otro lado, desalienta a los otros ciudadanos de participar en el sitio a través de ataques agresivos como por ejemplo *flaming* y *flooding*.

El *flaming* son tácticas usadas por los participantes para proferir insultos. El *flooding* es "inundar" un boletín en línea o un espacio de chat con texto "basura", como una forma de disminuir el tráfico del sitio y destruir cualquier potencial conversación racional.

Definitivamente la falta de reglas de ciertas comunidades en línea contrasta dramáticamente con las visiones ideales de un debate racional deliberado por proponentes democráticos. Pero aun así, Incluso el *flaming* y el *flooding* deben ser celebrados como momentos de ejercicio de la democracia donde incluso los sentimientos menos populares son evaluados como posibles verdades.

**c) *Cracking*:**

Es otra forma de llamar al *hacking* malicioso, pues trata de tomar acción en destruir los mecanismos del sitio de *crowdsourcing*,

probablemente a través de accesos ilegales y de la manipulación del código del sitio.

*Crowdslapping* destructivo desalienta el diálogo productivo de una aplicación de *crowdsourcing*, pero el *cracking*, previene a los individuos incluso de interactuar con el proyecto de *crowdsourcing*. Esto implica acciones como apagar chats y boletines, corromper archivos de datos o traer abajo al sitio.

**d) *Ignoring*:**

Es la forma más pacífica y efectiva que existe de *crowdslapping*. Pues los proyectos de *crowdsourcing* requieren de una cantidad de una multitud considerable de individuos que tratarán de resolver el problema dado. Sin este número de usuarios, el proceso falla.

**2.6.2. Propiedad intelectual y derechos de copia**

Cualquier sitio web, especialmente uno que maneje contenido generado por los usuarios, necesita tener una correcta declaración de términos de uso y otras políticas que protejan tanto la organización de *crowdsourcing* como a la multitud. Las compañías más exitosas de *crowdsourcing* tienen políticas para proteger ambas partes de forma justa, y estas políticas son fáciles de encontrar y de entender.

### **2.6.3. Prácticas de negocio injustas**

La mayoría de modelos de negocios nuevos tienen aspectos negativos, como el mercado tiene mercados negros. El *outsourcing* no es la excepción, y el *outsourcing* basado en inteligencia humana distribuida da lugar a una forma específica de manipulación de mercado, que puede ser causa de preocupación legal.

Dentro de estos escenarios podemos considerar aquellas iniciativas de *outsourcing* que manipulan los resultados de la multitud (*crowd*) a través de tráfico comprado y dirigido, así como la generación sistemática de datos ingresados por humanos, los cuales han sido pagados para que genere una tendencia específica.

### **2.7. Problemas éticos**

A menudo se ha acusado al *crowdsourcing* de ser una forma de esclavitud digital y de explotación de masas. Superficialmente, el *crowdsourcing* es una forma fácil para obtener labor de calidad de forma barata y rápida.

La organización de *crowdsourcing* se beneficia del trabajo de las multitudes sin ofrecer los tipos de recompensas monetarias que son la norma en un acuerdo tradicional de trabajo.

En ocasiones se dice que el uso de masas limita a la clase profesional, desechando años de labor de las asociaciones profesionales para aumentar los salarios, proteger a los trabajadores, y establecer estándares éticos para la profesión

### 2.7.1. Derechos laborales

Las discusiones de amateurismos en la cobertura popular de la prensa acerca del *crowdsourcing* puede servir como una forma de distraer a las personas que muchos profesionales ya tenían dificultades en vivir a través de su profesión incluso antes de la llegada del *crowdsourcing*, por ejemplo, carteros vs. correo electrónico, y ya en el ámbito de *outsourcing*, las traducciones y transcripciones (entre otras tareas) solicitadas en sitios como *Mechanical Turk*, que tienen una recompensa monetaria más baja que la que podría cobrar un traductor.

Un caso actual que pone de manifiesto el caso de problemas de derechos laborales es el caso de Austin, Texas, USA. Donde el ayuntamiento de esa ciudad requirió a Uber, que sus conductores pasaran el mismo proceso de verificación y revisión de datos que los conductores de taxis normales.<sup>19 20</sup>

Situaciones similares han sucedido en otras ciudades como México, Chile, Argentina, otros.<sup>21</sup>

### 2.7.2. Eficiencia

Aunque el *crowdsourcing* puede ser barato para las organizaciones, puede que no sea eficiente en un panorama general. Compañías como *InnoCentive*, por

---

<sup>19</sup>MCPHATE, Mike. *Uber and Lyft End Rides in Austin to Protest Fingerprint Background Checks*. *nytimes.com* [en línea]. <http://www.nytimes.com/2016/05/10/technology/uber-and-lyft-stop-rides-in-austin-to-protest-fingerprint-background-checks.html> [Consulta: 13 de octubre de 2016]

<sup>20</sup>MEYER, Jared, 2016, *By Losing Uber, Austin Is No Longer A Tech Capital*. *forbes.com* [en línea]. <https://www.forbes.com/sites/jaredmeyer/2016/05/11/by-losing-uber-austin-is-no-longer-a-tech-capital/#79eaf5e76bef> [Consulta: 21 de octubre de 2016]

<sup>21</sup>MEYER, David. *Uber's Troubles Pile Up in Argentina*. *Fortune.com* [en línea]. <http://fortune.com/2016/04/27/uber-argentina-crackdown/> [Consulta: 23 de octubre de 2016]

ejemplo, la multitud consiste principalmente de científicos profesionales entrenados con grados académicos. Esta compañía lanza retos que deben ser resueltos por estos científicos y los recompensa por sus soluciones. Aunque claro está, normalmente sólo necesitan una solución.

Dado esto, cuando se lanza un problema a resolver, docenas de científicos pueden gastar tiempo tratando de solventarlo. Unos cuantos de estos científicos propondrán soluciones, pero hay un gasto de tiempo asociado a los científicos que trataron de resolver dicho problema, especialmente aquellos que propusieron soluciones.

Las soluciones ganadoras son recompensadas con ciertos premios, pero dichos premios pueden ser considerados baratos en comparación con lo que cuesta tener un laboratorio funcional en casa. Definitivamente es eficiente para el *crowdsourcer* (*InnoCentive*), pero de alguna forma puede ser visto como una pérdida en términos de intelecto científico.

Muchas horas de trabajo científico profesional valioso son desperdiciadas en este sistema y resulta en la pregunta de qué otras cosas podría haber estado haciendo todo este talento científico, ¿qué otros problemas no fueron resueltos? visto de esta forma, el *crowdsourcing* puede ser un modelo ineficiente para solventar problemas difíciles, y este es un argumento que los modelos tradicionales usan contra el *crowdsourcing*.

### **2.7.3. Tiranía del acceso**

Aunque la penetración del internet es enorme en naciones industrializadas y tiene un crecimiento rápido en países en desarrollo, existe cierta brecha digital que previene a las multitudes (*crowd*) ser lo más diversas y accesibles.

Por ejemplo, en Estados Unidos existen muchas áreas rurales que aún no tienen conexiones de internet, los afroamericanos y los latinos de habla hispana están dentro de los menos conectados demográficamente.

Dado que el *crowdsourcing* es necesariamente un fenómeno en línea y no todos tienen acceso al internet, la aplicación de *crowdsourcing* no es accesible a todos.

Cuando una multitud no tiene diversidad, puede que no desempeñe también como si la tuviera, Por ejemplo, podría haber algún caso o proyecto de consorcio relacionado a la generación de alguna política, si la multitud no está representada diversamente, esta política podría no servir de forma adecuada aquellos sectores que no fueron representados dentro de la multitud

## **2.8. Usos y ejemplos**

Algunos de los usos y respectivos ejemplos del *crowdsourcing* pueden ser citados en las siguientes áreas:

### **2.8.1. En astronomía**

El *crowdsourcing* en astronomía fue usado a inicios del siglo XIX por el astrónomo Denison Olmsted. Después de haber despertado en una noche de finales de noviembre debido a una lluvia de meteoros, Olmsted notó un patrón en los meteoros. Debido a esto escribió un breve reporte en el periódico local, donde mencionaba que debido a que las estrellas fugaces no eran entendidas totalmente por los meteorólogos, era deseable coleccionar todos los datos relacionados a este fenómeno. Posteriormente, en un periódico de circulación

nacional, solicitó datos al respecto. Los lectores respondieron desde muchos estados.

Estas respuestas le ayudaron a trabajar una serie de descubrimientos científicos, entre ellos que la lluvia de meteoros pueden ser vista a nivel nacional, que caían del espacio como influencia de la gravedad, y que aparecían en ciclos anuales.

Un ejemplo actual es llevado a cabo por la NASA, el proyecto llamado *Star Date: M83* (<http://www.projectstardate.org/#/classify>), que aproximadamente a través de 3,000 fotos de cúmulos de estrellas, están tratando de mapear la galaxia M83.

### **2.8.2. En botánica**

Proyectos como plantnet-project (<http://identify.plantnet-project.org/>) y Tela Botanica (<http://www.tela-botanica.org/site:accueil>) permiten la identificación de plantas. A través de una aplicación se sube la foto de la planta en cuestión, a través de un algoritmo se comparan con aquellas que ya fueron identificadas, y las que son desconocidas son sometidas a identificación por los contribuyentes.

### **2.8.3. En entomología**

El mismo concepto ha sido usado en el proyecto Tela Insecta (<http://www.tela-insecta.net/>).

#### **2.8.4. En monitoreo del clima**

*Open Signal*, a través de una aplicación llamada *WeatherSignal*, permite recopilar información ambiental de los sensores de los dispositivos (teléfonos, etc.) donde está instalado y dar reportes de clima en tiempo real.

#### **2.8.5. En geología**

La *United States Geological Survey* es una entidad de los Estados Unidos de América, que proveen ciencia acerca de peligros naturales que afecten vidas, el agua, energía, minerales, y otros recursos naturales de los que la humanidad depende. En su sitio <http://earthquake.usgs.gov/data/dyfi/>, permite que las personas reporten eventos sísmicos que los contribuyentes percibieron.

#### **2.8.6. En procesos de digitalización**

Otro ejemplo conocido es *reCAPTCHA*, el cual solicita a las personas resolver *CAPTCHAs* para probar que son humanos, dichos *CAPTCHAs* los generan basados en libros viejos que no pueden ser descifrados por computadoras (OCR) y los resultados son usados como la digitalización de dichos libros.

#### **2.8.7. En aprendizaje**

Duolingo es una plataforma que permite el aprendizaje de idiomas. Para ganar dinero, esta plataforma adoptó un modelo de *crowdsourcing*, donde miembros del público fueron invitados a traducir contenido y a votar por traducciones. El contenido viene de organizaciones que pagan a Duolingo por traducirlo.

## 3. CAPTCHA

### 3.1. Conceptos

*CAPTCHA* es uno de los enfoques que un proyecto de *crowdsourcing* puede tomar. El término fue acuñado por Luis Von Ahn en 2003, quien lo define como: "un programa que puede generar un y calificar una prueba que: (A) la mayoría de humanos puede pasar, pero (B) los programas actuales de computadoras no pueden pasar"<sup>22</sup>.

*CAPTCHA* es el acrónimo para *Completely Automated Public Turing Test to Tell Computers and Humans Apart* (Prueba de Turing completamente automática y pública para diferenciar computadoras de humanos).

También hacen notar que no hay forma de probar que un programa no pueda pasar una prueba que un humano si puede, tomando como premisa que existe un programa que si lo hace, en referencia al ser humano. Pero lo que sí se puede demostrar es la evidencia que es difícil escribir un programa que pueda pasar el test en cuestión.

Mencionan que se puede establecer, para *CAPTCHA*, un enfoque similar al usado en la criptografía. Creen que un componente importante de la criptografía moderna es la práctica de definir precisa y claramente las suposiciones bajo las cuales los protocolos criptográficos son seguros, y que esto permite al resto de la comunidad probar estas suposiciones y tratar de quebrarlas.

---

<sup>22</sup> VON AHN, Luis, BLUM, Manuel y HOPPER, Nicholas J. *CAPTCHA: Using hard AI problems for security. International Conference on the Theory and Applications of Cryptographic Techniques* [en línea]. <http://repository.cmu.edu/cgi/viewcontent.cgi?article=1142&context=compsci> Springer, 2003. p. 294-311. [Consulta: 10 de octubre de 2016]

Al usar un problema de inteligencia artificial, como componente criptográfico, y definirlo de forma precisa y clara, permite a la comunidad hacer la misma validación. Es más, el definirlo claramente ayuda al desarrollo de la Inteligencia artificial, pues la comunidad tratará de desarrollar programas de IA que puedan solventar el problema definido.

Por lo tanto tiene lógica usar un problema de IA como componente criptográfico, es una situación de ganar-ganar, pues si los programas generados son incapaces de resolver el problema planteado, el problema de IA cumple su función criptográfica, pero si hay algún programa que pueda resolver el problema, definitivamente puede ser usado para resolver ese tipo de problemas de IA y similares.

## **3.2. Aplicaciones**

Algunas de las aplicaciones que más frecuentemente vemos en línea son:

### **3.2.1. Registro de correos electrónicos**

Al momento de solicitar la creación de un correo electrónico, la mayoría de estos servicios (Gmail, Yahoo Mail, Outlook, etc.) solicitan la resolución de un *CAPTCHA*, para poder continuar con el registro. Esto es utilizado incluso en los procesos de recuperación de contraseñas.

### **3.2.2. Motores de búsqueda**

Algunos motores de búsqueda, como Google, al identificar una búsqueda que contiene parámetros similares a los usados por los *bots*, solicitan resolver un *CAPTCHA*.

### **3.2.3. Servicios de compartición de archivos**

Servicios como *Mediafire*, que permiten subir y compartir archivos online, usan *CAPTCHA* como un mecanismo para impedir que *crawlers* puedan descargar de forma sistemática sus contenidos.

### **3.2.4. Formularios para compartir links**

Algunos formularios que permiten compartir links o porciones de texto, vía correo electrónico, usan *CAPTCHA* para evitar que *spammers* puedan usar su formulario como un medio para el envío de su *spam*.

## **3.3. Tipos de *CAPTCHA***

Von Ahn define 2 tipos de *CAPTCHA*: *MATCHA* y *PIX*.

### **3.3.1. *MATCHA***

En este tipo de *CAPTCHA*, se definen 2 imágenes, una original (sin transformación) y una transformada (que puede ser o no la primera imagen transformada). El programa sabe si las imágenes son iguales o no. Luego envía al *solver* el par de imágenes, preguntándole si son iguales o si son diferentes.

El usuario contesta, y el programa de *CAPTCHA* valida la respuesta:

- Si la primera imagen es igual a la segunda, y el usuario responde que son iguales, entonces el *MATCHA* es resuelto
- Si la primera imagen es igual a la segunda, y el usuario responde que son distintas, entonces el *MATCHA* no es resuelto
- Si la primera imagen es distinta a la segunda, y el usuario responde que son iguales, entonces el *MATCHA* no es resuelto
- Si la primera imagen no es igual a la segunda, y el usuario responde que son distintas, entonces el *MATCHA* se vuelve a plantear con otro par de imágenes

Como se puede ver, uno de los principales problemas de este enfoque es que si se cuenta con un programa que siempre diga que las imágenes son iguales, se puede obtener una probabilidad de 0.5 de pasar la prueba, lo cual sería poco aceptable.

### **3.3.2. PIX**

En el caso de *PIX*, si se tiene una imagen que contiene una palabra, y la palabra que contiene es mapeada a dicha imagen, el programa de *CAPTCHA* sabe cuál es la palabra que está asociada a la imagen.

Luego el programa envía la imagen al *solver*, y solicita que indique o escriba cual es la palabra asociada:

- Si la palabra ingresada por el usuario es igual a la palabra asociada a la imagen, por el programa de *CAPTCHA*, entonces el *PIX* es resuelto.
- En caso contrario, se envía un nuevo *CAPTCHA* para ser resuelto.

En este tipo de *CAPTCHA*, las probabilidades de resolver al azar un *PIX* son extremadamente bajas.

### **3.4. Ejemplos**

Algunos ejemplos de implementaciones de *CAPTCHA* son:

#### **3.4.1. *RECAPTCHA***

Es una implementación de *PIX*, pero usa normalmente 2 imágenes de palabras. Una de control, la cual tiene una palabra conocida asociada, y otra, que es la imagen a resolver, extraída de otra fuente, de la cual no se sabe su palabra asociada.

Ambas imágenes de palabras son transformadas y presentadas a los usuarios, los cuales deben responder para ambas imágenes.

Como el usuario no sabe cuál es la imagen de control, trata de responder ambas imágenes de forma adecuada, de esta manera, al responder la imagen de control, el *PIX* es resuelto, y la otra respuesta es empleada para formar una base estadística de lo que la comunidad cree que es la imagen a resolver.

#### **3.4.2. *CAPTCHA* de reconocimiento de imagen**

En este formato, una serie de imágenes es presentada al *solver*. Este formato puede adoptar cualquiera de estos 3 tipos:

### 3.4.2.1. CAPTCHA de nombramiento de imagen

En este formato, se muestra un conjunto de imágenes, que tienen algo en común, lo cual debe ser escrito por el *solver*.

Figura 3. CAPTCHA de nombramiento de imagen



Fuente: CHEW, Monica y TYGAR, J. D. *Image recognition CAPTCHAs* [en línea] [http://people.ischool.berkeley.edu/~tygar/papers/Image\\_Recognition\\_CAPTCHAs/imagecaptcha-tr.pdf](http://people.ischool.berkeley.edu/~tygar/papers/Image_Recognition_CAPTCHAs/imagecaptcha-tr.pdf). Berkeley, California : University of California, 2004 [Consulta : 28 de octubre de 2016]. p.

5

Por ejemplo, en la imagen anterior se muestran 6 imágenes correspondientes a un astronauta, así que la respuesta correcta para resolver este CAPTCHA, sería la palabra "astronauta".

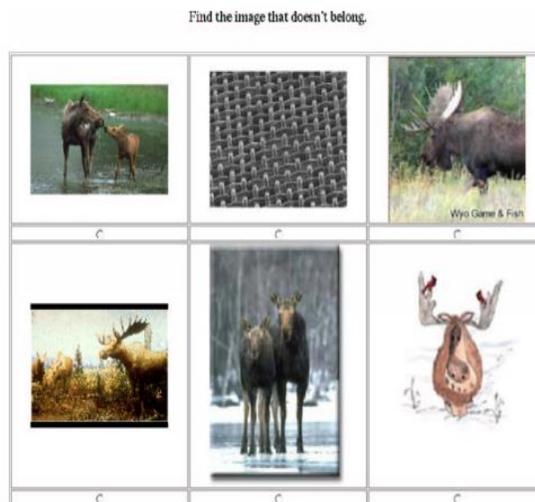
### 3.4.2.2. CAPTCHA de distinción de imágenes:

En este formato, se presentan 2 conjuntos de imágenes, que pueden o no tener un tema en común. El *solver* debe indicar si tienen o no un tema en común y cual sería (o serían) el tema.

### 3.4.2.3. CAPTCHA de identificación de anomalía

En este formato, se presentan un conjunto de imágenes, las cuales a excepción de 1, corresponden a una misma temática. Entonces el *solver* debe indicar, cuál es la imagen que no está relacionada con las demás (cuál es la anomalía).

Figura 4. CAPTCHA de identificación de anomalía



Fuente: CHEW, Monica y TYGAR, J. D. *Image recognition CAPTCHAs* [en línea] [http://people.ischool.berkeley.edu/~tygar/papers/Image\\_Recognition\\_CAPTCHAs/imagecaptcha-tr.pdf](http://people.ischool.berkeley.edu/~tygar/papers/Image_Recognition_CAPTCHAs/imagecaptcha-tr.pdf). Berkeley, California : University of California, 2004 [Consulta : 28 de octubre de 2016]. p.

Por ejemplo, en la imagen anterior, hay 6 imágenes, y se pide identificar la anomalía (la que no pertenece). Dado que 5 de las imágenes son de alces, la imagen del centro superior (una pieza de tela), debe ser la identificada como anomalía, para poder pasar la prueba.

### 3.4.3. *CAPTCHA* de identificación de rompecabezas

Este tipo de *CAPTCHA* muestra una imagen, a la cual se le han abstraído una o más áreas (en forma de rompecabezas), las cuales son colocadas en otra ubicación. El programa generador del *CAPTCHA* sabe cuáles son las coordenadas correctas donde se deben colocar las piezas removidas.

El usuario debe arrastrar y ubicar adecuadamente las piezas a manera de completar la imagen (rompecabezas). Luego presiona un botón para validación, y el programa validará si la ubicación de las piezas es la correcta. En caso que es la correcta, el *CAPTCHA* habrá sido resuelto.

Figura 5. ***CAPTCHA* de identificación de rompecabezas**



Fuente: KEY CAPTCHA. *Free Key Captcha Sample* [en línea]  
<https://www.keycaptcha.com/products/> [Consulta: 28 de octubre de 2016].

#### 4. BREVE RESEÑA DEL ESTADO ACTUAL DE LOS CATÁLOGOS BIBLIOGRÁFICOS EN LAS BIBLIOTECAS DE GUATEMALA.

Los tipos de bibliotecas que existen en Guatemala son diversos, las hay nacionales, especializadas, escolares, infantiles y públicas, así como centros de documentación.

Algunas de estas bibliotecas, como las pertenecientes a las universidades, poseen centros regionales y sedes, pero los procesos de catalogación no siempre se llevan a cabo en dichas sedes, sino que se realiza en la biblioteca central, y luego el material catalogado, es enviado, junto con su catálogo a la sede respectiva, como ocurre en el caso de la Biblioteca de la Universidad Mariano Gálvez<sup>23</sup>.

Un caso distinto ocurre con la mayoría de centros regionales y sedes, que constituyen el Sistema Bibliotecario de la Universidad de San Carlos, los cuales realizan sus propios procesos bibliotecológicos, como se hace notar en sus Manuales de Normas y Procedimientos<sup>24</sup>.

Ahora bien, no todas las bibliotecas de Guatemala tienen la capacidad necesaria para poder elaborar sus propios catálogos. Esto puede ser por falta de herramientas, por desconocimiento o por falta de presupuesto para la contratación del personal calificado que realice esta tarea.

---

<sup>23</sup> CENTROS REGIONALES - BIBLIOTECAS DE CENTROS UNIVERSITARIOS. *umg.edu.gt* [en línea]. [https://www.umg.edu.gt/biblioteca/?page\\_id=368](https://www.umg.edu.gt/biblioteca/?page_id=368) [Consulta: 13 de octubre de 2016]

<sup>24</sup> UNIVERSIDAD DE SAN CARLOS DE GUATEMALA. *Manual de Organización Biblioteca Central* [en línea]. <http://ddo.usac.edu.gt/wp-content/uploads/2015/01/Manual-Espec%C3%ADfico-de-BIBLIOTECA.pdf> Guatemala : Universidad de San Carlos de Guatemala. [Consulta: 13 de octubre de 2016]

El Instituto Nacional de Estadística, dentro de los múltiples informes que desarrolla, posee como parte de sus Estadísticas Sociales a las Estadísticas de Bibliotecas y Centros de Documentación.

Dichas estadísticas las definen como: "Las estadísticas de bibliotecas y centros de documentación, son generadas de registros administrativos que se recopilan del formulario INE 199-B. Permite contar con información sobre las características de cada establecimiento registrado como fuente, los materiales de distintas áreas de conocimiento especializado que disponen, así como información de usuarios que visitan dichos centros de información y el personal que labora en los mismos, con el fin de apoyar a profesionales e investigadores y a público en general"<sup>25</sup>.

Dentro de uno de sus últimos informes de estadísticas de bibliotecas y centros de documentación<sup>26</sup>, correspondiente al año 2013, se puede observar en la hoja "Cuadro 6", la siguiente información:

---

<sup>25</sup>BIBLIOTECAS Y CENTROS DE DOCUMENTACIÓN. *ine.gob.gt* [en línea].  
<http://www.ine.gob.gt/index.php/estadisticas-continuas/bibliotecas-y-centros-de-documentacion>  
[Consulta: 29 de octubre de 2016]

<sup>26</sup>INSTITUTO NACIONAL DE ESTADÍSTICA. *Estadísticas de Bibliotecas y Centros de Documentación* [en línea].  
<http://www.ine.gob.gt/sistema/uploads/2014/07/09/XYsJjXB87HHnL8jT6A2SPugLPyVTLYmp.xlsx>  
Guatemala : INE, 2013. [Consulta: 29 de octubre de 2016].

Tabla II. **Estadísticas de tipos de catálogo en bibliotecas del país**

<b>Bibliotecas y centros de documentación por tipo de catálogo, según departamento, Año 2013</b>				
<b>Departamento</b>	<b>Tipo de catálogo</b>			
	<b>Fichas</b>	<b>Computarizado</b>	<b>Listado impreso</b>	<b>No dispone de catálogo</b>
<b>Total país</b>	<b>141</b>	<b>113</b>	<b>41</b>	<b>94</b>
Guatemala	53	67	9	25
El Progreso	2	1	-	4
Sacatepéquez	4	5	-	9
Chimaltenango	8	1	-	4
Escuintla	6	-	2	3
Santa Rosa	4	-	2	1
Sololá	3	7	1	3
Totonicapán	1	-	-	1
Quetzaltenango	8	6	5	3
Suchitepéquez	6	1	2	10
Retalhuleu	3	-	-	6
San Marcos	6	2	3	4

Continuación de Tabla II

Huehuetenango	4	3	-	8
Quiché	4	1	4	-
Baja Verapaz	2	-	2	2
Alta Verapaz	4	13	5	3
Petén	5	2	1	5
Izabal	3	1	-	1
Zacapa	4	2	1	2
Chiquimula	3	-	-	-
Jalapa	5	1	3	-
Jutiapa	3	-	1	-

Fuente: elaboración propia, basado en INSTITUTO NACIONAL DE ESTADÍSTICA. *Estadísticas de Bibliotecas y Centros de Documentación* [en línea].

<http://www.ine.gob.gt/sistema/uploads/2014/07/09/XYsJjXB87HHnL8JT6A2SPugLPyVTLYmp.xls>  
x Guatemala : INE.

Como puede observarse, casi un 25 % de las bibliotecas contabilizadas no posee un catálogo accesible.

Dado lo anterior, sería ideal que estas bibliotecas contaran con la capacidad de poder optar a un profesional de la bibliotecología que les ayudase con el proceso de la elaboración de un catálogo. Lo más probable es que no cuenten con esos recursos, la aplicación de un proyecto de *crowdsourcing* para la elaboración de un catálogo básico podría ser la solución.

## **5. DESARROLLO DEL MÓDULO WEB QUE USARÁ CROWDSOURCING PARA LA ALIMENTACIÓN DE CATÁLOGOS DE BIBLIOTECAS.**

### **5.1. Descripción**

El objetivo es desarrollar un prototipo de un módulo web, que pueda servir como un programa de *CAPTCHA*, que tenga como meta la recopilación de información bibliográfica que permita la creación de un catálogo bibliográfico, para una biblioteca que no tenga los mecanismos para realizar sus propios procesos de catalogación.

### **5.2. Funcionamiento**

El prototipo del módulo debe permitir mostrar las respectivas portadas o páginas de título de 2 publicaciones, para que el *solver*, colabore indicando el título o el nombre del autor asociados a las imágenes, los cuales serán ingresados en campos de texto dispuestos para este propósito.

Las imágenes y campos de texto de estas publicaciones funcionarían de la misma manera que lo hace *reCAPTCHA*. Siendo una de las imágenes, una imagen de control, y la otra, la imagen a solventar.

Posterior a la resolución exitosa del *CAPTCHA*, se puede crear una cookie que permita que la prueba de *CAPTCHA* no sea mostrada nuevamente a los usuarios, hasta pasado un tiempo o cuando cierren la ventana.

Los datos recopilados se irán almacenando en una base de datos, y comparándose entre sí, para encontrar la mayor ocurrencia estadística de

solución propuesta, es decir, que título o autor cree la mayoría de personas que se muestra en la imagen a resolver.

### **5.3. Alcances y límites**

- Se pretende hacer un análisis de los componentes requeridos para el desarrollo del módulo web de ingreso de información bibliográfica.
- Se espera desarrollar un prototipo de módulo web para el ingreso de información bibliográfica que haga uso del *crowdsourcing*, y que se apoye en los conceptos de *CAPTCHA* para el filtrado de la información ingresada.
- El prototipo del módulo web, si bien incluirá la lógica de estadísticas para el filtrado de información, no proporcionará la interfaz (*API, framework, etc.*) para su inclusión en sistemas web, pues eso saldría del foco de investigación.

### **5.4. Diagramas**

A continuación se presentan los diagramas relacionados al diseño del componente, que tienen como propósito brindar un panorama general del módulo desarrollado.

#### **5.4.1. Diagrama de clases**

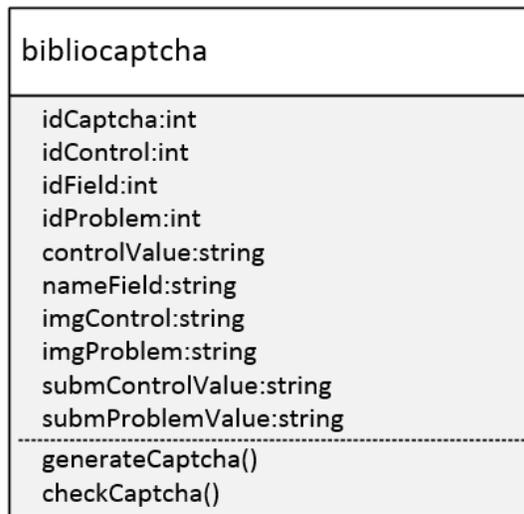
##### **5.4.1.1. Método generateCAPTCHA()**

Proporciona la lógica para generar el *CAPTCHA*, tanto a nivel de la base de datos, así como la generación del *CAPTCHA* en la página web donde se incluye.

#### 5.4.1.2. Método checkCAPTCHA()

Se encarga de revisar los datos ingresados por los usuarios, y los compara contra los valores de control, para determinar si el usuario ingresó la información correctamente.

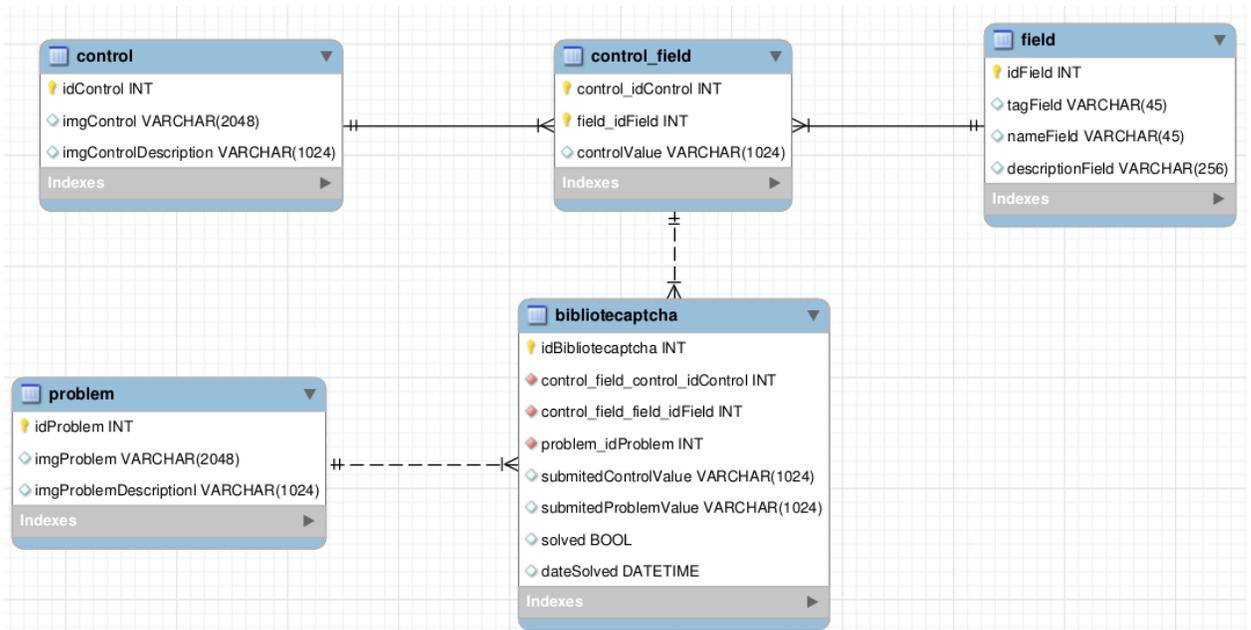
Figura 6. Diagrama de clases



Fuente: elaboración propia.

## 5.4.2. Diagrama Entidad Relación

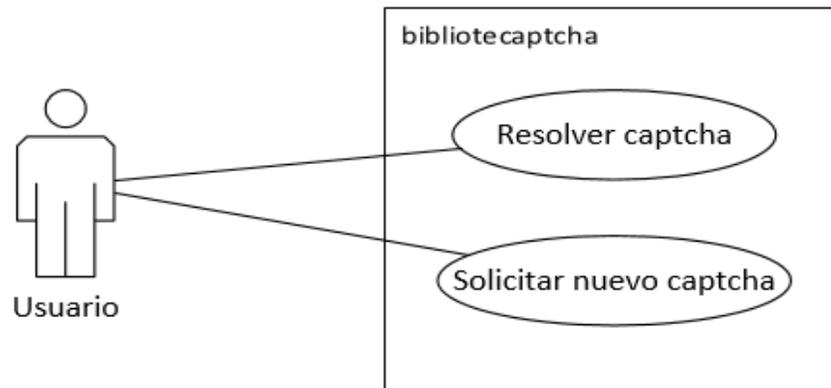
Figura 7. Diagrama de Entidad Relación



Fuente: elaboración propia.

### 5.4.3. Diagrama de casos de uso

Figura 8. Diagrama de casos de uso



Fuente: elaboración propia.

#### 5.4.3.1. Descripción de los casos de uso

- Caso de uso: Resolver *CAPTCHA*
  - Objetivo: Solventar el problema presentado por el *CAPTCHA* para avanzar en el flujo normal de las páginas que el *CAPTCHA* protege.
  - Actores: Usuario.
  - Precondiciones: Ninguna.
  - Flujo básico
    - El sistema muestra las imágenes de *CAPTCHA* a resolver.
    - El usuario ve las imágenes que se le presentan.
    - Determina que información respectiva a la imagen se le solicita. (FA1).
    - El usuario ingresa la información en sus respectivos campos.
    - El sistema ejecuta la validación de los datos ingresados (VA1).

- El sistema redirecciona al usuario a la página que el *CAPTCHA* esté protegiendo.
  - FA1
    - El usuario determina que no quiere o no sabe contestar el *CAPTCHA*, y solicita salir del proceso.
    - El sistema regresa al usuario a la pantalla inicial.
  - VA1
    - El sistema compara los datos ingresados con los datos de control.
    - Si los datos de control no coincide con los datos ingresados, el sistema despliega un mensaje y regresa al paso 1 del flujo básico.
    - El sistema retorna al paso 6 del flujo básico.
- Caso de uso: Solicitar nuevo *CAPTCHA*
  - Objetivo: Solicitar al sistema un nuevo *CAPTCHA* para resolver.
  - Actores: usuario.
  - Precondiciones: El *CAPTCHA* ya debe haber sido generado una vez.
  - Flujo básico
    - El usuario determina que necesita otro *CAPTCHA*
    - El usuario presiona el botón de renovar *CAPTCHA*
    - El sistema le proporciona un nuevo *CAPTCHA*

## 6. PRUEBAS REALIZADAS

Las pruebas realizadas en orden para validar los objetivos propuestos en este trabajo de graduación, son:

- Tratar de implementar un módulo web, que haciendo uso del concepto de *CAPTCHA*, permita la recopilación de información bibliográfica para alimentar un catálogo de una biblioteca
- Usar el módulo web implementado y verificar que la información recopilada sea adecuada para alimentar un catálogo de una biblioteca.

### 6.1. Implementación de prototipo de módulo web

La primera prueba consiste en tratar de implementar el prototipo de un módulo web que permita la recopilación de información bibliográfica para el catálogo de una biblioteca.

Para esto se generaron 3 pantallas (páginas) web, programadas en PHP, apoyadas en una base de datos basada en el esquema definido en el capítulo anterior.

La primera pantalla implementada muestra instrucciones al usuario acerca de lo que encontrará en las pantallas siguientes, y proporciona lineamientos básicos a seguir. Normalmente, un módulo *CAPTCHA* carece de esta pantalla inicial, pues se encuentra instalado en algún sistema, donde funge como mecanismo anti robots, anti *crawler*, etc., como paso intermedio de un formulario de registro o incluso de una pantalla de búsqueda. De esta manera, el incentivo del usuario es obtener su registro, o los resultados de su búsqueda, y se preocupa por contestar correctamente el *CAPTCHA*. En este caso, se carece de

este incentivo, y siendo este un experimento controlado, es necesario proveer de estas guías a las personas que formarán parte del experimento.

Figura 9. **Pantalla inicial del módulo web *CAPTCHA***



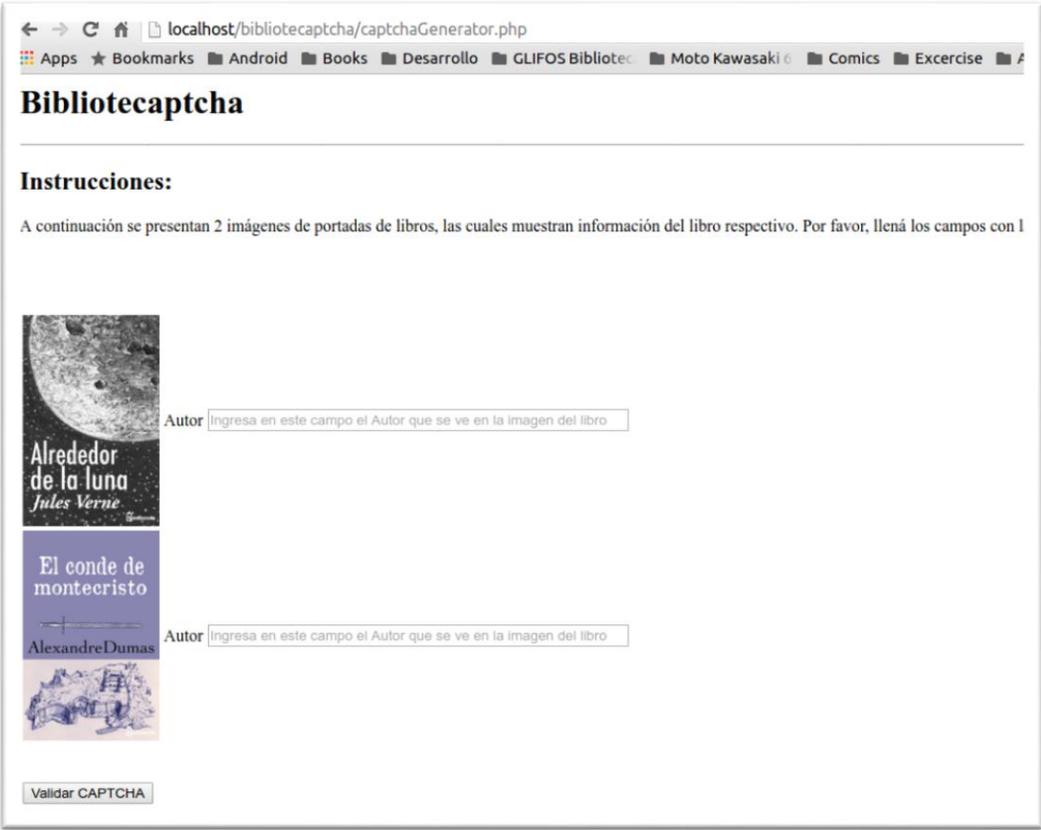
Fuente: elaboración propia.

Posterior a la pantalla inicial, cuando se presiona el botón **Ir**, se redirecciona el flujo del módulo web hacia una pantalla donde se mostrarán dos imágenes, portadas de libros, que contienen información acerca de su título y autor. De forma aleatoria, esta página mostrará una imagen correspondiente a la

parte de control del *CAPTCHA (PIX)* y otra imagen correspondiente al título del cual se desea obtener información.

También de forma aleatoria, se escoge acerca de que campo (título o autor) se solicitará información al usuario. El nombre del campo solicitado aparece a la derecha de la imagen, y a su derecha aparece un campo de entrada donde se puede ingresar la información pertinente al campo del cual se solicita información.

Figura 10. **Pantalla para la resolución del *CAPTCHA (PIX)***



The screenshot shows a web browser window with the address bar displaying 'localhost/bibliotecaptcha/captchaGenerator.php'. The browser's bookmark bar includes 'Apps', 'Bookmarks', 'Android', 'Books', 'Desarrollo', 'GLIFOS Bibliotec', 'Moto Kawasaki', 'Comics', 'Exercise', and 'A'. The page title is 'Bibliotecaptcha'. Below the title, there is a section titled 'Instrucciones:' followed by the text: 'A continuación se presentan 2 imágenes de portadas de libros, las cuales muestran información del libro respectivo. Por favor, llená los campos con 1'. Below this, there are two book covers. The first cover is for 'Alrededor de la luna' by Jules Verne, with the author's name 'Jules Verne' visible at the bottom. To the right of the cover is a text label 'Autor' and an input field containing the text 'Ingresa en este campo el Autor que se ve en la imagen del libro'. The second cover is for 'El conde de montecristo' by Alexandre Dumas, with the author's name 'Alexandre Dumas' visible at the bottom. To the right of the cover is a text label 'Autor' and an input field containing the text 'Ingresa en este campo el Autor que se ve en la imagen del libro'. At the bottom left of the page, there is a button labeled 'Validar CAPTCHA'.

Fuente: elaboración propia.

Por último, la tercera pantalla, posee la lógica del negocio necesaria para validar la información que el usuario ingrese. Revisará que los datos ingresados en el campo de control coincidan con los valores de control almacenados en la base de datos. En la medida que estos sean verdaderos, se procederá a almacenar también la información ingresada, asociada a la imagen de la cual se necesitan obtener los datos bibliográficos.

En caso que la información de control sea válida, la pantalla redireccionará el flujo del módulo web hacia la pantalla que el usuario espera ver (lista de resultados, registro completo, etc.). En caso que la información de control ingresada no sea válida, se regresará al usuario a la pantalla anterior (resolución de *CAPTCHA*) para que resuelva otro *CAPTCHA*.

Figura 11. **Pantalla de procesamiento y despliegue de mensaje**



Fuente: elaboración propia.

## 6.2. Pruebas de uso del módulo web implementado

Se procedió a alojar el módulo web en una dirección IP con acceso público, <http://107.170.78.177/biblioteCAPTCHA>, de manera que se pudiera enviar el link del módulo web para que pudiera ser probado por personas.

Debido a que la prueba se está haciendo bajo demanda, y no por la propia voluntad del *crowd*, es que se usó una pantalla inicial, con una serie de instrucciones y lineamientos.

Vale la pena notar que los links generados para que las personas probaran el módulo, incluían una llave, e.g.:

<http://107.170.78.177/biblioteCAPTCHA/index.php?key=c81e728d9d4c2f636f067f89cc14862c>

De esta forma el link se usa una sola vez. Esto es debido a que normalmente, cuando un usuario resuelve un *CAPTCHA*, lo quiere resolver adecuadamente, en orden para continuar a la página que le interesa. En este caso, al carecer del incentivo para las personas que prueban el módulo, se podría dar el caso que contestaran de forma arbitraria el *CAPTCHA*, de forma reiterativa, así que solo se le proporcionó un link a cada persona, vía correo electrónico.

Se distribuyeron 40 links únicos, a diferentes personas, para su uso. Para lograr un muestreo representativo, a los 40 links se les asignó el mismo título de control, “Alrededor de la luna” de Jules Verne, y el título del cual se desea obtener información bibliográfica, “El conde de Montecristo” de Alexandre Dumas. En ambos casos, lo que varió, fue la selección aleatoria del campo a contestar, título o autor del libro.

Finalmente, al terminar de ingresar la información y enviarla a procesamiento, se despliega un *pop-up* de una encuesta realizada en SurveMonkey, la cual la persona puede decidir contestar o no.

Figura 12. **Pop-up de encuesta de SurveyMonkey**



Fuente: elaboración propia.

Dicha encuesta posee 3 preguntas:

- **1. ¿Te pareció fácil de llenar la información en la página que visitaste, a través de link que te enviamos?**

Tiene como propósito determinar la dificultad de uso de la pantalla de *CAPTCHA*

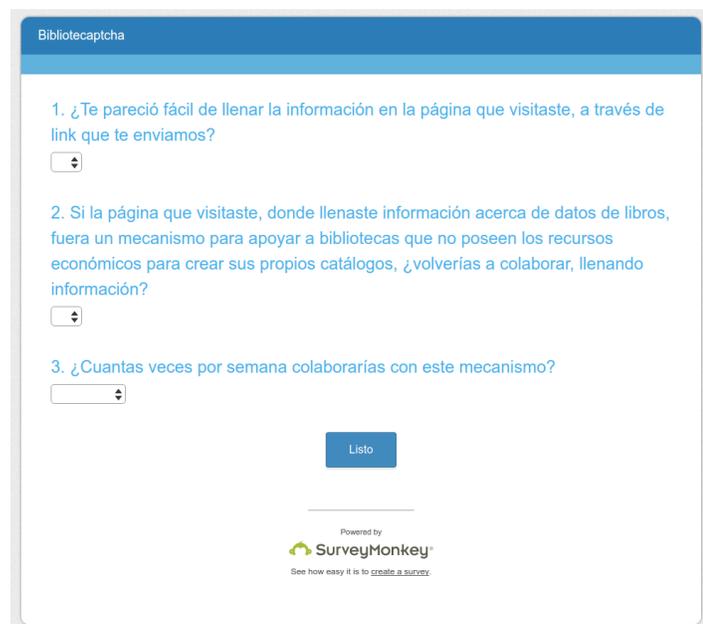
- **2. Si la página que visitaste, donde llenaste información acerca de datos de libros, fuera un mecanismo para apoyar a bibliotecas que**

**no poseen los recursos económicos para crear sus propios catálogos, ¿volverías a colaborar, llenando información?**

El propósito de esta pregunta es determinar la disposición de las personas a colaborar en este proyecto de *crowdsourcing*, de una forma voluntaria, sin un incentivo inmediato.

- **3. ¿Cuántas veces por semana colaborarías con este mecanismo?**  
Esta sirve para determinar el nivel de disposición por parte de las personas en colaborar, podríamos suponer que lo pueden hacer también en la misma medida si hay una compensación por hacerlo (esto al integrar el módulo web con una pantalla de búsqueda o un formulario electrónico).

Figura 13. **Encuesta de usabilidad y disposición de colaboración con el proyecto de *crowdsourcing***



The image shows a survey form titled "Bibliotecaptcha" with a blue header. It contains three questions, each followed by a dropdown menu:

1. ¿Te pareció fácil de llenar la información en la página que visitaste, a través de link que te enviamos?
2. Si la página que visitaste, donde llenaste información acerca de datos de libros, fuera un mecanismo para apoyar a bibliotecas que no poseen los recursos económicos para crear sus propios catálogos, ¿volverías a colaborar, llenando información?
3. ¿Cuántas veces por semana colaborarías con este mecanismo?

At the bottom of the form is a blue button labeled "Listo". Below the button, it says "Powered by SurveyMonkey" with the SurveyMonkey logo and a link: "See how easy it is to create a survey."

Fuente: elaboración propia.



## **7. RESULTADOS**

A continuación se muestran los resultados de las pruebas realizadas anteriormente:

### **7.1. Implementación de prototipo de módulo web**

La implementación del prototipo fue exitosa, se pudo implementar las pantallas descritas anteriormente, así como se pudo diseñar e implementar la base de datos que almacenaría la información del módulo web. El módulo web resultó ser portable, ya que para cambiarlo del ambiente de producción al ambiente de pruebas, solo fue necesario volcar la base de datos (datos y esquema), restituirla en el servidor de base de datos de la dirección de prueba, y copiar el módulo web del ambiente de producción hacia el ambiente de prueba, solo manteniendo su ruta de directorios.

Podría decirse que el resultado de esta prueba fue exitoso, ya que produjo un módulo web portable y funcional.

### **7.2. Pruebas de uso del módulo web implementado**

De los 40 links enviados a personas para que pudieran probar el módulo web, se obtuvieron los siguientes resultados:

Tabla III. **Resultados de generación de CAPTCHAS**

Conteo de CAPTCHAS	Campo a resolver	Resueltos	Procesados
5	Autor	NO	SI
15	Autor	SI	SI
3	Título	NO	NO
2	Título	NO	SI
15	Título	SI	SI

Fuente: elaboración propia.

Para tener una mejor comprensión de los resultados, estableceremos las siguientes definiciones:

- Campo a resolver: Se refiere al campo sobre el cual se ingresará información con el propósito de resolver el CAPTCHA
- Procesados: Se refiere a si los usuarios llegaron a la pantalla de resolución del CAPTCHA e ingresaron información en los campos
- Resueltos: Se refiere a aquellos CAPTCHAS en que sus valores de control fueron ingresados correctamente, y por consecuencia, los valores del libro a procesar tienen un alto grado de fiabilidad.

De la totalidad de CAPTCHAS generados, se puede observar lo siguiente:

- Se produjeron 5 CAPTCHAS que tenían como meta, la resolución del autor del libro en evaluación, los cuales fueron procesados, pero no resueltos satisfactoriamente
- Se produjeron 15 CAPTCHAS que tenían como meta, la resolución del autor del libro en evaluación, los cuales fueron procesados, y fueron resueltos satisfactoriamente

- Se produjeron 3 *CAPTCHAS* que tenían como meta, la resolución del título del libro en evaluación, los cuales no fueron procesados
- Se produjeron 2 *CAPTCHAS* que tenían como meta, la resolución del título del libro en evaluación, los cuales fueron procesados, pero no resueltos satisfactoriamente
- Se produjeron 15 *CAPTCHAS* que tenían como meta, la resolución del título del libro en evaluación, los cuales fueron procesados, y resueltos satisfactoriamente

Viéndolo como un conjunto de totales, podemos generar la siguiente tabla

Tabla IV. **Totales de *CAPTCHAS* generados, procesados y resueltos**

<i>CAPTCHAS</i> de título generados	<i>CAPTCHAS</i> de autor generados	<i>CAPTCHAS</i> de título procesados	<i>CAPTCHAS</i> de autor procesados	<i>CAPTCHAS</i> de título resueltos	<i>CAPTCHAS</i> de autor resueltos
20	20	17	20	15	15

Fuente: elaboración propia.

De la tabla anterior, también podemos observar que de los 40 *CAPTCHAS* generados, 37 fueron procesados y 30 fueron resueltos. Además podemos calcular las proporciones de *CAPTCHAS* procesados y resueltos, de la siguiente forma:

Tabla V. **Porcentajes de CAPTCHAS procesados y resueltos**

<b>% de CAPTCHAS de título generados que fueron procesados</b>	<b>% de CAPTCHAS de autor generados que fueron procesados</b>	<b>% de CAPTCHAS de título procesados que fueron resueltos</b>	<b>% de CAPTCHAS de autor procesados que fueron resueltos</b>
<b>85</b>	100	88.23529412	75

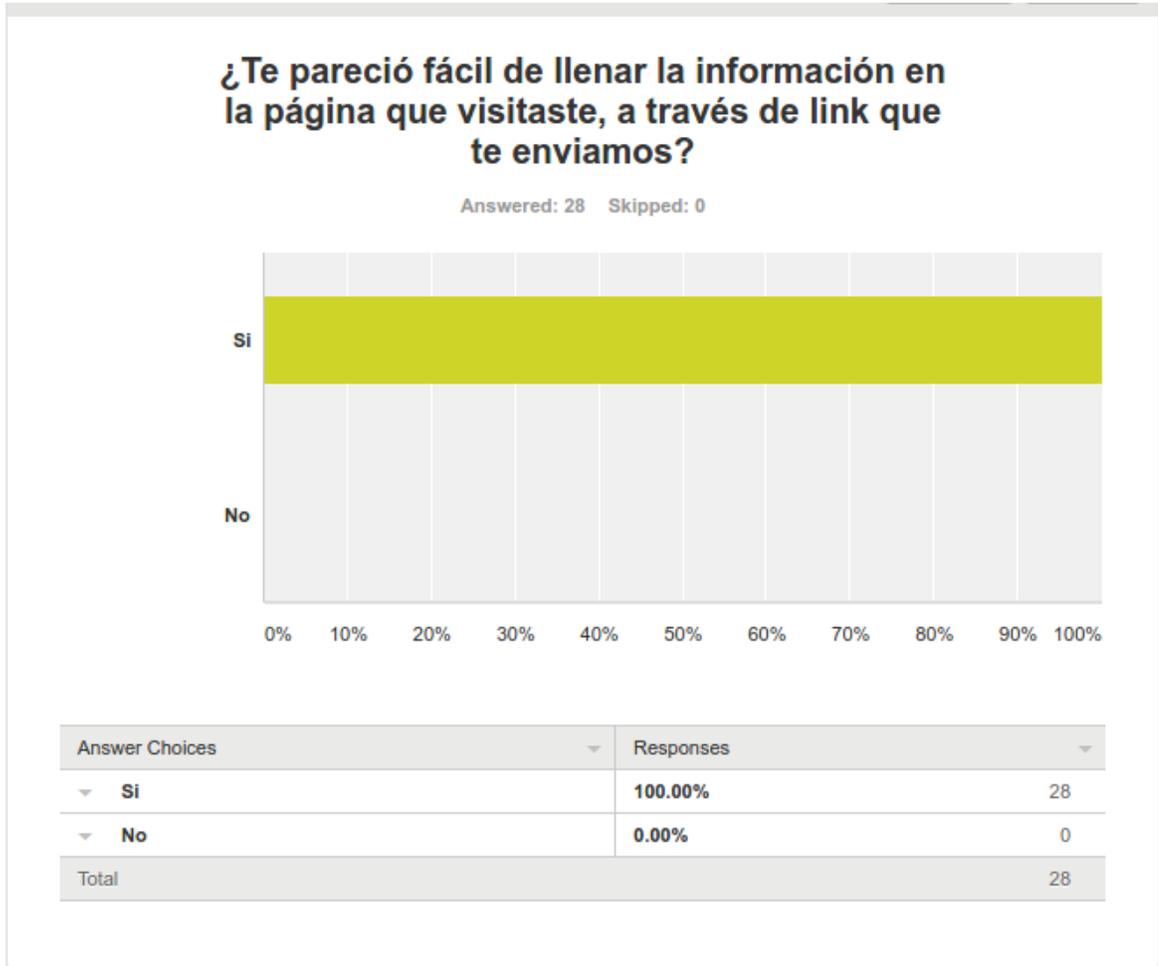
Fuente: elaboración propia.

De la tabla anterior, podemos observar que:

- El 85 % de los CAPTCHAS de títulos generados fueron procesados.
- El 100 % de los CAPTCHAS de autor generados fueron procesados.
- El 88,24 % de los CAPTCHAS de títulos procesados, fueron resueltos satisfactoriamente.
- El 75 % de los CAPTCHAS de autores procesados, fueron resueltos satisfactoriamente.
- El 92,5 % de los CAPTCHAS generados, fueron procesados.
- El 81,08 % de los CAPTCHAS procesados fueron resueltos satisfactoriamente.

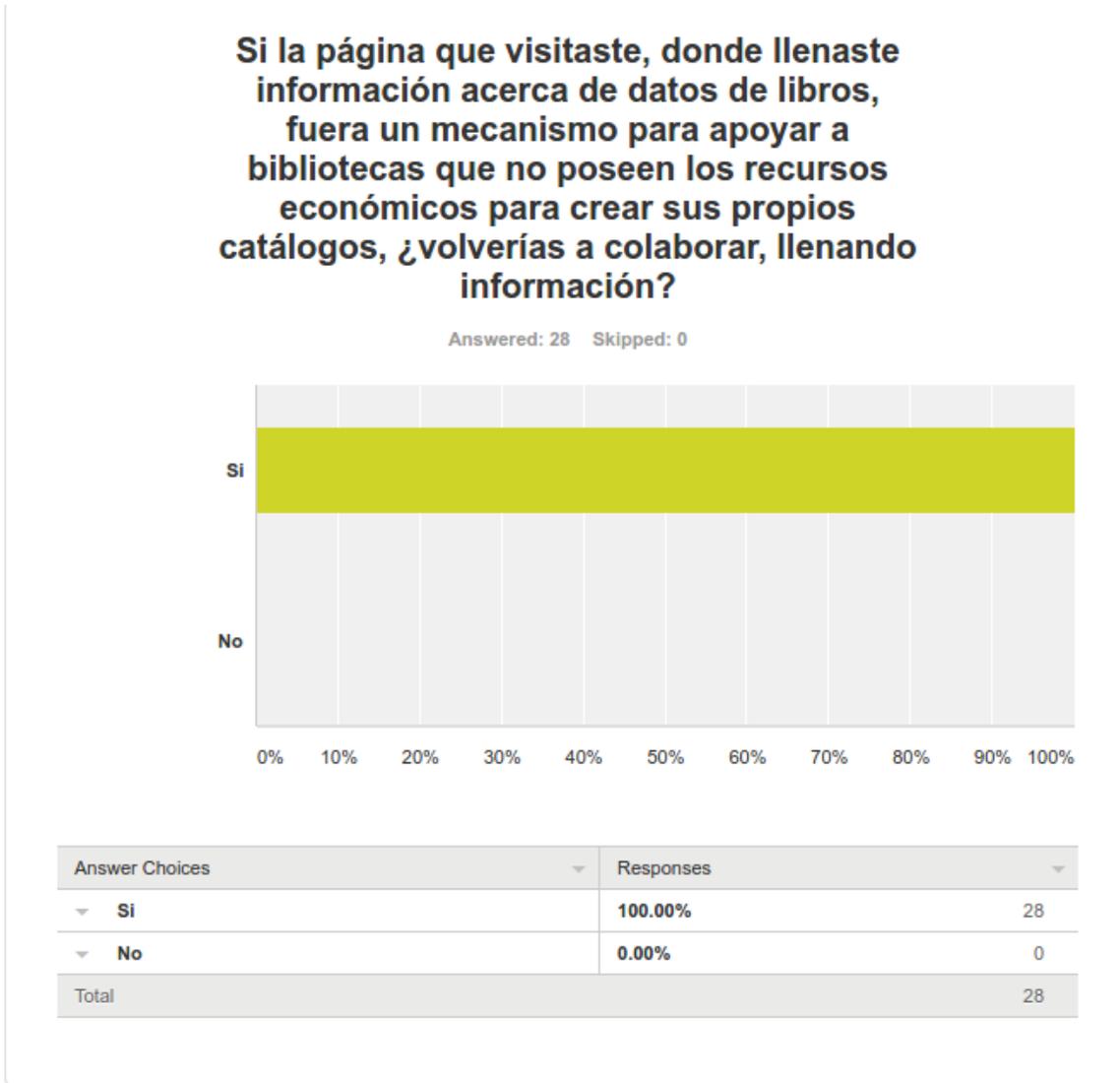
Adicionalmente, de las 40 personas que probaron el módulo web, 28 contestaron la encuesta al final de la prueba, con los siguientes resultados:

Tabla VI. Respuestas a la pregunta Nro.1 de la encuesta



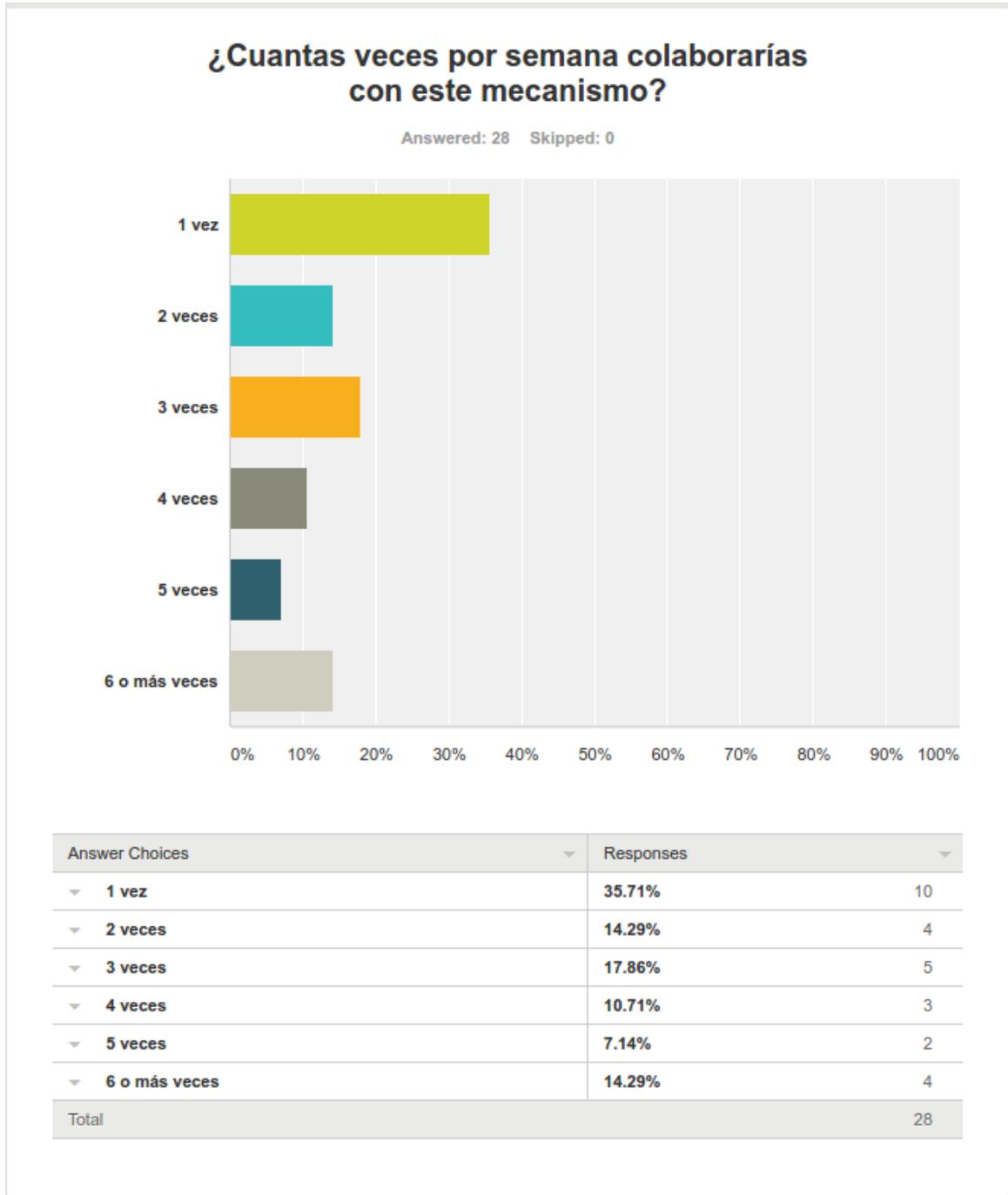
Fuente: elaboración propia.

Tabla VII. **Respuestas a la pregunta Nro.2 de la encuesta**



Fuente: elaboración propia.

Tabla VIII. Respuestas a la pregunta Nro.3 de la encuesta



Fuente: elaboración propia.



## CONCLUSIONES

La catalogación bibliográfica es una forma de organización de la información, que es de gran utilidad, principalmente en la creación de catálogos de consulta, donde las bibliotecas pueden poner sus acervos a disposición de los usuarios.

Por otro lado, el estado actual de los catálogos de las bibliotecas en nuestro país, no es el mejor. Como se podía observar en la estadística, había un buen porcentaje (aproximadamente la tercera parte) de bibliotecas que no tenían catálogos creados, otro porcentaje que los tenía físicos. Que si bien es mejor que nada, no es la mejor opción en una época donde las alternativas electrónicas han probado ser más eficientes y baratas.

El *crowdsourcing* es una herramienta con un gran potencial, las formas de trabajarlo y explotarlo son variadas, especialmente si se logran los motivadores apropiados para involucrar a una comunidad.

Dado lo anterior, se podría concluir que:

1. La obtención de información procesada por una comunidad es viable a través del uso del *crowdsourcing*.
2. Es factible hacer uso del *crowdsourcing* para proveer de procesos a organizaciones que de otra forma no tienen los recursos necesarios para costear dichos procesos.

3. Si los procesos del *crowdsourcing* son los apropiados, se consigue encontrar los motivadores intrínsecos necesarios y el proyecto de *crowdsourcing* tiene un enfoque de beneficio común, se puede lograr una base grande de *solvers* que trabajen en dicha actividad de *crowdsourcing*.
4. Es factible la creación de un módulo web que haga uso del *crowdsourcing* para la obtención de información bibliográfica para la creación de un catálogo.
5. De las estadísticas obtenidas en las pruebas del módulo web, se puede concluir que este es un mecanismo eficiente de obtención de información bibliográfica, pues el 81,08 % de los *CAPTCHAS* que los *testers* trataron de resolver, fueron resueltos de forma satisfactoria.

## RECOMENDACIONES

1. Es recomendable que el proyecto de *crowdsourcing* de *CAPTCHA* como mecanismo para generar catálogos bibliográficos se ejecute sobre plataformas ya establecidas, que sean lo suficientemente solicitadas, para lograr una alta tasa de participación del usuario.
2. Es recomendable que la implementación de este módulo web se amplíe de forma que pueda constituir un *API*, y tenga así una mayor capacidad de acoplamiento y facilidad de implementación.
3. Es recomendable hacer uso del mismo enfoque y propósito de este proyecto, pero explorando otros mecanismos para la obtención de la información, como por ejemplo el uso del protocolo Z39.50, aunado a la consulta por ISBN, para poder hacer catalogación por copia desde otros catálogos que ya tengan la información requerida por la biblioteca que está construyendo su catálogo.



## BIBLIOGRAFÍA

1. BOWERS, Fredson. *Principles of bibliographical description*. Winchester : ST Paul's Bibliographics and Oak Knoll, 1994.
2. BRABHAM, Daren C. *Crowdsourcing*. Cambridge, Massachusetts : The MIT Press, 2013.
3. CHEW, Monica y TYGAR, J. D. *Image recognition CAPTCHAs* [en línea]  
[http://people.ischool.berkeley.edu/~tygar/papers/Image\\_Recognition\\_CAPTCHAs/imagecaptcha-tr.pdf](http://people.ischool.berkeley.edu/~tygar/papers/Image_Recognition_CAPTCHAs/imagecaptcha-tr.pdf). Berkeley, California : University of California, 2004 [Consulta : 28 de octubre de 2016].  
p. 5
4. CUTTER, Charles A. *Rules for a printed dictionary catalogue*. Washington : Govt. Print. Off., 1889.
5. DECI, Edward L., RYAN, Richard M. y WILLIAMS, Geoffrey C. *Need satisfaction and the self-regulation of learning. Learning and Individual Differences* [en línea].  
[http://www.academia.edu/download/44801628/Need\\_satisfaction\\_and\\_the\\_self-regulation20160416-28775-cexrp1.pdf](http://www.academia.edu/download/44801628/Need_satisfaction_and_the_self-regulation20160416-28775-cexrp1.pdf). Elsevier BV, 1996. Vol. 8, no. 3, p. 165-183. p. 167. [Consulta: 1 de octubre de 2016]

6. ESTELLÉS-AROLAS, Enrique y GONZÁLEZ-LADRÓN-DE-GUEVARA, Fernando. *Towards an integrated crowdsourcing definition. Journal of Information Science*. SAGE Publications, 2012, Vol. 38, no. 2, p. 189-200.
7. GAO, H, Yao, D, Liu, H, Liu, X y Wang, L. *A novel image based CAPTCHA using jigsaw puzzle*. IEEE 13th International Conference on, 2010, 351-356.
8. VON HIPPEL, Eric. *Democratizing innovation*. London : MIT Press., 2005.
9. HOWE, Jeff. *Crowdsourcing: A Definition*. [en línea]. [http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing\\_a.html](http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html). 2006. [Consulta: 5 de agosto de 2016].
10. INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS, & INSTITUTIONS. Section on Cataloguing. Standing Committee. *Functional requirements for bibliographic records* [en línea]. [http://www.ifla.org/files/assets/cataloguing/frbr/frbr\\_2008.pdf](http://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf) IFLA Study Group on the Functional Requirements for Bibliographic Records, 2009. p.79. [Consulta: 7 de octubre de 2016].
11. JOUDREY, Daniel N, TAYLOR, Arlene G y MILLER, David P. *Introduction to cataloging and classification*. Santa Barbara, California : Libraries Unlimited, 2015. p.3-5.

12. SIEGEL, Michael D y MADNICK, Stuart E. *A metadata approach to resolving semantic conflicts*. Cambridge, Mass. : Sloan School of Management, Massachusetts Institute of Technology, 1991. p. 3.
  
13. STATEMENT OF INTERNATIONAL CATALOGUING PRINCIPLES. *IFLA Meetings of Experts on an International Cataloguing Code* [en línea]. [http://www.ifla.org/files/assets/cataloguing/icp/icp\\_2009-en.pdf](http://www.ifla.org/files/assets/cataloguing/icp/icp_2009-en.pdf). IFLA Cataloguing Section and IFLA Meetings of Experts on an International Cataloguing Code, 2009. P. 3-4. [Consulta: 3 de octubre de 2016].
  
14. TERRANOVA, Tiziana. *Network Culture: Politics for the Information Age*. London : Pluto Press, 2004. p. 3, p. 62.
  
15. TERWIESCH, Christian y Xu, Yi. *Innovation Contests, Open Innovation, and Multiagent Problem Solving*. [en línea]. [https://www.researchgate.net/profile/Christian\\_Terwiesch/publication/220534500\\_Innovation\\_Contests\\_Open\\_Innovation\\_and\\_Multiagent\\_Problem\\_Solving/links/00b495257417332845000000.pdf](https://www.researchgate.net/profile/Christian_Terwiesch/publication/220534500_Innovation_Contests_Open_Innovation_and_Multiagent_Problem_Solving/links/00b495257417332845000000.pdf)  
 Institute for Operations Research and the Management Sciences : 2008, Vol. 54, no. 9. p. 1529-1543. p. 1542 [Consulta: 16 de agosto de 2016]
  
16. VON AHN, Luis, BLUM, Manuel y HOPPER, Nicholas J. *CAPTCHA: Using hard AI problems for security*. *International Conference on the Theory and Applications of Cryptographic Techniques* [en línea]. <http://repository.cmu.edu/cgi/viewcontent.cgi?article=1142&context>

=compsci Springer, 2003. p. 294-311. [Consulta: 10 de octubre de 2016]

17. VON AHN, L., MAURER, B., MCMILLEN, C., ABRAHAM, D. y BLUM, M. *reCAPTCHA: Human-Based Character Recognition via Web Security Measures*. [en línea].  
[https://www.cs.cmu.edu/~biglou/reCAPTCHA\\_Science.pdf](https://www.cs.cmu.edu/~biglou/reCAPTCHA_Science.pdf) . 2008, Vol. 321, no. 5895, p. 1465-1468. p.1467 [Consulta: 8 de septiembre de 2016]