



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

**DISEÑO DE INVESTIGACIÓN PARA LA CREACIÓN DE UN SISTEMA  
TECNOLÓGICO INNOVADOR PARA LA DISTRIBUCIÓN GRATUITA DE  
CONTENIDOS TRADICIONALMENTE IMPRESOS GENERANDO INGRESOS A  
LOS DUEÑOS DE LOS DERECHOS DE AUTOR A TRAVÉS DE PUBLICIDAD**

**Mauricio Haroldo Rojas Morales**

Asesorado por el MSc. Ing. Luis Fernando Espino Barrios

Guatemala, febrero de 2014

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERÍA

**DISEÑO DE INVESTIGACIÓN PARA LA CREACIÓN DE UN SISTEMA  
TECNOLÓGICO INNOVADOR PARA LA DISTRIBUCIÓN GRATUITA DE  
CONTENIDOS TRADICIONALMENTE IMPRESOS GENERANDO INGRESOS A  
LOS DUEÑOS DE LOS DERECHOS DE AUTOR A TRAVÉS DE PUBLICIDAD**

PROTOCOLO DE TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA  
FACULTAD DE INGENIERÍA  
POR

**MAURICIO HAROLDO ROJAS MORALES**

ASESORADO POR EL MSC. ING. LUIS FERNANDO ESPINO BARRIOS

AL CONFERÍRSELE EL TÍTULO DE

**INGENIERO EN CIENCIAS Y SISTEMAS**

GUATEMALA, FEBRERO DE 2014

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA  
FACULTAD DE INGENIERÍA



**NÓMINA DE JUNTA DIRECTIVA**

DECANO	Ing. Murphy Olympo Paiz Recinos
VOCAL I	Ing. Alfredo Enrique Beber Aceituno
VOCAL II	Ing. Pedro Antonio Aguilar Polanco
VOCAL III	Inga. Elvia Miriam Ruballos Samayoa
VOCAL IV	Br. Walter Rafael Véliz Muñoz
VOCAL V	Br. Sergio Alejandro Donis Soto
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

**TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO**

DECANO	Ing. Murphy Olympo Paiz Recinos
EXAMINADOR	Ing. Juan Álvaro Díaz Ardavin
EXAMINADOR	Ing. Oscar Alejandro Paz Campos
EXAMINADOR	Ing. Pedro Pablo Hernández Ramírez
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

## 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:

**DISEÑO DE INVESTIGACIÓN PARA LA CREACIÓN DE UN SISTEMA  
TECNOLÓGICO INNOVADOR PARA LA DISTRIBUCIÓN GRATUITA DE  
CONTENIDOS TRADICIONALMENTE IMPRESOS GENERANDO INGRESOS A  
LOS DUEÑOS DE LOS DERECHOS DE AUTOR A TRAVÉS DE PUBLICIDAD**

Tema que me fuera asignado por la Dirección de la Escuela de Estudios de Posgrado, con fecha 23 de febrero de 2013.



**Mauricio Haroldo Rojas Morales**



**USAC**  
TRICENTENARIA  
Universidad de San Carlos de Guatemala

Escuela de Estudios de Postgrado  
Facultad de Ingeniería  
Teléfono 2418-9142 / Ext. 86226



000025

AATT-MTIPP-0001-2014

Guatemala, 23 de enero de 2014.

Director:  
Marlon Antonio Pérez Turk  
Escuela de Ingeniería en Ciencias y Sistemas  
Presente.

Estimado Director:

Reciba un atento y cordial saludo de la Escuela de Estudios de Postgrado. El propósito de la presente es para informarle que se ha revisado los cursos aprobados del primer año y el Diseño de Investigación del estudiante **Mauricio Haroldo Rojas Morales** con carné número **2002-30358**, quien opto la modalidad del "PROCESO DE GRADUACIÓN DE LOS ESTUDIANTES DE LA FACULTAD DE INGENIERÍA OPCIÓN ESTUDIOS DE POSTGRADO". Previo a culminar sus estudios en la **Maestría de Tecnologías de la información y la Comunicación**.

Y si habiendo cumplido y aprobado con los requisitos establecidos en el normativo de este Proceso de Graduación en el Punto 6.2, aprobado por la Junta Directiva de la Facultad de Ingeniería en el Punto Decimo, Inciso 10.2, del Acta 28-2011 de fecha 19 de septiembre de 2011, firmo y sello la presente para el trámite correspondiente de graduación de Pregrado.

Sin otro particular, atentamente,

"Id y enseñad a todos"

MSc. Luis Fernando Espino Barrios  
Asesor (a)

MSc. Ing. Marlon Antonio Pérez Turk  
Coordinador de Área  
Aplicación y transferencia tecnológica

M.Sc. Luis Fernando Espino Barrios.  
Ingeniero en Ciencias y Sistemas  
Colegiado No. 9145

Dra. Mayra Virginia Castillo Montes  
Directora  
Escuela de Estudios de Postgrado



Cc: archivo  
/la

E  
S  
C  
U  
L  
A  
  
D  
E  
  
C  
I  
E  
N  
C  
I  
A  
S  
  
Y  
  
S  
I  
S  
T  
E  
M  
A  
S

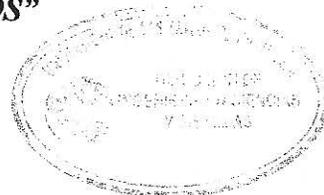
UNIVERSIDAD DE SAN CARLOS  
DE GUATEMALA



FACULTAD DE INGENIERÍA  
ESCUELA DE 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 **“DISEÑO DE INVESTIGACIÓN PARA LA CREACIÓN DE UN SISTEMA TECNOLÓGICO INNOVADOR PARA LA DISTRIBUCIÓN GRATUITA DE CONTENIDOS TRADICIONALMENTE IMPRESOS GENERANDO INGRESOS A LOS DUEÑOS DE LOS DERECHOS DE AUTOR A TRAVÉS DE PUBLICIDAD”**, realizado por el estudiante MAURICIO HAROLDO ROJAS MORALES, 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, 26 de febrero 2014



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 Ciencias y Sistemas, al trabajo de graduación titulado: **DISEÑO DE INVESTIGACIÓN PARA LA CREACIÓN DE UN SISTEMA TECNOLÓGICO INNOVADOR PARA LA DISTRIBUCIÓN GRATUITA DE CONTENIDOS TRADICIONALMENTE IMPRESOS GENERANDO INGRESOS A LOS DUEÑOS DE LOS DERECHOS DE AUTOR A TRAVÉS DE PUBLICIDAD**, presentado por el estudiante universitario: **Mauricio Haroldo Rojas Morales**, procede a la autorización para la impresión del mismo.

IMPRÍMASE.

Ing. Murphy Olympo Paiz Reinos  
Decano



Guatemala, febrero 2014

/cc

## **ACTO QUE DEDICO A:**

### **Mis padres**

Haroldo Rojas y Mei Lay Morales por su apoyo a lo largo de toda mi vida.

### **Mi esposa**

Claudia Archila ya que sin su compañía todo sería mucho más complicado

### **Mi familia**

Por ser siempre un lugar seguro al cual regresar, por querer mi bienestar y jamás demostrar egoísmo. A Chyntia Rojas, por eliminar la soledad en momentos difíciles y por aceptarme sin el menor prejuicio. También a aquellos que aun sin estar físicamente cerca, jamás abandonarán mi alma.

### **Mis amigos**

Pues no podría vivir sin aquellos ideales que en mi mente forjaron. Me mostraron un nuevo mundo; puedo decir que existo por todo lo que me han dado sin que yo haya otorgado ni lo más mínimo en retorno. Si no fuera por ustedes ni siquiera respirar sería viable: Umberto Eco, Jostein Gaarder, John Tolkien, Dave Grohl, el grupo CLAMP e incontables, que son los gigantes de cuyos hombros tomo la altura para contemplar un poco mejor las estrellas.

## **AGRADECIMIENTOS A:**

### **Mis catedráticos**

Por compartir su conocimiento sin egoísmo en cada etapa de mi formación académica. Al profesor Luis Macario, Erick Abac, Erick Arango, Hugo Pineda, Marlon Pérez, Alberto Vetorazzi, Jorge Mazariegos, Javier Gramajo y muchos otros que iluminaron mi visión del mundo.

### **Universidad de San Carlos de Guatemala**

Por la oportunidad de conocer a excelentes personas y por introducirme en inmejorable forma a la educación superior.

## ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES .....	VII
LISTA DE SÍMBOLOS .....	IX
GLOSARIO .....	XI
RESUMEN .....	XVII
1. INTRODUCCIÓN .....	1
2. ANTECEDENTES .....	3
3. PLANTEAMIENTO DEL PROBLEMA .....	9
3.1. Pregunta central .....	10
3.2. Preguntas auxiliares .....	11
4. JUSTIFICACIÓN .....	13
5. OBJETIVOS .....	15
6. NECESIDADES A CUBRIR Y ESQUEMA DE SOLUCIÓN.....	17
6.1. Esquema de la solución.....	18
6.1.1. Investigación y definición de modelo de negocio....	18
6.1.2. Desarrollo de plataforma .....	19
6.1.3. Pruebas y conclusiones de estudio .....	19
6.1.4. Creación de propuesta de negocio .....	19

7.	ALCANCES.....	21
7.1.	Perspectiva investigativa.....	21
7.1.1.	Alcances descriptivos.....	21
7.1.2.	Alcances explicativos .....	21
7.2.	Perspectiva técnica .....	22
7.3.	Perspectiva de resultado.....	22
8.	MARCO TEÓRICO .....	25
8.1.	Libros y revistas en línea, consumo en aumento .....	25
8.2.	Principales mecanismos para acceder contenidos en línea....	25
8.3.	Copyright en los contenidos tradicionalmente escritos.....	26
8.3.1.	Modelo de negocio actual de las editoriales y otros productores de contenidos .....	27
8.3.2.	Riesgos para productores de contenidos ante la distribución ilegal de productos de información.....	27
8.4.	Manejo de contenidos en línea .....	28
8.5.	La publicidad en internet como método alternativo de adquisición de ingresos.....	28
8.5.1.	Colocación de la publicidad.....	29
8.5.2.	Publicidad personalizada y referente a contenidos específicos .....	30
8.6.	Técnicas para la generación de palabras clave .....	31
8.6.1	Generación de palabras clave según hábitos de usuario, segmentación .....	31
8.6.2.	Palabras clave de contenido combinadas con dominios específicos .....	32
8.6.3.	Generación de palabras clave en tiempo real .....	33
8.7.	Extracción de contenidos en formato PDF .....	33

8.8.	Modelo de negocio de nuevos medios basados en publicidad .....	34
8.8.1.	Sitios para compartir archivos.....	34
8.8.2.	Blogs.....	34
8.8.3.	Redes sociales .....	35
8.9.	Tecnologías para optimización de buscadores para garantizar acceso a publicaciones.....	35
8.9.1.	Técnicas de optimización de motores de búsqueda SEO .....	36
8.9.1.1.	Análisis de textos para generación de palabras clave.....	36
8.9.1.1.1.	<i>Big data, map reduce,</i> Hadoop.....	36
8.10.	Paradigma social .....	37
8.10.1.	Manejo de redes de usuarios para ampliar la base de usuarios y para generación de meta datos .....	37
9.	PROPUESTA DE ÍNDICE DE CONTENIDOS .....	39
10.	METODOLOGÍA.....	41
10.1.	Variables e indicadores .....	41
10.1.1.	Problema: distribución ilegal gratuita de contenidos tradicionalmente impresos por internet.....	41
10.1.2.	Solución.....	41
10.1.3.	Subvariables .....	41
10.1.3.1.	Ganancias por contenidos gratuitos publicados en internet para publicadores .....	42

10.1.3.2.	Ganancias por contenidos gratuitos publicados en internet para poseedores del derecho de autor.....	42
10.1.3.3.	Pérdidas por contenidos gratuitos publicados en internet para dueños del copyright.....	42
10.1.4.	Matriz de operacionalización de variables.....	42
10.2.	Diseño, tipo y alcances del estudio.....	44
10.3.	Fases del estudio.....	44
10.3.1.	Observación.....	45
10.3.2.	Análisis del entorno.....	45
10.3.3.	Técnicas para recolección de información.....	45
10.3.4.	Revisión documental.....	46
10.3.5.	Trabajo de campo.....	46
10.3.6.	Población y muestra.....	46
11.	TÉCNICAS DE ANÁLISIS DE INFORMACIÓN.....	49
12.	CRONOGRAMA.....	51
12.1.	Tareas.....	52
13.	FACTIBILIDAD DEL ESTUDIO.....	55
13.1.	Factibilidad operativa.....	55
13.2.	Recursos humanos.....	55
13.3.	Acceso a información.....	55
13.4.	Permisos.....	55
13.5.	Equipo.....	56
13.6.	Infraestructura.....	56

13.7.	Aspectos éticos, legales y de seguridad industrial .....	56
13.8.	Conclusión de la factibilidad operativa .....	56
13.9.	Factibilidad técnica .....	57
13.9.1.	Recursos desde el punto de vista técnico .....	57
13.9.2.	Recursos tecnológicos.....	57
13.9.3.	Extensión del trabajo de investigación.....	58
13.9.4.	Tiempo disponible para realizar lo que se propone .....	58
13.9.5.	Conclusión de la factibilidad técnica .....	58
13.10.	Factibilidad económica .....	59
13.10.1.	Recursos económicos .....	59
13.10.2.	El costo del tiempo .....	59
13.10.3.	El costo de la realización .....	59
13.10.4.	El costo de adquirir nuevos recursos.....	60
13.10.5.	Costo total.....	60
13.10.6.	Conclusión de la factibilidad económica .....	61
14.	BIBLIOGRAFÍA .....	63



## ÍNDICE DE ILUSTRACIONES

### FIGURAS

1.	Cronograma .....	51
----	------------------	----

### TABLAS

I.	Variables .....	43
II.	Tipo y alcances del estudio .....	44
III.	Tareas programadas .....	52
IV.	Costos .....	60



## LISTA DE SÍMBOLOS

<b>Símbolo</b>	<b>Significado</b>
\$	Dólar estadounidense
%	Porcentaje
Q	Quetzal



## GLOSARIO

<b>Algoritmo</b>	Recibe el nombre de algoritmo el grupo finito de operaciones establecidas de forma ordenada que posibilita la solución de un problema.
<b>Amazon.com</b>	Empresa de origen estadounidense, que brinda el servicio de venta de productos a través de internet.
<b>Banner</b>	Anuncio publicitario que se coloca en páginas web; por lo regular de forma rectangular.
<b>Big data</b>	<i>Big data</i> o bien “grandes datos” se refiere a los sistemas que manejan grandes cantidades de datos de diversos tipos.
<b>Blog</b>	Es una página web, donde uno o varios autores tienen la libertad de publicar textos o imágenes de manera cronológica.
<b>Bookos.org</b>	Sitio en internet que funcionaba como una librería electrónica donde sus usuarios podían subir y consultar libros en línea.

<b>Buscador</b>	Es una herramienta electrónica que permite a los usuarios de internet realizar búsquedas de datos almacenados en diferentes páginas web.
<b>Clic</b>	En términos informáticos, se refiere a la acción de pulsar algún botón del mouse o ratón de la computadora.
<b>Copyright</b>	<i>Copyright</i> o derecho de autor, se refiere a la regulación legal que protege a los autores por la creación de una obra musical, científica, artística, literario o didáctica, la cual puede estar publicada o ser inédita.
<b>Dispositivos móviles</b>	Aparatos pequeños, con acceso permanente o intermitente a una red, con memoria limitada.
<b>E-book</b>	Libro en formato electrónico que puede ser consultado a través de una computadora o de otros dispositivos. Actualmente se distribuyen en su mayoría por medio de internet y se adquieren in librerías en línea.
<b>Facebook</b>	Sitio Web creado por Mark Zuckerberg, para la interacción entre diferentes redes sociales.
<b>Freaders</b>	Nombre dado a la plataforma diseñada para realizar el presente estudio, su nombre proviene de las palabras inglesas “Free”, gratis y “Readers”, lectores.

<b>Gawker</b>	Es una empresa de comunicación en línea americana y blog, creado por Nick Denton.
<b>Google</b>	Es una compañía de origen estadounidense que se dedica a crear productos y servicios para internet, software y otras tecnologías. Es mejor conocida por su buscador, google.com.
<b>Hadoop</b>	Sistema de código abierto que permite almacenar, procesar y analizar una gran cantidad de datos.
<b>Html5</b>	Sus siglas significan: <i>Hyper Text Markup Language</i> , se utiliza para organizar y presentar contenidos para páginas web. En su versión 5 el lenguaje tiene capacidades de formato elaboradas y con alta interacción entre contenidos multimedia.
<b>Indización</b>	Es un proceso a través del cual se extraen conceptos representativos de un documento, dando como resultado un índice que puede ser utilizado como herramienta de búsqueda.
<b>Internet</b>	Es un grupo de redes de comunicación conectadas entre sí y de uso globalizado.
<b>Map reduce</b>	Es el modelo de programación que utiliza Google para dar soporte a la computación paralela en grandes cantidades de datos.

<b>Mega</b>	Sitio de internet para subir y descargar archivos, su uso es controversial debido a contenidos piratas subidos por sus usuarios. Su antecesor Megaupload fue clausurado por problemas legales.
<b>Metadata</b>	Es información sobre la propia información, es decir, descripciones sobre un conjunto de datos dado.
<b>Motor de búsqueda</b>	Sistema a través del cual se pueden localizar archivos en servidores web, también recibe el nombre de buscador.
<b>Mp3</b>	Es un formato de audio digital que se utiliza en computadoras o dispositivos móviles de audio.
<b>Netflix</b>	Compañía que proporciona servicio en línea para ver películas y series de televisión a través de internet disponibles con una cuota de suscripción mensual.
<b>Nube</b>	Se refiere al almacenamiento de grandes cantidades de datos en servidores en internet.
<b>Online</b>	Su traducción al español es: en línea y se refiere al estado de una persona conectada a alguna red de internet.
<b>P2P</b>	<i>Peer to Peer</i> , se refiere a redes entre pares donde un usuario se conecta directamente a otro sin necesidad de un servidor intermedio.

<b>PDF</b>	Es el acrónimo de <i>Portable document format</i> o bien su traducción al español: formato de documento portátil, el cual fue desarrollado por Adobe Systems, es una forma de almacenamiento de documentos digitales.
<b>Pdf2htmlex</b>	Software de código abierto que transforma archivos del formato PDF al formato HTML.
<b>Piratería</b>	Se refiere al uso no autorizado o prohibido de obras protegidas por las leyes de derechos de autor.
<b>Rocket eBook</b>	Es un lector de libros digitales portátil del tamaño de un libro de papel.
<b>SEO</b>	Sus siglas en inglés significan <i>Search Engine Optimization</i> , es un grupo de medidas para optimizar el posicionamiento de una página web, en cualquier buscador.
<b>Servidor</b>	El término en informática se refiere a un ordenador remoto (nodo) que proporciona los datos solicitados por los navegadores de otros ordenadores.
<b>Softbook</b>	Fue el primer dispositivo creado específicamente para leer libros electrónicos en 1998.

<b>Software</b>	Se le llama así al conjunto de los componentes lógicos que son necesarios para la realización de tareas en un ordenador.
<b>Statscrop</b>	Es un sitio de Internet que ofrece estadísticas acerca de otros sitios en línea con detalles tales como valor aproximado y popularidad.
<b>Twitter</b>	Sitio web que utilizan sus usuarios para escribir mensajes de texto pequeños (hasta 140 caracteres) que aparecen en la página principal de usuario.
<b>Weblogs</b>	También conocido como Blog, es un sitio web donde se publican textos o anotaciones por uno o varios autores, de manera cronológica.

## RESUMEN

La piratería es un problema que aqueja a todos los contenidos disponibles en internet, debido en gran parte a la facilidad con la que se pueden reproducir y distribuir; facilidades que surgen como consecuencia del avance en las tecnologías de la información.

En el caso de los contenidos tradicionalmente impresos como los libros y las revistas, el problema tiene igual incidencia y en ocasiones aún más grave ya que al tratarse de texto, la distribución es mucho más sencilla. Las palabras codificadas en forma digital ocupan poco espacio en disco haciendo que un libro completo pueda transmitirse en cuestión de segundos.

Con el auge de internet también surgen diversas estrategias dirigidas a prevenir la distribución ilegal de los contenidos, por ejemplo, en algunos países surgen leyes para penalizar la publicación no autorizada, otros contenidos son codificados buscando evitar su copia y existen sitios web que crean algoritmos para encontrar copias ilegales. Sin embargo los sitios con contenidos ilícitos continúan surgiendo e incluso desarrollando métodos para ganar dinero a través de aquella distribución ilegítima.

Uno de los problemas que ocasiona este negocio ilegal basado en la distribución de contenidos, es que sus creadores no reciben remuneración alguna por su trabajo, remuneración que recibirían si el contenido hubiera sido consumido de manera legítima. Como consecuencia peligra la continuidad de la producción de ciertos textos; por ejemplo, disminuyen las razones para

escribir un nuevo libro o una editorial puede verse obligada a cesar la producción de una revista.

La solución que se plantea a este problema, es ofrecer al público una alternativa igual o superior a lo que ofrecen los sitios actuales que publican contenidos ilegales, pero de una manera legal y que retribuya a los poseedores de los derechos de autor. Para conseguir esta alternativa se emplearán técnicas para generar ingresos mediante la publicidad, basándose en las palabras contenidas en los materiales tradicionalmente impresos.

Las técnicas para producir ganancias se emplean ya en medios digitales tales como los blogs, donde escritores publican contenidos para luego obtener dinero cuando los lectores ingresan en algún anuncio del sitio. Lo que se busca es hacer converger tecnologías que exploten el texto que ya se encuentra en los escritos; los sitios ilegales actuales emplean la publicidad para generar ingresos, pero no sacan el potencial oculto en cada una de las páginas de información que se ofrecen. Al emplear dicho potencial se generarán ingresos que irán directamente a los poseedores del *copyright* que empleen la plataforma para la publicación.

El estudio basado en el presente diseño de investigación, evaluará la efectividad de la solución que se plantea mediante pruebas ejecutadas sobre un sistema funcional que genere publicidad en base a la información contenida en los textos publicados. El objetivo es estimar el beneficio que podrían obtener los usuarios de esta nueva tecnología, para así corroborar si se trata de una respuesta viable para la publicación de medios tradicionalmente impresos ante los nuevos retos que plantean las tecnologías de la información.

# 1. INTRODUCCIÓN

Freaders es un sistema que proveerá contenidos gratuitos en internet, reportando ingresos para los poseedores del *copyright* a través de mecanismos de publicidad implementados de manera personalizada para los usuarios, según lo que lean y sus características demográficas.

La razón de ser de Freaders, es retribuir a los productores de contenidos para que puedan seguir entregando materiales de alta calidad. Además, Freaders contará con características que permitirán que la información ubicada en textos pueda ser encontrada y accedida fácilmente; esta característica no solo beneficiará a los nuevos productores de información, sino que permitirá que se generen ingresos a partir de contenidos que actualmente están ociosos, por ejemplo, libros que no están actualmente disponibles para buscar en sus páginas desde internet y que por lo tanto no son consumidos por sus lectores potenciales.

Para que este sistema funcione, se requiere la convergencia de una serie de tecnologías en una sola plataforma disponible en internet. Estos elementos comprenden la publicidad en internet, la optimización para buscadores y los paradigmas de redes sociales.

Para conseguir crear este sistema, se realizará una investigación que comprenderá un estudio del trasfondo requerido para una innovación tecnológica en el área de la publicación. En esta parte del estudio se investigará sobre las mejores tecnologías para cumplir con la visión de

Freaders; estas tecnologías se relacionan con la publicación en línea, la publicidad en internet y el análisis de textos.

Con este trasfondo establecido, la investigación incluirá el análisis y diseño del sistema innovador así como pruebas de la plataforma creada funcionando en línea.

Finalmente, la investigación contendrá la evaluación de los beneficios potenciales obtenidos según las pruebas funcionales donde se concluirá que tanto sería el beneficio proyectado para quienes decidan usar la plataforma para publicar.

## 2. ANTECEDENTES

La tecnología del siglo XXI, ha causado una transición desde los medios tradicionales de información hacia medios electrónicos consumidos por medio de computadoras y dispositivos móviles.

Ha ocurrido con la música, desde los discos de vinilo a los mp3, con las películas desde las cintas hasta Netflix. En el caso de los medios escritos, la humanidad se encuentra precisamente en el momento en que el cambio empieza a volverse cada vez más real y popular, del papel a la pantalla.

Los libros electrónicos o E-books aparecieron en 1971 con el Proyecto Gutenberg de Michael Hart's, sin mucha relevancia ya que aún se carecía de medios de difusión funcionales y era necesario estar frente a la computadora para poder leer, incómodo comparado con un portátil libro de papel. Internet creció y surgieron los dispositivos portátiles para leer E-books a partir de 1998 con el Rocket eBook y el Softbook, el medio empezó a ser cada vez más aceptado (Wu & Mitchell, 2010).

Estos E-books llegan a los lectores a través de compras por internet y por redes para compartir archivos, en ocasiones de manera legal y en otras ignorando los derechos del autor. Algunos sitios ocupan una zona gris de legalidad; por ejemplo bookos.org, un sitio que ofrece actualmente más de dos millones de libros con la posibilidad de que los mismos usuarios suban el contenido.

Un mercado importante se mueve en sitios como este; el 21 de enero de 2013 el sitio Statscrop publicó un reporte en el que valuaban a bookos.org en \$ 101 750,00, una cantidad interesante si se considera que el sitio tiene solo siete meses de estar disponible en internet. Este ingreso proviene de 7 140 visitantes y 48 552 páginas vistas por día que pueden generar ingresos de \$ 182 diarios (Statscrop.com, 2013).

En contraste, el reporte de transparencia generado por Google en referencia a bookos.org, revela que tienen más de 6 000 archivos con solicitud de ser removidos por violar derechos de autor de los cuales solo algunos han sido removidos, además, Google revela un porcentaje de indización menor al 1 % (Google.com, 2013). Eso significa que solo esa cantidad del contenido del sitio puede ser encontrada por el popular buscador.

Mientras tanto, las editoriales han tenido que adaptarse al nuevo modelo de negocio nacido a raíz de los E-books, todo esto de manera abrupta y sin que haya una directriz generalizada en el medio. Se enfrentan retos donde los costos de producción decrecen al desaparecer los gastos de impresión y almacenaje, además de reducirse los de distribución. Debido a esto una versión digital es más barata y el porcentaje de ganancias recibidas por cada libro para cada autor pasa de un 25 % en el papel a un 8 % por copia digital (Costa-Knufinke, 2010). Se continúa sin un consenso para la respuesta a esta disrupción.

Otro mercado interesante existe en los blogs, los populares reciben tráfico importante de visitas que les provee de ingresos en base a la publicidad; esto si y solo si tienen lectores fieles que consumen sus contenidos constantemente y si cuentan con una estrategia en la que los anuncios reciban clics de lectores. Un solo autor puede generar ingresos de hasta \$ 5 000,00 mensuales por

producir contenido de calidad accesible desde buscadores en internet (Smith, 2005).

Una estrategia empleada por empresas dedicadas a la publicidad como Google –Ad Words– es revisar las palabras clave de los blogs y de esta forma enviar publicidad relacionada con el contenido generando ingresos por clic desde el anunciante hasta el proveedor del contenido (Colwell, 2010). Los medios escritos en papel y los E-books ofrecidos exclusivamente a través de software propietario o en sitios cerrados, desperdician información que bien podría estar generando ingresos a los escritores y editores por medio de metadatos, es decir, palabras clave que los hicieran accesibles desde buscadores y que a su vez generaran ingresos a través de programas como Ad Words.

Otros medios que incluso ya existen de manera electrónica, también sufren de este desperdicio al existir únicamente en forma de imágenes. Esto pasa principalmente con revistas, por ejemplo en alguna publicación subida como imagen para preservar su formato original podría aparecer la historia de algún acontecimiento; texto que no podría ser indexado o analizado sin sus correspondientes metadatos.

La generación y optimización de metadatos es un tema ampliamente tratado cuando se busca la optimización de sitios para motores de búsqueda, una estrategia de mercadeo en línea. Su importancia estiba en que el 90 % del tráfico de internet pasa por los motores de búsqueda; es una práctica común el que las personas empleen como punto de partida su buscador favorito al iniciar cualquier investigación o actividad en internet, se estima que para el 2012 el volumen de búsquedas era de 300 billones al mes (Patel, Patel, & Patel, 2012).

Por lo tanto, no basta con tener un texto publicado en internet si no hay una forma en la que este texto aparezca entre los primeros 40 resultados de los buscadores más populares, lo más probable es que dicho texto jamás sea consultado.

Existen varias tecnologías relacionadas con la generación de metadatos relevantes para determinados textos, siendo destacadas aquellas que pueden segmentar las palabras clave generadas al introducir a la ecuación datos personales de los usuarios; esta metodología proviene del uso de campos predefinidos de texto que se añaden a algoritmos conocidos para la generación de palabras clave (Kuppusamy & Aghila, A Model for Personalized Keyword Extraction from Web Pages using Segmentation, 2012).

Además de ser la parte fundamental para el sostenimiento del proyecto, la publicidad dirigida también cuenta con varios estudios y tecnologías relacionadas; un estudio psicológico en el que se analiza si realmente los usuarios de una página de internet miran los anuncios y si los recuerdan, ha demostrado que los usuarios recuerdan más aquellos anuncios congruentes con el contenido que están consumiendo que aquellos anuncios incongruentes (Hervet, Guérard, Tremblay, & Chrourou, 2011).

El presente estudio se basa en estos antecedentes al notar la disrupción de los medios digitales y que detrás de los contenidos tradicionalmente impresos se encuentran autores que podrían beneficiarse con otros modelos de negocio y tecnologías que hasta ahora les son ajenos; como la accesibilidad desde buscadores de internet por medio de metadatos que optimicen la publicidad capaz de generar ingresos.

Se trata de una innovación tecnológica diferenciada de otros proyectos de publicación de contenidos en línea, puesto que busca retribuir directamente a los poseedores del *copyright* aun cuando se publique de manera gratuita. Los estudios que se relacionan específicamente son aquellos que tratan el tema del *copyright* en la era digital, como en el caso del trabajo de Veiksa en 2012 (Veiksa, 2012), donde se evalúa la validez de los mecanismos tradicionales de publicación y de *copyright* ante los cambios que han ocurrido en los medios de distribución. En este caso el acercamiento será práctico y no teórico como en el citado estudio.



### 3. PLANTEAMIENTO DEL PROBLEMA

Existe una distribución y consumo no controlados de contenidos tradicionalmente impresos, de manera gratuita en internet sin generar ingresos para los dueños legítimos del *copyright*.

Por un lado es un problema para los dueños del *copyright* ya que no perciben ingresos, estos se quedan únicamente con quienes distribuyen sus contenidos. Por otro lado es un problema para los lectores ya que en ocasiones los textos ofrecidos no son fiables, además de ser difíciles de conseguir debido a las barreras creadas en intentos de frenar la distribución ilegal.

El 21 de enero de 2013 el sitio Statscrop publicó un reporte en el que se valuaba a bookos.org, un sitio que provee E-books gratis sin retribuir a los dueños del *copyright*, en \$ 101 750,00, esto cuando el sitio tenía solo siete meses de estar disponible en internet. Este ingreso proviene de 7 140 visitantes y 48 552 páginas vistas por día capaces de generar ingresos de \$ 182 diarios. Si no se plantea una solución, estos flujos de ingresos se mantendrán sin que los autores perciban algún beneficio lo que conlleva un detrimento en la producción de contenidos, de todos esos ingresos diarios la mayor parte podría corresponder a los dueños del *copyright*.

Específicamente, el problema radica en la distribución ilegal en internet de contenidos tradicionalmente impresos.

Este problema, de carácter global, surge desde el inicio de internet y se agrava con las mejoras tecnológicas que facilitan la distribución de la información. Las consecuencias van desde problemas económicos para los productores de contenidos hasta el endurecimiento de leyes relacionadas con *copyright*, que podrían afectar otros factores relativos a la libertad de comunicación en internet.

Mientras tanto, emergen negocios que aprovechan los contenidos para generar ingresos sin otorgar nada a los creadores. Los productores de contenidos están perdiendo oportunidades con alto potencial de dar a conocer lo que generan a la vez que incrementan sus ingresos.

El problema se limita a la falta de aplicación de técnicas para análisis de datos y generación de metadatos, para la producción de publicidad en una plataforma donde se puedan publicar contenidos tradicionalmente impresos. Es una situación que degenera en la falta de un modelo de negocio para los poseedores de *copyright*, que emplee las ventajas tecnológicas disruptivas que provee la publicación en línea como la monetización en base a publicidad.

Por lo tanto, el problema a resolver es específicamente la pérdida económica que sufren los productores de contenidos a raíz de la distribución ilegal.

### **3.1. Pregunta central**

¿Cómo puede una innovación tecnológica ayudar a reducir las pérdidas en que incurren los propietarios de *copyright* por la distribución ilegal gratuita de contenidos tradicionalmente impresos en internet?

### 3.2. Preguntas auxiliares

- ¿Es posible mostrar publicidad directamente relacionada con lo que se está leyendo para generar una cantidad mayor de ingresos sin cobrar directamente al lector?
- ¿Cómo pueden aportar valor los metadatos de un texto determinado al emplearse como palabras clave en internet?
- ¿Cómo puede lograrse un mayor volumen de lectores para un texto en internet?
- ¿Pueden los dueños del *copyright* obtener beneficios al publicar contenidos con acceso gratuito para los lectores?
- ¿Cuál sería el beneficio obtenido al publicar de manera gratuita con publicidad en internet?
- ¿Qué estrategias de manejo de usuarios pueden emplearse para que una innovación tecnológica que otorga contenidos gratuitos funcione y sea sostenible?



## 4. JUSTIFICACIÓN

Se puede ofrecer contenidos tradicionalmente impresos de manera gratuita a los usuarios sin afectar negativamente a los productores de contenidos, incluso es posible generar un flujo de ingresos para dichos productores a través del consumo libre empleando estrategias de publicidad correctamente orientadas.

Por medio de una investigación en la línea de la innovación tecnológica es posible crear una plataforma donde los contenidos puedan publicarse generando ingresos para sus productores.

Por otro lado, los sitios que publican material sin retribuir a los creadores de los contenidos, constantemente remueven títulos debido a solicitudes de los dueños del *copyright*, por lo que los contenidos disponibles tienden a mermar a menudo, no hay una disponibilidad continua ni asegurada para los usuarios con respecto a lo que se puede encontrar.

Estas mismas restricciones hacen que los contenidos de los libros publicados en algunos de estos sitios no puedan ser buscados desde internet, ocasionando que algunos textos que podrían ser interesantes para determinadas investigaciones pasen desapercibidos.

Así, esta investigación en la línea de tecnologías de la información y la comunicación para innovar en industria es necesaria al promover una mejora tecnológica en la que convergen el análisis de datos, la publicación en línea, herramientas de búsqueda en internet y la publicidad para beneficiar a

escritores, investigadores y otros propietarios de *copyright* al permitirles distribuir sus materiales a través de internet de manera expedita y obteniendo beneficios económicos.

Al existir amplia accesibilidad a contenidos tradicionalmente impresos, existirá también una mejor distribución del conocimiento que beneficiará a lectores y a escritores.

## 5. OBJETIVOS

### **General**

Diseñar un sistema tecnológico innovador que reduzca las pérdidas por publicaciones ilegales al generar ingresos a propietarios de derechos de autor, mediante la publicidad personalizada, mostrando contenidos tradicionalmente impresos de manera gratuita a través de una plataforma en internet.

### **Específicos**

1. Diseñar una plataforma en internet para publicación de contenidos tradicionalmente impresos con la capacidad de mostrar anuncios relacionados a lo que se está leyendo.
2. Aplicar mecanismos de análisis de textos para la generación de palabras claves que puedan usarse para mostrar anuncios y optimizar la compatibilidad con motores de búsqueda.
3. Aplicar técnicas de optimización para motores de búsqueda (SEO) de tal manera que los contenidos publicados en la plataforma web sean fácilmente encontrados y consultados.
4. Explicar la relación entre la publicación gratuita en internet y la obtención de ganancias por medio de publicidad dirigida.

5. Evaluar los beneficios potenciales obtenidos por los propietarios de *copyright* al participar promoviendo contenidos a través del sistema diseñado.
  
6. Implementar una plataforma web con capacidad para manejar usuarios, presentar contenidos, analizarlos y generar publicidad específica para cada individuo.

## **6. NECESIDADES A CUBRIR Y ESQUEMA DE SOLUCIÓN**

Al ofrecer contenidos gratis de manera legal se enfrenta directamente la imperante necesidad global de encontrar una solución al problema de la piratería a través de internet.

Se proveerá de un nuevo esquema para poder publicar contenidos en línea que a su vez permitirá generar ingresos por medio de la publicidad. La solución podrá ser consumida por pequeños escritores o por editoriales, ubicando los documentos en un sistema fácil de encontrar y de navegar desde los buscadores de internet.

De este modo, no solo se solventa la necesidad de publicar en internet de una manera más eficiente; a la vez, se provee de una nueva estrategia de mercadeo donde ya no es necesario obtener ingresos directamente del lector ni tampoco a través de un sitio de publicaciones masivo, sino que cada autor puede ganar solo con que sus contenidos sean consumidos.

La autopublicación podrá generar ingresos, incluso si estos son muy bajos, constituirán una diferencia notoria contra los modelos de publicación actuales donde se requiere que un lector efectivamente pague el total del libro. La necesidad de este nuevo modelo de negocio, surge debido a que al comprar un libro, el lector corre el riesgo de desconocer si el contenido del mismo realmente será de su utilidad y agrado, razón por la cual se dificultan las ventas de nuevos escritores; en contraparte, esta solución plantea que los contenidos de los libros puedan ser buscados y consultados contra demanda y que cada una de estas consultas tenga la posibilidad de generar ganancias. Es decir que

un lector podrá buscar el tema de su interés y acceder directamente a la sección que busca del libro sin necesidad de comprar todo el texto.

Es un estudio y ensayo indispensable en la actualidad, dado el incremento rápido de nuevas tecnologías, no puede continuarse sin un modelo sostenible para la distribución de conocimiento en todas las esferas que cuenten con acceso a internet. Las tecnologías están disponibles pero se requiere la investigación para aportar congruencia entre las mismas y definir su correcto funcionamiento en conjunto.

## **6.1. Esquema de la solución**

La solución se plantea a través de los siguientes elementos:

### **6.1.1. Investigación y definición de modelo de negocio**

Se definirá un modelo de negocio innovador donde se provean contenidos tradicionalmente impresos gratis, de manera legal y generando ingresos a los poseedores del *copyright*.

A través de una investigación documental se establecerán las tecnologías necesarias para crear una plataforma que pueda soportar lo exigido por el plan de negocio, es decir, generar ingresos a través de publicidad en torno a contenidos tradicionalmente impresos.

### **6.1.2. Desarrollo de plataforma**

Por tratarse de un producto de software, es viable crear la plataforma para hacer pruebas incluso antes de un lanzamiento comercial o de contar con alianzas estratégicas.

Se creará el sistema donde se pueda:

- Subir contenidos
- Generar palabras clave automáticamente
- Generar publicidad en base a palabras clave
- Navegar contenidos
- Manejar usuarios
- Manejar información de hábitos de uso del sistema por parte de usuarios
- Generar publicidad personalizada

### **6.1.3. Pruebas y conclusiones de estudio**

Se realizarán pruebas funcionales sobre la plataforma para concluir si el sistema será funcional en el mercado.

### **6.1.4. Creación de propuesta de negocio**

Con los resultados previos, se generará una propuesta de negocio con el objetivo de financiar el proyecto o planificar un lanzamiento comercial. El alcance de esta investigación no abarca dicho lanzamiento.



## **7. ALCANCES**

### **7.1. Perspectiva investigativa**

Desde el punto de vista de la investigación, se propone establecer un punto de convergencia entre diversos conceptos relacionados con la tecnología que en la actualidad no se funden para proveer una respuesta al problema planteado.

#### **7.1.1. Alcances descriptivos**

La investigación abarcará lo necesario para crear un sistema tecnológico innovador para publicación de contenidos en línea, así como lo relacionado a las diferentes técnicas requeridas para generar publicidad dirigida a partir de materiales tradicionalmente impresos y su aplicación en una plataforma en internet.

En este sentido, se realizará una investigación documental acerca de estas técnicas y mecanismos preexistentes que se pueden emplear en tal innovación.

#### **7.1.2. Alcances explicativos**

Tras contar con el sistema para pruebas, se estudiarán los datos que produzca el mismo para así predecir los beneficios potenciales que se podrán obtener en base a publicidad al publicar contenidos tradicionalmente impresos.

Esto permitirá explicar y documentar la relación entre la publicación en este tipo de sistema y las ganancias obtenidas.

## **7.2. Perspectiva técnica**

El estudio involucrará el desarrollo de una plataforma de prueba por lo que desde el punto de vista técnico se involucra el uso de tecnologías necesarias para el análisis de textos, publicación en internet y aquellas que sean necesarias para la generación de publicidad en línea. También, se involucra el estudio de los resultados generados tras contar con la plataforma de pruebas.

## **7.3. Perspectiva de resultado**

En cuanto a la solución, el resultado será una plataforma en internet donde se podrán publicar textos que serán analizados automáticamente para producir anuncios relacionados con el contenido a través de la generación de palabras clave relevantes. Además, esta plataforma manejará usuarios cuyos datos serán combinados en el análisis de generación de palabras clave para producir anuncios aún más relevantes.

Cada usuario también podrá publicar sus propios contenidos bajo el modelo de autopublicación.

La expectativa es contar con un sistema funcional donde se puedan realizar pruebas en línea de lo plasmado en el estudio. Como resultado de las pruebas se presentarán estadísticas revelando los beneficios potenciales por uso del sistema y también en relación al análisis de textos.

Estos resultados serán útiles para los potenciales usuarios del sistema así como para los escritores y productores de contenidos tradicionalmente impresos en general ya que se plantea como una respuesta sostenible al problema de la piratería en internet.



## **8. MARCO TEÓRICO**

### **8.1. Libros y revistas en línea, consumo en aumento**

El consumo de libros electrónicos y otros contenidos tradicionalmente impresos por medio de internet se disparó debido, en gran medida, al advenimiento de los dispositivos móviles que eliminaron barreras previas tal como la obligación de tener que estar en un lugar fijo para poder consumirlos.

Además, la disponibilidad de internet de banda ancha permite la obtención de los contenidos haciendo que sea mucho más fácil adquirir una copia digital que una física; tal cómo se esboza en el estudio de Wu & Mitchell al denotar el importante incremento en los lectores tras el establecimiento de estos métodos de obtención en los dispositivos móviles por parte de Google (Wu & Mitchell, 2010) .

Esto implica que existe un volumen de lectores de medios en línea cada vez mayor, tanto para contenidos obtenidos de forma legal como ilegal.

### **8.2. Principales mecanismos para acceder contenidos en línea**

El alza en la disponibilidad de dispositivos móviles ha ocasionado un aumento en el consumo de los libros electrónicos debido a la comodidad que ofrecen para su consumo.

Las prácticas para obtenerlos van desde tiendas como Amazon.com y hasta redes de distribución de archivos p2p; de acuerdo con el estudio realizado por Cordón-García *et al*, estas redes ya se encuentran en uso incluso en los mercados de habla hispana (Cordón-García, Gómez-Díaz, & Alonso-Arévalo, 2011).

La mayor parte de los lectores de contenidos en línea de la actualidad, realizan la lectura desde dispositivos móviles en los cuales los libros y otros documentos son descargados desde sitios de distribución de archivos de persona a persona.

### **8.3. *Copyright* en los contenidos tradicionalmente escritos**

Legalmente, los contenidos cuentan con un propietario de *copyright* que por ende puede autorizar o no el copiado de la información. Sin embargo este derecho se ve fácilmente obtenido debido a las tecnologías de la información modernas que hacen que los datos se puedan copiar y transmitir casi de inmediato. En esta nueva sociedad de la información se hace imperativo el surgimiento de nuevos mecanismos para contrarrestar la piratería, incluso se consideran cambios a las normas usuales de *copyright* para apegarse a las capacidades actuales.

“Los sitios de distribución ilegal de música y filmes ilegales jamás serán totalmente eliminados, esto es tecnológicamente imposible” (Veiksa, 2012). Más que eliminar los sitios de piratería, se hace necesario adaptarse a las nuevas tecnologías de tal forma que las pérdidas sean menores.

### **8.3.1. Modelo de negocio actual de las editoriales y otros productores de contenidos**

En la actualidad las editoriales funcionan distribuyendo contenidos de los escritores obteniendo parte de la ganancia y distribuyendo otra parte a los escritores.

Según un estudio realizado con editoriales españolas, en este modelo los autores ganan un promedio del 25 % del valor de cada libro vendido (Costa-Knufinke, 2010).

El problema de la distribución ilegal afecta al negar este porcentaje a los autores, además de poner en riesgo la continuidad del negocio de los editores y de los distribuidores de contenidos.

### **8.3.2. Riesgos para productores de contenidos ante la distribución ilegal de productos de información**

El riesgo se sitúa en que existe una caída en las ventas que produce un desestímulo en la producción de nuevos contenidos. Al dejar de ser un negocio redituable existen empresas que salen del mercado ocasionando una escasez en ciertos ámbitos.

También, estas salidas del mercado ocasionan desempleo, pérdidas para los inversionistas y en general un estancamiento en el crecimiento del conocimiento al reducirse la distribución de la información.

#### **8.4. Manejo de contenidos en línea**

Existen varios estándares para la definición de metadatos en diversos tipos de documentos para la organización en librerías digitales. La adopción de un sistema como este puede ser determinante para clasificar correctamente los libros y otros archivos contenidos en el sistema; por ejemplo, el estándar definido por la Biblioteca Digital de California establece que etiquetas debería tener un libro, como por quién, dónde, cuándo y fue escrito (California Digital Library Services Advisory group, 2007).

En una alternativa a la tradicional compra de libros, este tipo de información es indispensable pues permite poner a disposición de los lectores los contenidos que realmente podrían interesarles en un momento dado, promoviendo la agilidad en la adquisición de conocimiento e incrementando los ingresos potenciales por cada página leída.

#### **8.5. La publicidad en internet como método alternativo de adquisición de ingresos**

En internet, al igual que en varios medios de comunicación, uno de los principales mecanismos para monetizar los contenidos es la publicidad.

Sin embargo, internet permite ir un paso más adelante; en la televisión los anuncios son estandarizados y fijos para toda la población que vea la programación mientras que en internet cada usuario puede ver un anuncio distinto, apegado a sus preferencias personales.

### **8.5.1. Colocación de la publicidad**

Se ha demostrado que el éxito de la publicidad en internet depende de varios factores que van desde el contenido del propio anuncio, el tipo de página donde se encuentra y hasta el tipo de contenido que el usuario se encuentra consultando. Por ejemplo, un estudio realizado por Gho & Jang-han en Corea del Sur revela que un usuario mirará más los anuncios cuando se encuentra explorando, o buscando algo, que cuando ya está en la página a la que deseaba dirigirse (Gho & Jang-han, 2011). De tal suerte, es mucho más efectivo colocar anuncios en las páginas previas al consumo de un contenido en sí que el colocarlos al lado de dicho contenido, o en todo caso, emplear estrategias para que el usuario vuelva al estado en que se encuentra cuando está explorando.

Las estrategias de publicidad en línea involucran datos tales como los tipos y tamaños de *banners online* más usuales y el tema de la publicidad personalizada tal como se enumera en el estudio español realizado por Martínez-San-Emeterio en 2010 (Martínez-San-Emeterio, 2010). Al igual que los anuncios televisivos fueron la respuesta para distribuir gratuitamente contenidos de la industria fílmica, también este tipo de estrategias pueden ser la solución para la libre distribución de medios impresos sin perder ingresos para los autores.

Hay que considerar además, el cambio de paradigma actual en el cual los usuarios consultan cada vez más los contenidos desde dispositivos móviles que en la mayoría de los casos son comparativamente cada vez más pequeños. En este paradigma también el modelo de la publicidad en línea debe cambiar, un banner no puede usarse igual pues en muchos casos ocuparía toda la pantalla del móvil por sí solo.

La respuesta planteada se relaciona con nuevas innovaciones en la manera de anunciar; Thompson presenta un artículo en el que se plantea como ha cambiado el negocio de la publicidad a la par del surgimiento de un nuevo medio de comunicación, el móvil (Thompson, 2013). Un teléfono inteligente tiene una pantalla más pequeña pero aun así se emplea para leer contenidos tradicionalmente impresos; la manera en que la publicidad se le presenta ya no es con imágenes estáticas en el pie de la página, si no interactivamente al ingresar o salir del contenido.

#### **8.5.2. Publicidad personalizada y referente a contenidos específicos**

Actualmente se cuenta ya con publicidad inteligente que aparece según los contenidos que el usuario está visitando. Por ejemplo, si el usuario visita un blog en el que se escribe sobre videojuegos es posible que le aparezca un anuncio del videojuego de moda en un *banner* o en formato de texto.

Esto se consigue tras un estudio de los contenidos publicados por parte del anunciante. Google emplea esta estrategia y se basa en palabras claves encontradas con frecuencia en el sitio que distribuye los contenidos, "... los anunciantes apuntan únicamente a usuarios interesados" (Colwell, 2010) es el nivel de segmentación más granular que se puede conseguir en los medios actuales.

Los blogs a los que Google otorga anuncios son contenidos generados actualmente por escritores en internet, sin embargo no es imposible migrar esta misma tecnología hacia los libros tradicionales. Los libros que actualmente se consumen de manera gratuita e ilegal podrían contener anuncios Google, generados en base a sus contenidos que en muchos casos son aún más

interesantes o están mejor fundamentados que aquellos de los escritores contemporáneos en línea.

## **8.6. Técnicas para la generación de palabras clave**

Para contar con palabras clave que describan los contenidos y permitan realizar búsquedas que además asocien lo encontrado con cada usuario, es necesario emplear una serie de estrategias que generen una base de metadatos.

### **8.6.1 Generación de palabras clave según hábitos de usuario, segmentación**

Es posible generar palabras clave específicas para cada usuario dependiendo de los hábitos que demuestran al consultar los libros. En este caso las palabras clave generadas podrían almacenarse en el perfil de cada usuario para así generar anuncios a nivel de sitio según los intereses generales de cada persona. Este tipo de estudios se ha realizado incluso en bibliotecas físicas demostrando potencial para concluir sobre futuras acciones de los usuarios; tal es el caso de la investigación dirigida por Lim *et al*, en una biblioteca de Japón donde estas palabras clave en libros de papel indicó exitosamente la tendencia de los siguientes libros que los lectores consultarían (Lim, Jain, Nishihara, Kimura, & Ohsawa, 2012).

El emplear esta estrategia, permite generar anuncios que cuentan con una probabilidad mayor de generar ingresos para los poseedores del *copyright*.

### **8.6.2. Palabras clave de contenido combinadas con dominios específicos**

Existen modelos para extraer de manera automatizada, palabras clave de páginas web agregando datos personalizados según información recopilada de los usuarios. La utilidad de este acercamiento consiste en su uso para combinar las palabras clave generadas por el contenido, que podrían producirse cuando se suba el documento, junto con las palabras clave del usuario actual, de tal manera que puedan obtenerse los mejores anuncios posibles. Un modelo definido por Kuppusamy & Aghila emplea una “bolsa de perfil de usuario” con palabras claves específicas para cada persona que usa, en una combinación, las palabras de los documentos para así extraer metadatos más efectivos (Kuppusamy & Aghila, A Model for Personalized Keyword Extraction from Web Pages using Segmentation, 2012).

En este sentido, estudios psicológicos sugieren que los usuarios recuerdan más aquellos anuncios congruentes con el contenido que están consumiendo que aquellos anuncios incongruentes. Hervert *et al*, realizaron pruebas en este sentido al emplear un sensor que detectaba cuando los anuncios al lado de un texto eran vistos por lectores según el movimiento de los ojos; posteriormente se preguntaba a los sujetos sobre el contenido de la publicidad observada demostrando que, en efecto, lo más recordado era aquello que tenía alguna relación con el texto que se leía (Hervert, Guérard, Tremblay, & Chrourou, 2011) .

Los sitios actuales de publicación de contenidos no optimizan su publicidad con este tipo de técnicas impidiendo que se realice el potencial de lo publicado.

### **8.6.3. Generación de palabras clave en tiempo real**

También es posible generar palabras clave relevantes para determinado dominio de datos. Un mecanismo habitual para hacerlo es mediante el aprendizaje empleando diversas prácticas, el inconveniente de este tipo de métodos es que se requiere haber realizado varios análisis previos para que el sistema otorgue datos cualitativamente correctos.

Una alternativa definida por un modelo de Chen & Lin en 2011, es emplear una búsqueda en Google para extraer la relevancia de las palabras dadas, garantizando de esta manera que los resultados tengan importancia y que además se consiga dicho resultado en tiempo real (Chen & Lin, 2011).

Así, se denota que es factible generar un dominio de palabras clave sobre cualquier texto incluso sin contar con algún estudio preliminar sobre las palabras ahí contenidas.

### **8.7. Extracción de contenidos en formato PDF**

Una de las maneras de subir contenidos a la red es por medio del formato PDF que permite crear un documento tal y como si se hubiese impreso en papel. Existen diversos mecanismos para convertir un documento PDF a formato html5, un formato compatible con el análisis de contenido y con una amplia gama de dispositivos para visualización en navegadores de internet; estos mecanismos emplean varias perspectivas, útiles dependiendo de la solución que se pretenda implementar, por ejemplo, una solución que no requiere resultados inmediatos puede utilizar un mecanismo lento de procesamiento mientras que en otros casos se requiere procesamiento en tiempo real.

Una comparación entre estos métodos se encuentra en el sitio de conversión de formatos de Wang, PDF2htmlEX, en este caso las características del proyecto hacen que sea preferible la conversión durante la creación del documento (Wang, 2013).

De esta forma puede mejorarse y ampliarse la facilidad que existe ya actualmente para la publicación en línea, pero no simplemente para publicar sino también para generar ingresos a los propietarios de los derechos de autor.

## **8.8. Modelo de negocio de nuevos medios basados en publicidad**

En internet, existen en la actualidad medios que basan sus ingresos en la publicidad. Sus estrategias son similares a las empleadas por medios tradicionales como la prensa escrita pero empleando las ventajas que provee la tecnología.

### **8.8.1. Sitios para compartir archivos**

Sitios como Mega o Bookos.org, se basan en compartir archivos para atraer una amplia base de usuarios que luego genere ingresos mediante suscripciones o bien mediante la publicidad.

### **8.8.2. Blogs**

Los blogs basan la mayoría de sus ingresos en publicidad que se genera según los contenidos de los cuales se escribe, además se enfocan en optimizar sus características para poder ser encontrados a través de los diversos navegadores.

Para 2005, sitios como Weblogs y Gawker “generaban entre 200 y 2 000 dólares al mes” en publicidad para sus escritores de blogs (Smith, 2005).

Cada texto publicado en línea tiene el mismo potencial que estos blogs para obtener ingresos de la publicidad.

### **8.8.3. Redes sociales**

Las redes sociales obtienen ingresos mediante la publicidad y en ocasiones como en el caso de Facebook o Google +, cuentan con sus propios generadores de anuncios que se monetizan cuando los anunciantes solicitan un espacio.

Esto es viable para este tipo de sitios debido a la gran base de usuarios con que cuentan o en el caso de Google, por contar de antemano con una división dedicada al tema.

### **8.9. Tecnologías para optimización de buscadores para garantizar acceso a publicaciones**

Al realizar una búsqueda en internet, los motores de búsqueda tienen una forma jerárquica de mostrar los resultados basándose en técnicas que aproximen lo que se mostrará a lo que los usuarios están investigando. Se pueden optimizar los contenidos a ofrecer para que puedan tener lugares privilegiados en esa jerarquía de búsquedas.

### **8.9.1. Técnicas de optimización de motores de búsqueda SEO**

Las técnicas de optimización de motores de búsqueda, llamadas SEO por las siglas en inglés “Search Engine Optimization” se basan en contar con palabras clave y contenidos relevantes para los motores de búsqueda.

#### **8.9.1.1. Análisis de textos para generación de palabras clave**

Las palabras clave que se requieren en las técnicas SEO en muchos casos son redactadas por las personas que crean los contenidos. Una mecánica más efectiva y automática es analizar texto para luego aplicar minería de datos y así conseguir palabras clave relevantes.

##### **8.9.1.1.1. *Big data, map reduce, Hadoop***

El análisis de grandes cantidades de información, de libros completos o de datos complejos era un reto técnico casi insalvable pocos años antes de la realización del presente estudio.

Luego, los procesadores tuvieron un avance importante, así como los mecanismos de almacenamiento aunado al surgir de nuevos paradigmas como los servicios en la nube.

Todos estos factores hacen que el poder de procesamiento ya no sea un obstáculo para analizar grandes cantidades de información. Es posible adquirir un servicio de nube elástica y así contar con procesadores trabajando en demanda.

El uso de estas herramientas puede emplearse para el análisis de textos junto con la información de usuarios para generar palabras claves, entregables a un proveedor de anuncios que a su vez otorgue una oferta interesante para cada cliente.

## **8.10. Paradigma social**

Con el auge de las redes sociales, como Facebook y Twitter, se creó un nuevo paradigma en internet donde las personas tienden a compartir los contenidos que consumen y también a pasar gran parte de su tiempo en línea empleando alguno de estos sitios.

### **8.10.1. Manejo de redes de usuarios para ampliar la base de usuarios y para generación de meta datos**

Un mecanismo para ampliar la base de usuarios en una plataforma en internet es usar características sociales tales como permitir a los usuarios citar frases de los contenidos que leen y compartirlas directamente en Facebook o en Twitter; de esta forma los contactos del usuario que probablemente compartan algunos intereses tienen la posibilidad de conocer también la plataforma.



## 9. PROPUESTA DE ÍNDICE DE CONTENIDOS

ÍNDICE DE ILUSTRACIONES

LISTA DE SÍMBOLOS

GLOSARIO

RESUMEN

PLANTEAMIENTO DEL PROBLEMA Y PREGUNTAS DE INVESTIGACIÓN

OBJETIVOS

MARCO METODOLÓGICO

INTRODUCCIÓN

1. CAPÍTULO 1 – TRASFONDO DE UNA INNOVACIÓN TECNOLÓGICA PARA PUBLICACIÓN DE CONTENIDOS
  - 1.1. Plataforma
    - 1.1.1. Almacenamiento de datos
    - 1.1.2. Tecnologías a emplear
  - 1.2. Análisis de texto para metadatos
    - 1.2.1. Software para análisis de texto
  - 1.3. Técnicas de optimización para motores de búsqueda
  
2. CAPÍTULO 2 – ANÁLISIS Y DISEÑO DEL SISTEMA
  - 2.1. Estructuración de plataforma en internet
  - 2.2. Definición de tecnologías a emplear para publicar sistema
  - 2.3. Implementación de sistema
  
3. CAPÍTULO 3 – PRUEBAS DE PLATAFORMA FUNCIONANDO
  - 3.1. Análisis estadístico de datos procesados por el sistema

3.2. Estudio de métricas evaluadas en el sistema

4. CAPÍTULO 4 – EVALUACIÓN DE BENEFICIOS POTENCIALES  
OBTENIDOS SEGÚN PRUEBAS FUNCIONALES

4.1. Conclusión sobre beneficios al publicar en el sistema

CONCLUSIONES

RECOMENDACIONES

REFERENCIAS

BIBLIOGRAFÍA

ANEXOS

## **10. METODOLOGÍA**

### **10.1. Variables e indicadores**

Las variables relacionadas con el estudio se resumen en la tabla I y son las siguientes:

#### **10.1.1. Problema: distribución ilegal gratuita de contenidos tradicionalmente impresos por internet**

Se define como la existencia de contenidos distribuidos ilegalmente en internet que no reportan ingresos a los dueños del *copyright*.

#### **10.1.2. Solución**

Sistema con una plataforma en internet con anuncios dirigidos, técnicas sociales y optimización de motores de búsqueda para ofrecer contenidos gratuitos pero con ganancias para sus dueños.

Un sistema que otorga ingresos a los dueños del *copyright* aun cuando los contenidos se ofrecen de manera gratuita a los lectores.

#### **10.1.3. Subvariables**

Las subvariables relacionadas con el problema que se detectaron, se enfocan en las ganancias que pueden producir los contenidos gratuitos y como pueden afectar a los dueños del *copyright*.

#### **10.1.3.1. Ganancias por contenidos gratuitos publicados en internet para publicadores**

Porcentaje de la ganancia total potencial obtenida por quien distribuye los contenidos; es de tipo cuantitativo y su indicador es el porcentaje de ganancia obtenido.

#### **10.1.3.2. Ganancias por contenidos gratuitos publicados en internet para poseedores del derecho de autor**

Porcentaje de la ganancia total potencial obtenida por quien produce los contenidos o posee los derechos de autor; es de tipo cuantitativo y su indicador es el porcentaje de ganancia obtenido.

#### **10.1.3.3. Pérdidas por contenidos gratuitos publicados en internet para dueños del copyright**

Porcentaje de la ganancia total potencial pérdida por quien produce los contenidos o posee los derechos de autor; es de tipo cuantitativo y su indicador es el porcentaje perdido de la ganancia total potencial.

#### **10.1.4. Matriz de operacionalización de variables**

Las variables y subvariables se detallan para contar con un panorama amplio de los focos donde el estudio se concentra. En principio las variables son el problema y la solución pero estos luego se descomponen en características más específicas.

Tabla I. **Variables**

Variables	Definición Conceptual
<b>Problema:</b> Distribución ilegal gratuita de contenidos tradicionalmente impresos por internet.	La existencia de contenidos distribuidos ilegalmente en internet que no reportan ingresos a los dueños del <i>copyright</i> .
<b>Solución:</b> Sistema con una plataforma en internet con anuncios dirigidos, técnicas sociales y optimización de motores de búsqueda para ofrecer contenidos gratuitos pero con ganancias para sus dueños.	Un sistema que otorga ingresos a los dueños del <i>copyright</i> aun cuando los contenidos se ofrecen de manera gratuita a los lectores.
Ganancias por contenidos gratuitos publicados en internet (publicador)	Porcentaje de la ganancia total potencial obtenida por quien distribuye los contenidos
Ganancias por contenidos gratuitos publicados en internet (dueño de <i>copyright</i> )	Porcentaje de la ganancia total potencial obtenida por quien produce los contenidos o posee los derechos de autor.
Pérdidas por contenidos gratuitos publicados en internet (dueño de <i>copyright</i> )	Porcentaje de la ganancia total potencial pérdida por quien produce los contenidos o posee los derechos de autor.

Fuente: elaboración propia.

## 10.2. Diseño, tipo y alcances del estudio

El estudio tendrá un tipo de investigación cuantitativa con alcances de tipo experimentales y explicativos; esto puede observarse en la tabla II. Para su realización se trazan varias fases para cumplir con los resultados esperados.

Tabla II. **Tipo y alcances del estudio**

Tipo	Alcances
Investigación Cuantitativa	Experimentales, Explicativos
Se diseñará un sistema en línea que produzca palabras clave para mostrar publicidad específica en base a contenidos tradicionalmente impresos. El estudio incluirá resultados relacionados a la implementación de tal plataforma.	Se realizarán pruebas documentadas al sistema para predecir comportamientos en una potencial fase de producción y así determinar el beneficio esperado para autores que lo empleen. Desde el punto de vista explicativo, se explorará la relación entre la publicación gratuita en internet y la obtención de ganancias por medio de publicidad dirigida.

Fuente: elaboración propia.

## 10.3. Fases del estudio

Las fases del estudio se desglosan para conseguir implementar el sistema de pruebas planteado, así como para realizar las mediciones prudentes para contar con respuestas certeras en cuanto al problema que se pretende resolver.

### **10.3.1. Observación**

Mediante la observación se determinó la necesidad de una solución para el problema de la publicación gratuita no controlada de contenidos tradicionalmente impresos. Esta observación inicia y justifica el estudio planteado.

### **10.3.2. Análisis del entorno**

El estudio contendrá información sobre el estado actual de la publicación en línea de documentos tradicionalmente impresos y cómo este cambio en el paradigma de los medios impresos, afecta directamente a los autores y poseedores del *copyright*.

Aunque se plantea una plataforma innovadora, se estudian alternativas preexistentes, diferentes, que se emplean para acceder a contenidos en internet.

### **10.3.3. Técnicas para recolección de información**

Se emplearán fuentes secundarias a través de la investigación bibliográfica y la consulta a diversos sitios en internet que provean datos al día acerca de las publicaciones a través de medios electrónicos.

Por otro lado, se realizarán pruebas sobre el sistema generado lo cual producirá información relevante para la conclusión de la investigación, fundamentalmente, de si el sistema realmente podría constituir una alternativa redituable para la publicación de contenidos.

#### **10.3.4. Revisión documental**

Los documentos consultados y presentados en la bibliografía de este documento son de fuentes verificadas y, en su mayoría, de carácter científico para fundamentar así los supuestos que dan origen al estudio.

Esta bibliografía ha sido escogida según sus contenidos para sustentar la mayor parte de la información requerida para todas las fases del estudio, incluyendo los aspectos técnicos de la creación de una plataforma para pruebas.

#### **10.3.5. Trabajo de campo**

Tras la creación de una plataforma para pruebas, se realizarán ensayos del uso de la misma como si se tratase del sistema completo funcionando durante tres meses; estas pruebas arrojarán datos sobre la viabilidad de un producto final, además de información sobre rendimiento, eficiencia y retornos esperados para los poseedores de los derechos de autor.

Por otro lado, se estudiarán alternativas preexistentes de publicación en línea que se puedan adaptar a la misma plataforma.

#### **10.3.6. Población y muestra**

Los resultados de la investigación podrán ser empleados por el mercado masivo de usuarios de contenidos tradicionalmente impresos en internet.

Para efectos de la investigación, se procesaran por lo menos 5 documentos en el sistema de prueba, para así estudiar la capacidad de la plataforma. Además, se presentará el producto a los estudiantes de la Maestría en Tecnologías de la Información y la Comunicación.



## 11. TÉCNICAS DE ANÁLISIS DE INFORMACIÓN

Se emplearán fuentes secundarias, tales como estudios previos sobre técnicas relativas a la publicación y a la generación de publicidad dirigida y con un alto grado de granularidad.

Además, se estudiarán por medio de estadística descriptiva los resultados generados por la plataforma desarrollada, de manera que revelen el tipo de publicidad que podrá ser generada; por ejemplo, conteos de palabras clave relevantes para un texto determinado.

La estadística inferencial se empleará desde dos perspectivas, en primer lugar para estimar la relevancia de palabras clave para la generación de publicidad en documentos y por otro lado para estimar el beneficio y uso en general de la plataforma a desarrollar.

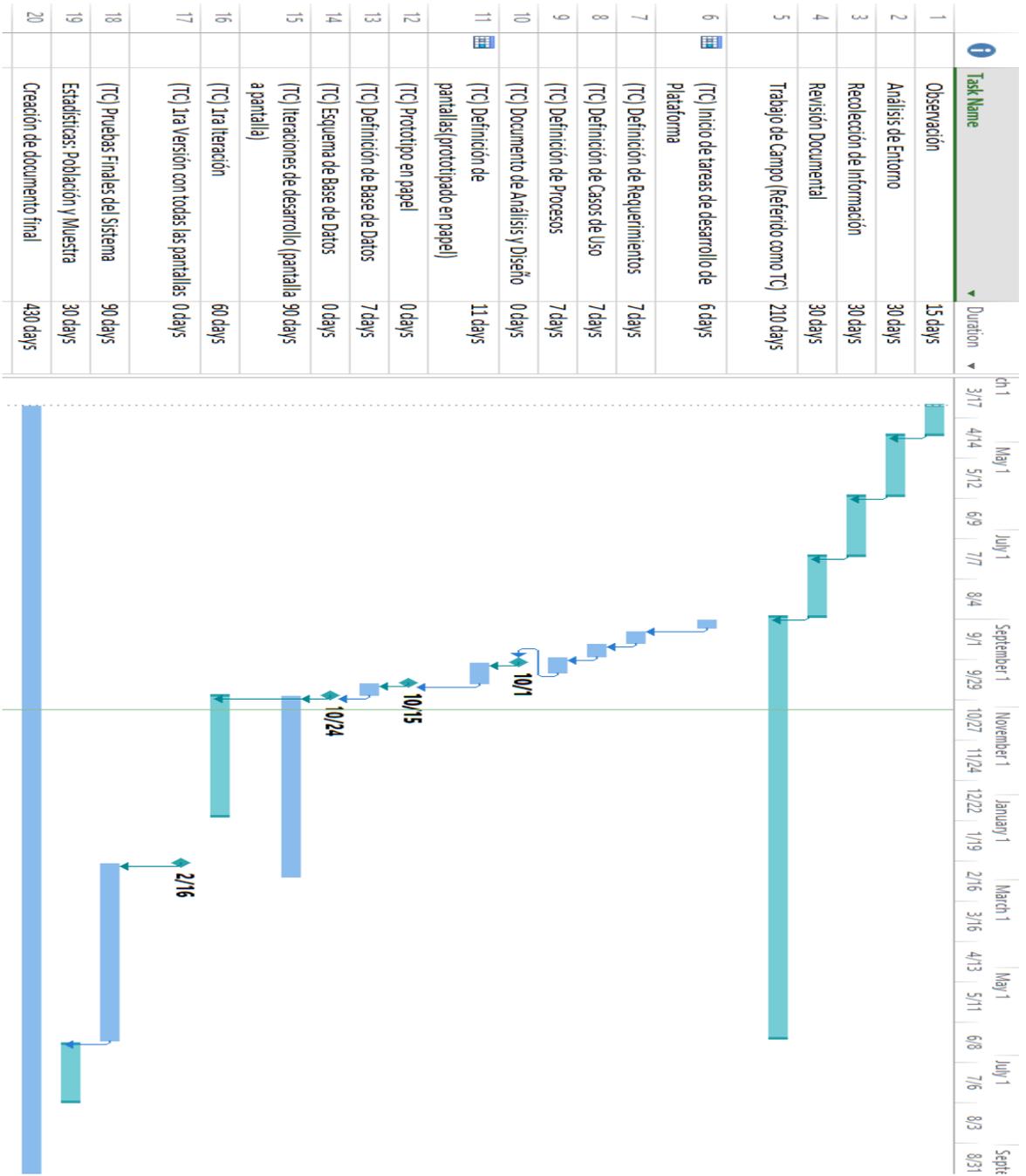
La relevancia de las palabras clave puede estimarse a través de modelos existentes para análisis de texto, mismos que se encuentran embebidos en software para análisis de datos preexistente. Se evaluará cual modelo es el óptimo para este caso o si debe definirse un modelo que combine otros preexistentes.

El modelo toma como entrada las palabras de un texto dado y obtiene como salida cuales de estas palabras son las más relevantes y que por lo tanto pueden emplearse para generar publicidad relacionada posteriormente.

El beneficio porcentual obtenido por publicar bajo el esquema previsto en este estudio deberá ser predecible mediante un modelo estadístico predictivo, tomando como variables de entrada la cantidad de publicaciones y la relevancia relativa a la publicidad disponible al ingresar en la plataforma diseñada.

## 12. CRONOGRAMA

Figura 1. Cronograma



Fuente: elaboración propia, con programa Microsoft Project 2013.

## 12.1. Tareas

En la tabla III se detallan las tareas que se programaron en el cronograma, estas tareas están enlazadas a las etapas del estudio que se definieron previamente y se plantean para finalizarse en el tiempo que dura la maestría.

Tabla III. **Tareas programadas**

Nombre de la Tarea	Duración	Inicio	Fin
Observación	15 días	Vie 5/4/13	Jue 25/4/13
Análisis de Entorno	30 días	Vie 26/4/13	Jue 6/6/13
Recolección de Información	30 días	Vie 7/6/13	Jue 18/7/13
Revisión Documental	30 días	Vie 19/7/13	Jue 29/8/13
Trabajo de Campo (Referido como TC)	210 días	Vie 30/8/13	Mie 18/6/14
(TC) Inicio de tareas de desarrollo de Plataforma	6 días	Dom 1/9/13	Vie 6/9/13
(TC) Definición de Requerimientos	7 días	Lun 9/9/13	Mar 17/9/13
(TC) Definición de Casos de Uso	7 días	Mie 18/9/13	Jue 26/9/13
(TC) Definición de Procesos	7 días	Vie 27/9/13	Lun 7/10/13
(TC) Documento de Análisis y Diseño	0 días	Mar 1/10/13	Mar 1/10/13
(TC) Definición de pantallas(prototipado en papel)	11 días	Mar 1/10/13	Mar 15/10/13
(TC) Prototipo en papel	0 días	Mar 15/10/13	Mar 15/10/13
(TC) Definición de Base de Datos	7 días	Mar 15/10/13	Mie 23/10/13
(TC) Esquema de Base de Datos	0 días	Jue 24/10/13	Jue 24/10/13

Continuación de la tabla III.

(TC) Iteraciones de desarrollo (pantalla a pantalla)	90 días	Jue 24/10/13	Mie 26/2/14
(TC) 1ra Iteración	60 días	Jue 24/10/13	Mie 15/1/14
(TC) 1ra Versión con todas las pantallas	0 días	Dom 16/2/14	Dom 16/2/14
(TC) Pruebas Finales del Sistema	90 días	Lun 17/2/14	Vie 20/6/14
Estadísticas: Población y Muestra	30 días	Lun 23/6/14	Vie 1/8/14
Creación de documento final	430 días	Vie 5/4/13	Mie 26/11/14

Fuente: elaboración propia.



## **13. FACTIBILIDAD DEL ESTUDIO**

### **13.1. Factibilidad operativa**

Recursos desde el punto de vista operativo:

### **13.2. Recursos humanos**

Investigador / desarrollador: Mauricio Haroldo Rojas Morales

Asesor: Luis Fernando Espino Barrios

### **13.3. Acceso a información**

Información de dominio público y de investigación bibliográfica, principalmente la planteada en este documento.

### **13.4. Permisos**

Todo el estudio y las pruebas se realizaran empleando documentos de dominio público o personales, el objetivo a cubrir es alcanzable mediante este tipo de ejemplificación.

### **13.5. Equipo**

Se empleará una computadora y un servidor gratuito en la nube para realizar las pruebas.

### **13.6. Infraestructura**

El servicio de publicación en la nube administra la infraestructura de red requerida para realizar las pruebas. Este servicio es gratuito y ya está contratado para las etapas cubiertas en este estudio.

### **13.7. Aspectos éticos, legales y de seguridad industrial**

Al trabajarse con documentos de dominio público no se tienen problemas legales en cuanto al uso de la información.

Desde el punto de vista técnico en la implementación de software para realizar pruebas funcionales, se empleará herramientas de código abierto, respetando las respectivas licencias según lo que estipulen en cada caso.

### **13.8. Conclusión de la factibilidad operativa**

El estudio es viable desde el punto de vista operativo debido a que las actividades que involucra pueden ser realizadas en su totalidad por el investigador y están planificadas para casar en el tiempo de investigación requerido.

Los recursos están totalmente disponibles e incluso ya se encuentran en uso por parte del investigador.

Las pruebas planteadas en el estudio garantizan que la solución será funcional.

### **13.9. Factibilidad técnica**

Desde el punto de vista técnico se presentan los argumentos que determinan si el proyecto es viable, estos argumentos van desde los recursos con los que se cuenta, la extensión del trabajo técnico a realizar y el tiempo que se invertirá en el mismo.

#### **13.9.1. Recursos desde el punto de vista técnico**

Desde este punto de vista, los recursos son todos aquellos que se emplearán para desarrollar la solución sobre la cual se realizarán las pruebas. Esta solución será una aplicación por lo que los elementos necesarios estarán relacionados con tecnología.

#### **13.9.2. Recursos tecnológicos**

El equipo, infraestructura y software necesarios para realizar el estudio se encuentran disponibles y en uso. El investigador conoce las herramientas necesarias para desarrollar el proyecto propuesto; aquellas tecnologías de reciente implementación que se utilizarán, encuentran sustento cognitivo en base a la investigación bibliográfica desarrollada.

### **13.9.3. Extensión del trabajo de investigación**

El trabajo de investigación está planificado para culminar en dos años, considerando la preparación preliminar y actividades de desarrollo de una plataforma de pruebas.

La calendarización planteada en este documento establece un período de cien días para el desarrollo de la plataforma, un tiempo apropiado para que un proyecto de esta magnitud pueda ser puesto a prueba según algunas metodologías.

### **13.9.4. Tiempo disponible para realizar lo que se propone**

El tiempo estimado para desarrollar el proyecto de acuerdo a lo calendarizado será de catorce horas semanales; distribuidas en horarios fuera de las horas laborales.

### **13.9.5. Conclusión de la factibilidad técnica**

El estudio es factible desde el punto de vista técnico, debido a que se cuenta con los recursos tecnológicos necesarios, así como el conocimiento para su aplicación en la tarea investigativa. La experiencia de más de siete años laborando en el área de la tecnología, provee al investigador con la habilidad necesaria para completar lo requerido en este sentido.

Aunado a esto, el tiempo ha sido programado de tal modo que el estudio pueda concluirse en tiempo considerando el grado de complejidad establecido y el tiempo disponible por parte de los involucrados en el proyecto.

## **13.10. Factibilidad económica**

Desde el punto de vista económico, la viabilidad del proyecto se evalúa considerando los costos en los que se incurriría para completar el estudio, tomando en cuenta el costo del tiempo y de los insumos a emplear.

### **13.10.1. Recursos económicos**

Se emplearan recursos propios del investigador, para realizar el estudio no es necesario recurrir a fuentes externas.

### **13.10.2. El costo del tiempo**

Costo por hora del investigador / desarrollador: Q 62,50, basado en promedio de ingresos aproximado

Costo total: Q 87 500,00 para el plazo total de 2 años en el proyecto considerando el tiempo de 14 horas semanales.

Costo por hora del asesor: Q 70,50

Costo total: Q 8 460,00 para el plazo total de 2 años en el proyecto considerando el tiempo de 2 horas semanales.

### **13.10.3. El costo de la realización**

Los costos por espacio y servicios son de aproximadamente Q 2 800,00 desglosados de la siguiente manera:

Espacio físico: Q. 1 700,00  
 Internet: Q 300,00  
 Electricidad y agua: Q. 600,00  
 Otros consumibles: Q. 200,00  
 Costo total para 2 años: Q. 67 200,00

#### **13.10.4. El costo de adquirir nuevos recursos**

El costo en este sentido es insignificante dada la infraestructura gratuita que se empleará en el proyecto y a que la mayor parte de la parte investigativa será desarrollada por el investigador.

#### **13.10.5. Costo total**

En la tabla IV se resumen los costos de cada insumo así como la sumatoria de todo lo necesario. Los insumos considerados en resumen son el tiempo del investigador, el tiempo del asesor, el espacio a utilizar y los servicios.

Tabla IV. **Costos**

Insumo	Costo
Investigador	Q. 87 500,00
Asesor	Q. 8 460,00
Espacio y servicios	Q. 67 200,00
Total	Q. 163 160,00

Fuente: elaboración propia.

### **13.10.6. Conclusión de la factibilidad económica**

El estudio cuenta con tres ventajas fundamentales desde el punto de vista económico:

- El principal recurso humano es el investigador por lo que el costo del personal, de ubicación e insumos, en su mayoría, es reabsorbido.
- Se emplean herramientas de código abierto que no tienen algún costo.
- Se utiliza infraestructura gratuita en internet para pruebas y publicación.

De este modo, este proyecto es viable desde el punto de vista económico con los recursos disponibles por parte del investigador.



## 14. BIBLIOGRAFÍA

- 1 California Digital Library Services Advisory group. (Junio de 2007). *CDL Guidelines for Digital Objects Version 2.0*. Obtenido de <http://www.cdlib.org/inside/diglib/guidelines>
- 2 Chen, P., & Lin, S. (2011). Word AdHoc Network: Using Google Core Distance to extract the most relevant information. *Knowledge-Based Systems*, 24(3), 393-405.
- 3 Colwell, T. (2010). Paid Online Marketing: Adwords Simplified. *Greater Games Industry Catalog*, (12), 18.
- 4 Cordon-García, J., Gómez-Díaz, R., & Alonso-Arévalo, J. (2011). LIBROS ELECTRÓNICOS: OFERTA COMERCIAL Y REDES P2P. *El Profesional De La Información*, 20(2), 149-158.
- 5 Costa-Knufinke, J. (2010). Adaptación de las editoriales españolas al libro electrónico. *El Profesional De La Información*, 19(1), 13-20.
- 6 Gho, K., & Jang-han, L. (2011). The Effect of Search Condition and Advertising Type on Visual Attention to Internet Advertising. *Cyberpsychology, Behavior & Social Networking*, 323-325.

- 7 Google.com. (Febrero de 2013). *Estadísticas de infracciones a copyright de bookos.org*. Obtenido de <http://www.google.com/transparencyreport/removals/copyright/domains/bookos.org>
- 8 Hervet, G., Guérard, K., Tremblay, S., & Chourou, M. (2011). Is banner blindness genuine? Eye tracking internet text advertising. *Applied Cognitive Psychology*, 25(5), 708-716.
- 9 Hervet, G., Guérard, K., Tremblay, S., & Chourou, M. (2011). Is banner blindness genuine? Eye tracking internet text advertising. *Applied Cognitive Psychology*, 25(5), 708-716.
- 10 Kuppusamy, K. S., & Aghila, G. (2012). A Model for Personalized Keyword Extraction from Web Pages using Segmentation. *International Journal of Computer Applications*, 42(1), 21-40.
- 11 Kuppusamy, K. S., & Aghila, G. (2012). A Model for Personalized Keyword Extraction from Web Pages using Segmentation. *International Journal of Computer Applications*, 42(1), 21-40.
- 12 Lim, C. P., Jain, L. C., Nishihara, Y., Kimura, A., & Ohsawa, Y. (2012). Extracting users'interest from book-browsing behaviors recorded with RFID. *International Journal Of Knowledge Based Intelligent engineering Systems*, 16(1), 17-23.
- 13 Martínez-San-Emeterio, M. (2010). Publicidad online en medios sectoriales: caso Grupo TecniPublicaciones. *El Profesional De La Información*, 19(2), 194-200.

- 14 Patel, S., Patel, J., & Patel, A. (2012). Statistical Analysis of SEO for Joomla, Drupal and Wordpress. *International Journal Of Computer Applications*, 521-5.
  
- 15 Sandieson, R. W., Kirkpatrick, L. C., & Zimmerman, W. (2010). Harnessing the Power of Education Research Databases with the Pearl-Harvesting Methodological Framework for Information Retrieval. *Journal Of Spaecial Education*, 44(3), 161-175.
  
- 16 Smith, S. (2005). In SEARCH of the Blog Economy. *Econtent*, 28(1/2), 24-29.
  
- 17 Statscrop.com. (Febrero de 2013). *Estadísticas de acceso a sitio bookos.org*. Obtenido de <http://www.statscrop.com/www/bookos.org>
  
- 18 Summers, N. (2012). Click this Ad Already! *Newsweek*, 159(13/14), 58-64.
  
- 19 Thompson, D. (2013). The Incredible Shrinking Ad. *Atlantic Monthly* (10727825), 24-28.
  
- 20 Veiksa, I. (2012). Is the existing concept of copyright still justified in the information society? *Current Issues Of Business & Law*, 7(2), 248-262.
  
- 21 Wang, L. (Junio de 2013). *Comparisson of pdf files online approaches*. Obtenido de Github Wang Lu: <https://github.com/coolwanglu/pdf2htmlEX/wiki/Comparison>

- 22 Wu, A., & Mitchell, A. M. (2010). MassManagement of E-Book Catalog Records: Approaches, Challanges, and Solutions. *Library Resources & Technical Services*, 54(3), 164-174.