

IDENTIFICACIÓN AUTOMÁTICA DE TENDENCIAS COMERCIALES A TRAVÉS DEL  
ANÁLISIS DE SITIOS DE E-COMMERCE

JORGE LUIS GALVIS QUINTERO

UNIVERSIDAD AUTÓNOMA DE BUCARAMANGA

UNIVERSITAT OBERTA DE CATALUNYA

MAESTRÍA EN SOFTWARE LIBRE

BOGOTÁ, D.C.

2015

IDENTIFICACIÓN AUTOMÁTICA DE TENDENCIAS COMERCIALES A TRAVÉS DEL  
ANÁLISIS DE SITIOS DE E-COMMERCE

JORGE LUIS GALVIS QUINTERO

Trabajo de grado presentado como requisito para optar al título de:

Magister en Software Libre

Directora:

Ph. D. LILIANA CALDERÓN BENAVIDES

UNIVERSIDAD AUTÓNOMA DE BUCARAMANGA

UNIVERSITAT OBERTA DE CATALUNYA

MAESTRÍA EN SOFTWARE LIBRE

BOGOTÁ, D.C.

2015

Nota de Aceptación

---

---

---

---

---

Jurado

---

Jurado

Bogotá, Enero de 2015.



## Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

This is a human-readable summary of (and not a substitute for) the [license](#).

[Disclaimer](#)

### You are free to:

**Share** — copy and redistribute the material in any medium or format

The licensor cannot revoke these freedoms as long as you follow the license terms.

### Under the following terms:



**Attribution** — You must give **appropriate credit**, provide a link to the license, and **indicate if changes were made**. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.



**NonCommercial** — You may not use the material for **commercial purposes**.



**NoDerivatives** — If you **remix, transform, or build upon** the material, you may not distribute the modified material.

**No additional restrictions** — You may not apply legal terms or **technological measures** that legally restrict others from doing anything the license permits.

### Notices:

You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable **exception or limitation**.

No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as **publicity, privacy, or moral rights** may limit how you use the material.

## **Dedicatoria**

*A Stella, mi madre.*

## **Agradecimientos**

Los autores expresan sus agradecimientos a:

Ingeniera Liliana Calderón Benavides, directora de tesis, quién orientó de forma proactiva el desarrollo de esta investigación.

Ingeniero Daniel Arenas, coordinador de maestría, quién con su apoyo en trámites académicos permitió la culminación del plan de estudios.

## Resumen

En esta investigación se construyó una aplicación Web que permite hacer seguimiento de los productos ofertados en cuatro plataformas de comercio electrónico con presencia en Colombia (Mercado Libre Colombia, Amazon, eBay y Linio Colombia) por un periodo específico de tiempo, con el objetivo de predecir sus comportamientos en el futuro reciente y comparar estos contra los comportamientos reales. El problema que intenta resolver esta investigación es: ¿cómo determinar de forma automática las tendencias comerciales que surgen en sitios de comercio electrónico con presencia en Colombia?

El seguimiento a los productos se hizo a través de tareas programadas (crons) que extraen datos específicos (precio, cantidad vendida, cantidad disponible) de dichas plataformas a través del consumo de sus APIs o con la técnica de Web Scraping, una vez recolectados los datos y, con el método estadístico de los mínimos cuadrados, se proyectó su comportamiento por el mismo número de días que duró su recolección, cuando una tarea programada termina su objetivo de extracción, se visualizan los resultados con respecto a las variables extraídas en la página de estadística individual de cada producto.

Los entregables obtenidos como resultado de la ejecución de las fases de la investigación fueron:

La especificación DOM/API de cada sitio seleccionado: contiene una descripción general del funcionamiento de cada sitio seleccionado a nivel de API o a nivel de DOM, esto de acuerdo a la estrategia de extracción seleccionada para cada sitio.

Modelo del repositorio de datos: describe el modelo utilizado para almacenar los datos a extraer y los datos a graficar.

Scripts de extracción y proyección de comportamientos: son un conjunto de programas es-

critos en Python que permiten la extracción de datos desde los sitios seleccionados, además de calcular su comportamiento en el futuro reciente.

Interfaz Web: es una aplicación Web escrita en el framework de desarrollo Django, que permite visualizar los resultados que calcula el script de proyección de comportamiento de productos, además permite administrar las tareas programadas, los sitios y las categorías.

Palabras claves: Tendencias comerciales, comercio electrónico, búsqueda de tendencias comerciales, tendencias comerciales en Internet, análisis de tendencias de e-commerce, identificación automática de tendencias comerciales, e-commerce tendencias, repositorios de tendencias comerciales, software de búsqueda de tendencias comerciales, software de tendencias comerciales, algoritmos de tendencias comerciales, método de los mínimos cuadrados.



## Contenido

	<b>pág.</b>
1. Introducción	17
1.1 Antecedentes	17
1.2 Definición del problema	19
1.3 Justificación	19
1.4 Objetivos	20
1.4.1 Objetivo general	20
1.4.2 Objetivos específicos	20
1.5 Estructura del documento	20
2. Revisión bibliográfica	23
2.1 Estado del Arte	23
2.1.1 eCommerce, comportamientos y tendencias	24
2.1.2 eCommerce y clasificación	31
2.1.3 Consulta y extracción Web	32
2.2 Marco Teórico	35
2.2.1 e-Commerce	37
2.2.2 Tendencias comerciales	38
2.2.3 Information Retrieval (IR, recuperación de información)	38
2.2.4 Application Programming Interface (API)	38
2.2.5 Crawler	39
2.2.6 Buscadores Web	41
2.2.7 Ingeniería de Software	42

2.2.8 Metodologías ágiles	46
2.2.9 Métodos de tendencia lineal	46
2.2.9.1 Método gráfico (diagrama de dispersión)	46
2.2.9.2 Regresión lineal simple	47
2.2.9.3 Regresión lineal por mínimos cuadrados	47
3. Marco Legal	49
4. Método de la Investigación	53
4.1 Enfoque metodológico	53
4.2 Definición de población y muestra	53
4.3 Descripción general de los sitios	54
4.3.1 Mercado libre Colombia.	54
4.3.2 Tiendas Monomi	55
4.3.3 OLX Colombia	55
4.3.4 Linio Colombia	55
4.3.5 eBay	55
4.3.6 Amazon	55
4.3.7 Tmart	55
4.3.8 BangGood	56
4.3.9 Dafiti	56
4.3.10 Éxito.	56
4.3.11 Falabella Colombia	56
4.3.12 Alibaba	56
4.3.13 Alkosto	56

4.3.14 HomeCenter	57
4.4 Selección de sitios Web	57
4.4.1 Criterios de selección para los Sitios Web de comercio electrónico a modelar	57
4.4.2 Método de selección de sitios: matriz de selección	58
4.4.3 Proceso de selección de sitios de comercio electrónico	59
4.5 Técnicas de Recolección de Datos	68
4.6 Método de Proyección de Comportamiento: Mínimos Cuadrados	69
4.6.1 Criterios de selección de métodos estadísticos.	69
4.6.2 Selección del método estadístico.	70
4.7 Fases desarrolladas para la construcción del producto de software	70
4.8 Algunos Aspectos de Diseño	72
4.9 Plan de Presentación de Resultados	73
5. Resultados de la Investigación	74
5.1 Análisis DOM y API de los Sitios Seleccionados	74
5.1.1 Mercado libre	74
5.1.2 eBay	78
5.1.3 Amazon	81
5.1.4 Linio Colombia	86
5.2 Modelo del Repositorio de Datos	89
5.3 Aspectos de Diseño de la Aplicación	91
5.3.1 Aspectos visuales de la interfaz Web de consulta	91
5.3.2 Aspectos visuales de la interfaz Web de administración	93
5.3.3 Modelo de clases de la aplicación	95

5.3.4 Consideraciones de despliegue y código fuente	99
5.3.5 Ejemplos de categorías y productos	99
5.3.6 Descripción y uso de la herramienta	102
5.3.6.1 Listar sitios	102
5.3.6.2 Listar categorías	103
5.3.6.3 Listar tareas programadas	104
5.3.6.4 Listar productos	105
5.4 Descripción de resultados para un producto específico	107
6. Conclusiones	113
7. Recomendaciones y Trabajos Futuros	116
Bibliografía	118
Anexos	125

## Lista de Figuras

	pág.
Figura 1. Estructura del estado del arte	23
Figura 2. Estructura del marco teórico	36
Figura 3. Recta de regresión de Y sobre X	47
Figura 4. Recta de regresión de X sobre Y	47
Figura 5. Coeficiente de correlación	47
Figura 6. Cálculo de parámetros por mínimos cuadrados	48
Figura 7. Coeficiente de determinación	48
Figura 8. Coeficiente de correlación	48
Figura 9. Interfaz principal del Applications Manager de Mercado Libre	75
Figura 10. Detalle de la aplicación registrada	75
Figura 11. Dashboard principal cuenta de desarrollador en eBay	79
Figura 12. Proceso de registro en el Product Adversiting API Developer program	82
Figura 13. Registro como Associate User	83
Figura 14. Categoría tecnología/teléfonos en Linio Colombia	86
Figura 15. Producto específico en Linio Colombia	88
Figura 16. Modelo relacional del repositorio para las entidades sitios, categorías, tareas y productos	
Figura 17. Esquema JSON para el almacenamiento de lecturas	90
Figura 18. Interfaz de categorías de la aplicación de consulta	92
Figura 19. Interfaz de estadística de productos	93
Figura 20. Interfaz de administración de las categorías	94

Figura 21. Formulario de registro y edición de categorías	95
Figura 22. Herencia entre los modelos de la aplicación	96
Figura 23. Relaciones entre los modelos de la aplicación	96
Figura 24. Clases de las vistas públicas de la aplicación de consulta	97
Figura 25. Clases de los formularios de la aplicación de consulta	97
Figura 26. Clases de las vistas privadas de la aplicación de consulta	98
Figura 27. Módulo de sitios	103
Figura 28. Módulo de categorías	104
Figura 29. Módulo de tareas programadas	105
Figura 30. Módulo de productos	106
Figura 31. Página de estadísticas	106
Figura 32. Proyección de precios para el producto 321647917600 en eBay	108
Figura 33. Proyección de ventas para el producto 321647917600 en eBay	109
Figura 34. Proyección de inventario para el producto 321647917600 en eBay	110
Figura 35. Tendencia de precio para el producto 321647917600 en eBay	111
Figura 36. Tendencia de venta para el producto 321647917600 en eBay	111
Figura 37. Tendencia de inventario para el producto 321647917600 en eBay	112

## Lista de Tablas

pág.

Tabla 1. Compañías de proyectos de plataformas de comercio electrónico.	
Tabla 2. Aplicaciones	
Tabla 3. Lenguajes y tecnologías de base de datos	
Tabla 4. Muestra de tiendas electrónicas	
Tabla 5. Criterios de selección de tiendas electrónicas	
Tabla 6. Matriz de comparación de criterios	
Tabla 7. Evaluación de criterios por sitio	
Tabla 8. Ponderación para el factor términos de uso	
Tabla 9. Ponderación para el factor disponibilidad de una API	
Tabla 10. Ponderación para el factor presencia en Colombia	
Tabla 11. Ponderación para el factor desarrollado sobre un CMS/Framework	
Tabla 12. Ponderación genera	
Tabla 13. Técnica de extracción por sitio seleccionado	
Tabla 14. Herramientas de desarrollo utilizadas en la implementación de la aplicación	
Tabla 15. Esquema de consulta a la API de Mercado Libre	
Tabla 16. Listado parcial de APIs de eBa	
Tabla 17. Detalle URL de servicio para la API de búsqueda de eBa	
Tabla 18. Detalle URL de servicio para la API de compra de eBa	
Tabla 19. Detalle URL de servicio para la API de consulta de Amazo	
Tabla 20. Principales categorías de Mercado Libre	
Tabla 21. Listado de algunos productos de Mercado Libre	

## Lista de Anexos

pág.

- Anexo A. Términos de uso de los sitios tomados como muestra
- Anexo B. Lista de APIs utilizadas por plataforma
- Anexo C. Esquema de licenciamiento de la documentación y de la aplicación Web
- Anexo D. Documento JSON devuelto por la API de Mercado Libre
- Anexo E. Documento JSON devuelto por la API de eBay
- Anexo F. Documento XML devuelto por la API de Amazon
- Anexo G. Estructura general del DOM de categorías en Linio Colombia
- Anexo H. SQL generado a partir del modelo relacional
- Anexo I. Instrucción JavaScript de ejemplo para crear un documento en la base de datos no relacional
- Anexo J. Manual de usuario de la interfaz de consulta
- Anexo K. Manual de usuario de la interfaz de administración



## 1. Introducción

### 1.1 Antecedentes

Esta investigación surge del interés por conocer las tendencias comerciales en sitios Web de comercio electrónico con presencia en Colombia, esta misma hace parte del proyecto global titulado “Planteamiento y desarrollo de una plataforma integral de gestión de la información” del grupo de investigación Tecnologías de la Información de la Universidad Autónoma de Bucaramanga. El proyecto tiene un ámbito nacional y se desarrolló en la Ciudad de Bogotá, con este se pretende, en gran medida, dar respuesta al interrogante: ¿cómo identificar de forma automática las tendencias comerciales que se presentan en los sitios de comercio electrónico en Colombia? La solución propuesta consistió en el desarrollo de una aplicación Web que haga uso de la técnica de extracción de datos “Web Scrapping” y haga uso de las interfaces de programación (APIs) disponibles en los sitios de comercio electrónico estudiados, para así, hacer seguimiento al comportamiento de sus productos.

La aplicación se desarrolló en cuatro grandes etapas:

La primera de ellas consistió en seleccionar cuatro plataformas de comercio electrónico con presencia en Colombia para su posterior estudio. Para la selección de los sitios a analizar se tomaron como muestra catorce tiendas electrónicas publicadas como populares por la Cámara Colombiana de Comercio Electrónico en su reporte de abril de 2014 (solo se tuvieron en cuenta plataformas de propósito general), a estas catorce tiendas se les aplicó el método de matriz de selección con un conjunto de criterios que tienen en cuenta el proceso de desarrollo de los scripts de extracción de datos (términos de uso del sitio Web, disponibilidad de una interfaz de programación, presencia en Colombia, etc.), finalmente la matriz de selección permitió escoger a las plataformas: eBay, Amazon, Mercado Libre Colombia y Linio Colombia. Por cada uno de estos sitios

seleccionados se procedió a realizar el análisis de su funcionamiento a nivel de publicación de ofertas que permitió determinar la mejor estrategia para la extracción de datos.

La segunda etapa consistió en modelar e implementar un repositorio que almacenara los datos que los scripts (crons) iban extrayendo desde las mismas tiendas seleccionadas, en el repositorio se tendrían las lecturas hechas por los scripts para cada uno de los productos involucrados en las tareas automáticas de extracción de datos. El modelo implementado contiene un componente relacional que almacena los datos asociados a los sitios, las categorías, las tareas de extracción y los productos, además cuenta con un componente no relacional que permitió almacenar las lecturas de los productos con una estructura sencilla y de fácil manipulación.

La tercera etapa cubrió la búsqueda y selección de un método estadístico que permitiera proyectar el comportamiento de los productos con base en lecturas previas respecto a variables como precio, cantidad disponible y cantidad vendida. El método seleccionado fue el de los mínimos cuadrados. En esta fase también se construyeron los scripts de extracción de datos y los scripts de proyección de comportamiento. Ambos tipos de scripts fueron implementados en el lenguaje Python.

La cuarta y última fase se concentró en el desarrollo de la interfaz gráfica para la visualización de resultados y para la administración de entidades (sitios, categorías, tareas y productos). Se construyó una aplicación Web en Django (framework de desarrollo Web para el lenguaje de programación Python) con dos componentes: el primero permite listar/filtrar sitios, categorías, tareas y productos, además de desplegar la página de estadísticas para cada producto involucrado en una tarea de extracción de datos, el segundo componente permite la administración (agregar, modificar, eliminar, consultar) de sitios, categorías y tareas de extracción. Es esta la interfaz Web, en la que los usuarios podrán ver los resultados de los scripts de extracción y proyección de

comportamiento, ya que estos scripts no son visibles ni administrables.

## **1.2 Definición del problema**

Si el lector es un comprador frecuente en una determinada tienda virtual, ¿Ha podido observar tendencias comerciales en dicha tienda por categorías? ¿ha podido observar tendencias comerciales con respecto a los productos que usualmente compra? Justamente esta investigación intenta dar una respuesta aproximada a este tipo de interrogantes. El proyecto abarca el desarrollo de una aplicación Web que permita de manera automática establecer tendencias comerciales (comportamiento de productos) en algunos sitios de comercio electrónico previamente seleccionados y modelados, esto debido a que no hay evidencia encontrada por el autor que sugiera la existencia (en Colombia) de una herramienta de software con características similares.

Es posible que se establezcan tendencias sobre sitios de comercio electrónico por parte de compañías especializadas en el área, pero seguramente son tendencias calculadas con intervención humana. El problema que se intenta resolver en esta investigación es: ¿cómo determinar de forma automática las tendencias comerciales que surgen en sitios de comercio electrónico con presencia en Colombia?

## **1.3 Justificación**

La aplicación desarrollada permitirá identificar posibles tendencias comerciales en los sitios de comercio electrónico seleccionados, además permitirá visualizar el comportamiento futuro de los productos con base en lecturas hechas previamente sobre las plataformas. A nivel de documentación y de código fuente, desarrolladores y/o investigadores podrán construir obras derivadas utilizando los modelos diseñados para la implementación del producto.

Una versión o derivación del producto de software obtenido les permitirá a las plataformas de e-Commerce un soporte de facto para establecer promociones o estrategias comerciales que estén

más ajustadas a su mercado.

Esta investigación buscó desarrollar una primera versión de una herramienta pionera en su categoría, es probable que no ofrezca funcionalidades determinantes para ser utilizadas por compañías de mercadeo o análisis de ventas, pero si será un referente para desarrollos posteriores.

## **1.4 Objetivos**

**1.4.1 Objetivo general.** Determinar de forma automática las tendencias comerciales que se registran en sitios de Comercio electrónico en Colombia.

### **1.4.2 Objetivos específicos:**

Establecer el esquema básico de funcionamiento de cuatro (4) sitios Web de comercio electrónico con presencia en Colombia.

Implementar un repositorio dinámico que almacene patrones comerciales con base en el esquema de funcionamiento de los sitios Web seleccionados.

Implementar un algoritmo de inspección y registro estadístico que se apoye en los datos registrados en el repositorio que permita determinar las tendencias comerciales en los sitios de comercio electrónico seleccionados.

Implementar una interfaz Web que permita visualizar las tendencias comerciales.

## **1.5 Estructura del documento**

En este documento se encuentra la especificación teórica y metodológica de la investigación, organizado en 7 capítulos como sigue:

En el capítulo uno se encuentra la introducción, donde se contextualiza un poco la investigación y se presenta de forma breve el problema y los objetivos que se buscan completar al final del estudio.

En el capítulo dos se encuentra la revisión bibliográfica, que se compone del estado del arte

y del marco teórico. En el estado del arte se documentan investigaciones afines al proyecto; algoritmos de recuperación de información en la Web, categorización de productos, especificidad de consultas de usuario, etc. Además se referencian técnicas de extracción de datos Web y estudios clásicos en el área del comercio electrónico. En el marco teórico se detallan algunos conceptos relevantes del “Information Retrieval” y del e-Commerce, como su propia definición, sus objetivos, además de temas directamente asociados al desarrollo del proyecto como APIs y Web Scrapping.

En el capítulo tres se tiene el marco legal, que brinda un resumen de la normatividad que afecta al desarrollo del proyecto, menciona directrices desde derechos de autor hasta licencias de software.

En el capítulo cuatro se detalla el proceso de desarrollo de la propuesta, se mencionan las técnicas y tecnologías utilizadas para completar cada uno de los objetivos específicos, es en este capítulo dónde se especifica la selección de sitios Webs junto a su estructura de funcionamiento, el diseño e implementación de los script de extracción de datos y proyección de comportamiento, además de la interfaz gráfica para la visualización de resultados.

Los resultados aparecen en el capítulo cinco, dónde se pueden apreciar las características del producto de software final construido, junto a algunas descripciones de las plataformas estudiadas.

En el capítulo seis se encuentran las conclusiones, las cuales permiten observar una pequeña retrospectiva acerca de los objetivos planteados con respecto a los resultados obtenidos.

Finalmente, en el capítulo siete se mencionan algunas recomendaciones generales y algunas propuestas para trabajos derivados y/o futuros.

## 2. Revisión bibliográfica

### 2.1 Estado del Arte

Son numerosos los estudios realizados en áreas asociadas al e-Commerce, estos van desde aspectos históricos, pasando por el análisis y mejoramiento de consultas, hasta llegar a técnicas de categorización. Transversal a estos frentes de acción, se encuentran los tópicos de extracción (eCommerce queries) y clasificación de ítems comerciales, que justamente son temas fundamentales en el desarrollo de este trabajo de investigación. A continuación se muestra un mapa conceptual que resume el contenido de este capítulo.

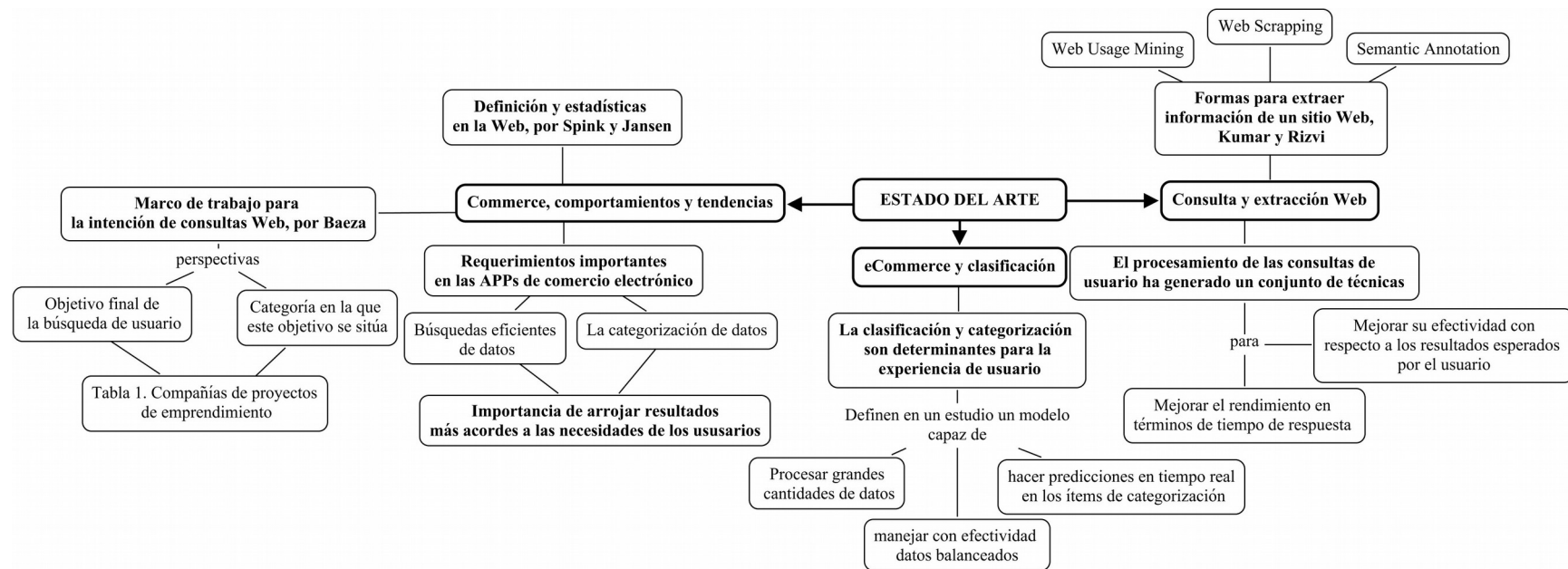


Figura 1. Estructura del estado del arte

**2.1.1 eCommerce, comportamientos y tendencias.** Spink (2002) describe la Web como una fuente de información y como una herramienta de uso dominante en los negocios, menciona que los tópicos de búsqueda se han movido de temas relacionados al entretenimiento y sexo a temas de comercio y personas, por ejemplo, en las estadísticas que presenta se puede identificar una disminución de búsquedas en temas relacionados a la pornografía y un aumento en temas asociados a viajes, gustos, ofertas de empleo, economía, compra y venta de artículos.

El estudio menciona que los usuarios tienden a revisar pocas páginas de resultados por consulta hecha.

Estas conclusiones se basan en el análisis de consultas de usuarios (1 millón de consultas) que desarrolló sobre el portal Web Excite ([www.excite.com](http://www.excite.com)) entre los años 1997 a 2001.

En un estudio sobre compartimientos de usuarios en Web Queries (utilizando nueve motores de búsqueda) desarrollado por Spink y Jansen, concluye entre otras afirmaciones que:

Los usuarios cada vez revisan menos resultados por consulta hecha.

Las búsquedas en Estados Unidos son más especializadas que las hechas en Europa, en otras palabras, hacen más uso de operadores avanzados de consulta.

En este mismo estudio se detallan tendencias importantes en cuanto a términos de búsqueda, se identifica una disminución de búsquedas relacionadas con sexo y pornografía y un aumento en términos como comercio, viajes, economía y empleo.

Como dato particular, la investigación demuestra que las consultas asociadas a computadores o internet han disminuido notablemente (del 12.5 % al 9.7% en AltaVista).

En Colombia, una encuesta realizada en septiembre de 2013 (<http://tcanalysis.com/blog/posts/infografia-la-compra-online-en-colombia>) por la cámara colombiana de comercio electrónico (CCCE) y la compañía The cocktail analysis (<http://tcanalysis.com/es>) señala que:

El 76% de los internautas realizan compras presenciales con previo asesoramiento vía Internet (efecto ROPO: Research Online Purchase Offline), Estas consultas se dan mayormente en categorías como electrónica/informática, moda, alimentación, servicios de telecomunicaciones y cultura.

Las compras en línea en el país han incrementando; un 52% de usuarios consultados han hecho al menos una compra electrónica para productos como tiquetes aéreos, boletos de cine, teléfonos inteligentes, boletos de conciertos, entre otros. 2 de cada 10 usuarios que no realizan compras en línea lo harán dentro del corto plazo.

El mayor obstáculo para compras en línea por parte de los usuarios son las formas de pago, aún existe un alto grado de desconfianza con respecto a los pagos electrónicos.

Por otra parte, el informe de la CCCE de abril de 2014 revela que:

Los sitios de comercio electrónico mas visitados en Colombia son MercadoLibre, Amazon y Ebay. Para este tipo de portales el 51.9% de la audiencia masculina.

Los sitios de banking online más visitados son los del grupo Bancolombia y el Banco de Bogotá, con un 52.2% de audiencia femenina.

Despegar.com.co, avianca.com y lan.com son los sitios más visitados a la hora de comprar tiquetes aéreos, con un 51.2% de audiencia femenina.

Con respecto a la intención de compra, Baeza (2006) desarrolló un marco de trabajo para la identificación automática de los intereses de usuario (intenciones) en las consultas Web, este proceso de identificación se desarrolló desde dos perspectivas: el objetivo final de la búsqueda de usuario y la categoría en la que este objetivo se sitúa. Dentro del análisis de logs de consultas de usuarios que se realizaron, se identificaron las siguientes categorías para los términos de búsqueda



da: artes, juegos, compras, salud, sexo, noticias, ciencia, computadores, deportes, niños y jóvenes, hogar y recreación.

Existen millares de aplicaciones (software) de comercio electrónico, tanto en interfaces Web cómo móviles, algunas de ellas de alcance mundial, continental y otras sectorizadas a zonas geográficas o países. Detrás de dichas aplicaciones hay, desde grandes compañías, hasta proyectos de emprendimiento, por mencionar algunas se tienen:

**Tabla 1. Compañías de proyectos de plataformas de comercio electrónico.**

Nombre	País/ciudad de origen	Propietario/ Empresa	Presencia en Colombia	Presencia Global	Enfoque
Mercado Libre.	Latino América.	Mercado Libre.	Sí.	Presencia en Latino América y Portugal.	Cliente a cliente, artículos en general.
OLX.	New York / Buenos Aires.	OLX, Inc.	Sí.	106 países. En 40 idiomas.	Cliente a cliente, clasificados online para venta y compra de artículos usados y nuevos.
Monomi.	Colombia.	Monoku.	Sí.	Colombia y México.	Pyme a cliente por cada sitio generado, artículos en general.
Linio Colombia.	Colombia.	Linio Colombia S.A.S	Sí.	Colombia, México, Perú y Venezuela.	Empresa a cliente, artículos en general.
EBay.	Estados Unidos.	eBay Inc.	No. Sin embargo es posible comprar desde Colombia.	+ 30 países.	Cliente a cliente, artículos en general.
Amazon.	Estado Unidos.	Amazon.com, Inc.	No. Sin embargo es posible comprar desde Colombia.	Estados Unidos, Reino Unido, Francia, Canada, Alemania, Italia, España, Australia, Brasil, Japón, China, India, México, entre otros.	Empresa a cliente, libros (nuevos, usados), lectores de libros digitales, películas, música, juegos, computadoras, servicios de almacenamiento, etc. ver más en: <a href="http://www.amazon.com/gp/site-directory/ref=sa_menu_top_fullstore">http://www.amazon.com/gp/site-directory/ref=sa_menu_top_fullstore</a>
Taobao[9].	China.	Alibaba Group.	No. Se aceptan tarjetas de créditos internacionales pero el autor no logró certificar el soporte de transporte para Colombia.	Principalmente en China, Taiwan, Macau.	Cliente a cliente, artículos en general.
Rakuten.	Japón.	Rakuten, Inc.	No. Se aceptan tarjetas de créditos internacionales pero el autor no logró certificar el soporte de transporte para Colombia.	Japón.	Empresa a cliente, productos de hogar, deportivos, alimentos y bebidas, dispositivos electrónicos, libros.

Este tipo de compañías (las listadas arriba y similares), que tienen gran parte de su mercado basado en las operaciones que ofrece su plataforma, se han preocupado por entender las dinámicas del comercio electrónico desde el estudio de su propia estructura, esto es: de entender las interacciones de sus usuarios, de cómo organizar sus productos y cómo visualizarlos. Shen (2011) por ejemplo describe a eBay cómo una red compleja, en la que es vital conocer tanto los tipos de comportamientos internos (ej: nuevos clientes, nuevos vendedores, número de clientes en un segmento de productos), como identificar posibles tendencias que se estén generando (ej: aumento de las ventas de tenis y uniformes de baloncesto por las semifinales de la liga profesional de los Estados Unidos (NBA) o aumento de las ventas de lentes de sol por el verano próximo, etc.). También menciona la importancia de conocer qué estructura tecnológica soporta dicha red (ej: granja de servidores, localización de ellos, funcionamiento de aplicaciones, etc.), qué grado de confianza y qué reputación existe y cómo su evolución puede generar nuevas oportunidades de negocio. En eBay se estudian dichos comportamientos y tendencias aplicando métodos estadísticos para grandes cantidades de datos.

Un requerimiento importante en las aplicaciones de comercio electrónico y en general en cualquier aplicación Web son las búsquedas eficientes de datos y la categorización de estos. Con frecuencia los compradores novatos en Internet no utilizan directamente las plataformas o servicios de compra en línea, sino que llegan al producto que desean comprar desde un motor de búsqueda, cobrando así importancia la influencia o afinidad comercial de los resultados que estos producen. Lewandowski (2011) analiza la influencia comercial de tres de los más grandes motores de búsqueda en la Web: Google, Yahoo! y Bing. Según el estudio los tres motores arrojan resultados con influencia comercial, esto es, publicidad en forma de texto o imagen, o enlaces patrocinados. De los tres Google y Yahoo! son los más afines a resultados comerciales (enlaces a

dominios .com), siendo Google el más eficaz (preciso) y a la vez el que más publicidad despliega en formato de texto. Existen iniciativas como “DuckDuckGo” que implementa un motor de búsqueda neutral, con poca o nula influencia comercial y con cero publicidad.

Si los grandes motores de búsqueda se han preocupado por introducir intención comercial en sus resultados, cuanto más lo habrán hecho las propias plataformas de comercio electrónico. Los compradores en Internet después de haber adquirido cierta experiencia utilizan directamente los sistemas de búsqueda ofrecidos por las tiendas en las que hacen sus transacciones, tradicionalmente estos sistemas utilizan algoritmos de búsquedas basados en la relevancia de la consulta (o parte de ella), sin embargo Long (2012) ha demostrado la efectividad de nuevos algoritmos y modelos de puntuación (en este caso Best-Selling Prediction) con respecto a la identificación de productos con más probabilidad de compra en comparación a modelos clásicos (ej. Vector y Boolean Model, Capítulo 3 en Modern Information Retrieval). Este nuevo framework de ranking (Best-Selling Prediction) reduce la brecha entre la relevancia obtenida en la consulta y una opción real de compra. Su esquema de trabajo consiste en que, a partir de una categorización tradicional, realiza un re-cálculo con dos nuevos factores: El histórico de transacciones y el volumen de ventas. Con esto, según el estudio, el usuario obtiene resultados más acordes a su necesidad de compra, ya que, en efecto, entre los primeros resultados siempre encontrará los productos con más movimiento comercial convenientes para él.

Volviendo a los motores de búsqueda, Calderón-Benávides (2011) desarrolló un marco para identificar la intención del usuario a partir de la cadena dada en la consulta, utilizando técnicas de minería de datos. El estudio menciona que cada búsqueda de usuario requiere o incluye una necesidad y que de acuerdo a esta se deben mostrar los mejores resultados.

Por otra parte Zhou y Mobasher (2006), resaltan la necesidad de estudiar/determinar los inte-

reses y las preferencias de los usuarios en la Web, para tal fin proponen un enfoque que mezcle el componente probabilístico (density-based) y el componente de clasificación (distance-based), este es MFA; Mixture of Factor Analyzers. Con este enfoque, según los autores, varias características se pueden co-relacionar al tiempo dentro de un cluster de usuarios, permitiendo así modelar su navegación.

Na & Gen & Hang & Jiaying (2013) establecieron una correlación entre las búsquedas en sitios de comercio electrónico y sus respectivas órdenes, esto es: a través del análisis empírico de las órdenes de compra y las búsquedas hechas en un e-commerce site chino (octubre de 2011 y febrero de 2012) construyeron un modelo de predicción de compra de productos que se comporta bien dentro de los siete días siguientes. Según los autores, el modelo ayuda a mejorar la administración de inventarios en empresas que venden por Internet. Este estudio hace énfasis en que la predicción de ordenes de compra puede ser más efectiva si se toman en consideración las búsquedas hechas por los usuarios.

En resumen se observan dos tendencias: la creciente iniciativa a modelar y desarrollar nuevos algoritmos que arrojen resultados especializados (filtros con información específica de contexto), y la necesidad de definir modelos que predigan tanto como sea posible la necesidad que origina la búsqueda, El comercio electrónico como lo menciona Xiaojing, &Weiqing, & Liwei (2012) tiene el poder generar nuevas dinámicas en grandes economías, como la China. Estos autores destacan del eCommerce entre otras características las siguientes:

Diversidad de mercados: permite realizar transacciones en mercados internacionales.

Diversidad de contenidos: cada vez más productos y servicios son accesibles desde tiendas en línea.

Sitio y medios virtuales para compra: todo el proceso de compra y venta se desarrolla a tra-

vés de Internet.

**2.1.2 eCommerce y clasificación.** La clasificación y categorización de datos en sistemas de comercio electrónico y en general en cualquier sistema Web es un factor determinante para la experiencia de usuario, un usuario satisfecho con el “look and feel” de una aplicación de e-commerce puede tender a comprar o consultar más productos. Si un ítem de venta es difícil de localizar en una plataforma, seguramente su promedio de venta no será el mejor.

Dan (2011) menciona que una mejor categorización (entendiéndola como una clasificación de productos bajo algún criterio de organización) ayuda tanto a vendedores como compradores, y para lograr este objetivo, proponen un mecanismo basado en grafos (y no en jerarquías tipo hoja/árboles) que gobierne los segundos niveles de categorización (sub-categorías: categorías que están dentro de una categoría padre) y que soporten grandes volúmenes de información. Además demuestra que este tipo de clasificación ayuda a eliminar ítems duplicados, por ejemplo, que un nuevo smartphone de marca Samsung no se registre en dos categorías con nombres “New Samsung devices” y “new samsung mobile devices” ya que para efectos prácticos son las mismas. Una categoría en el dominio del comercio electrónico es un nombre o etiqueta que agrupa un conjunto de productos con ciertas características comunes, *tecnología* puede ser un ejemplo de categoría, ya que esta etiqueta puede agrupar productos de computación y de telecomunicaciones. En general, en este estudio definen un modelo de categorización capaz de: 1) procesar grandes cantidades de datos, 2) manejar con efectividad datos balanceados y 3) hacer predicciones en tiempo real en los ítems de categorización.

Con la evolución de las aplicaciones de e-commerce la cantidad de datos a manejar se ha incrementado considerablemente, y los costos que se requieren para soportar su administración también lo han hecho, de modo que a mayor cantidad de datos más costos habrá en su manteni-

miento. Una de las áreas más complejas dentro del ecosistema del comercio electrónico es el análisis de logs (Web Usage Mining) tanto de consultas de usuario como de operaciones internas, en estas, la supresión de información duplicada puede hacer la diferencia. Este problema ha sido abordado por Da-sheng y Qin-fen (2009), quienes desarrollaron un algoritmo eficiente para la clasificación de textos (similitud de unas cadenas con respecto a otras), aunque tiene ciertas limitaciones para cadenas superiores a 1000 palabras e inferiores a 100, se desenvuelve muy bien dentro del rango de 100 a 1000 palabras. Este algoritmo puede ser utilizado en los clientes Web de e-commerce (Web Content Mining) para detectar duplicidad en las categorías (agrupaciones de productos), de modo que los usuarios siempre obtengan una clasificación más limpia y tratada.

**2.1.3 Consulta y extracción Web.** Cómo se mencionó anteriormente, las consultas eficientes de datos son un aspecto crucial en los sistemas Web en general y aún más en las plataformas de comercio electrónico. Recuperar ítems de relevancia para el usuario de acuerdo a sus necesidades o intenciones hace la diferencia entre un servicio de comercio en línea eficaz o no, y esto es claro para todos los actores y compañías que se desempeñan en el área. La preocupación por el buen procesamiento de las consultas de usuario ha generado un conjunto de técnicas tanto para mejorar el rendimiento en términos de tiempo de respuesta como para mejorar su efectividad con respecto a los resultados esperados por el usuario.

La especificidad de una consulta garantiza o no la obtención de mejores resultados. Hafernik y Jansen (2013), estudiaron cómo la longitud y las partes de la oración pueden afectar la calidad de los ítems obtenidos, en el estudio se tomó un conjunto de logs (queries hechos en el servicio de información de AOL: <http://www.aol.com/>) y para cada uno de estos se evaluó su especificidad, pasándolo por un filtro de 9 indicadores: 1) URL's contenidas en la consulta. 2) Sitios o lu-

gares con términos adicionales en la consulta. 3) Consultas que intentan comparar múltiples ítems. 4) Consultas que contienen ideas distintivas. 5) Consultas que contienen una pregunta con respuesta clara. 6) Consultas que contienen una solicitud de direcciones, instrucciones o tips. 7) Consultas que tienen fechas específicas con términos adicionales. 8) Consultas que contienen un número y términos adicionales. 9) Consultas que contienen un nombre y términos adicionales. Al final se publica un cuadro comparativo con porcentajes de consultas en términos de partes de oración.

Un aporte relevante en el área del análisis de consultas de usuario es el desarrollo de aplicaciones para la segmentación y re-formulación de los String Queries, esto es, aplicaciones capaces de dividir la cadena de consulta, analizarla y ordenarla de tal forma que obtenga los mejores resultados dentro del conjunto de productos ofertados, intentando predecir la intención del comprador, un ejemplo de este tipo de aplicación es “QSEGMENT”, desarrollada por eBay para mejorar su sistema interno de búsqueda.

Las variables tiempo y contexto también han sido tenidas en cuenta a la hora de estudiar la relevancia de resultados, Newsha y Sun (2013), profundizaron en cómo las características de contexto y tiempo puede afectar la relevancia de los resultados y la prontitud de la respuesta, estos investigadores crearon 20 indicadores que ayudan a determinar qué tanta afinidad tienen los resultados de una consulta con respecto a los eventos de la actualidad dado un contexto. El estudio menciona que la correcta predicción de si una consulta está o no relacionada a un evento reciente puede ser crítica para el mejor entendimiento y direccionamiento de la intención de búsqueda del usuario. Concluyen en que las características de contexto son más efectivas que las de tiempo y que la afinidad de una consulta con respecto a los sucesos actuales puede ser determinada con alta exactitud pasándola por el conjunto de indicadores propuestos por ellos.



La extracción de información en plataformas de comercio electrónico también ha ido tomando un papel protagónico, con el auge de los servicios Web y de las aplicaciones nativas para móviles, los sistemas de comercio electrónico se han visto en la obligación de desarrollar mecanismos para publicación de datos en formatos legibles, sea en forma de base de datos públicas, APIs vía JSON/XML, bases de datos, etc. Sin embargo dichas formas de comunicación requieren acuerdo técnico y garantías de compatibilidad entre las partes y esto no siempre se logra. En efecto surgen alternativas como la extracción de datos directamente desde la lectura y análisis del árbol de nodos HTML (DOM) de las aplicaciones (Web Scrapping). Ouyang, Dong y Zou (2009), estudiaron esta forma de extracción, en particular profundizaron en cómo extraer información de ítems comerciales desde el tag HTML table, que se utiliza para publicar información en forma ordenada (filas y columnas). La complejidad de este estudio radicó en la dificultad para determinar si una tabla HTML ubicada en cualquier parte del DOM contenía información comercial o no.

Kumar y Rizvi (2011), exponen tres formas para extraer información de un sitio Web: Web Usage Mining, Web Scrapping y Semantic Annotation. En este estudio se define al Web Usage Mining como el método que consiste en la extracción de patrones de usuario relevantes desde los logs de servidores Web usando técnicas de minería de datos, al Web Scrapping como la técnica de extracción automática de datos desde el HTML de un sitio Web hecha por un software y al Semantic Annotation como el proceso de enriquecer un documento con el fin de establecer conexiones entre el texto y sus descripciones semánticas.

Un ejemplo interesante de la extracción Web y de la Web Content Mining (en el contexto de la Ingeniería del Software) la presenta Dumitro (2011), donde mostró el funcionamiento de un sistema recomendador de funcionalidades o “features”, el cual, a partir de unas características

iniciales deseadas (del producto de software) obtiene todo un modelo de requerimientos básicos y algunas recomendaciones extras, lo anterior con base tanto en la comparación de características dentro de la documentación interna del producto, como de aquellos productos que ofrece la competencia. En otras palabras, este sistema es capaz de determinar que funcionalidades ofrecen los productos de la competencia a través de textos publicitarios en la Internet, para que al momento de la especificación de características para el producto nuevo (del mismo dominio o tipo), se pueda recomendar funcionalidades no pensadas.

Sieg, Mobasher y Burke (2007), identificaron un problema en la personalización de búsquedas de usuario; la mayoría de motores de búsqueda no tienen en cuenta el contexto del usuario. Para corregir esta dificultad proponen un método que hace un “re-ranking” de resultados basados en ontologías de perfiles de usuario. Concluyeron que las búsquedas personalizadas son más efectivas cuando se usan ontologías de perfiles de usuario, además resaltaron que las búsquedas ambiguas se resuelven con la evidencia semántica que brindan los perfiles de usuario.

## **2.2 Marco Teórico**

La base teórica de la investigación está altamente influenciada por el área de la computación llamada Information Retrieval, se definen tópicos relacionados a buscadores Web, extracción de datos, naturaleza de APIs, tipos de plataformas de comercio electrónico, entre otros. A continuación un mapa conceptual con el contenido del capítulo

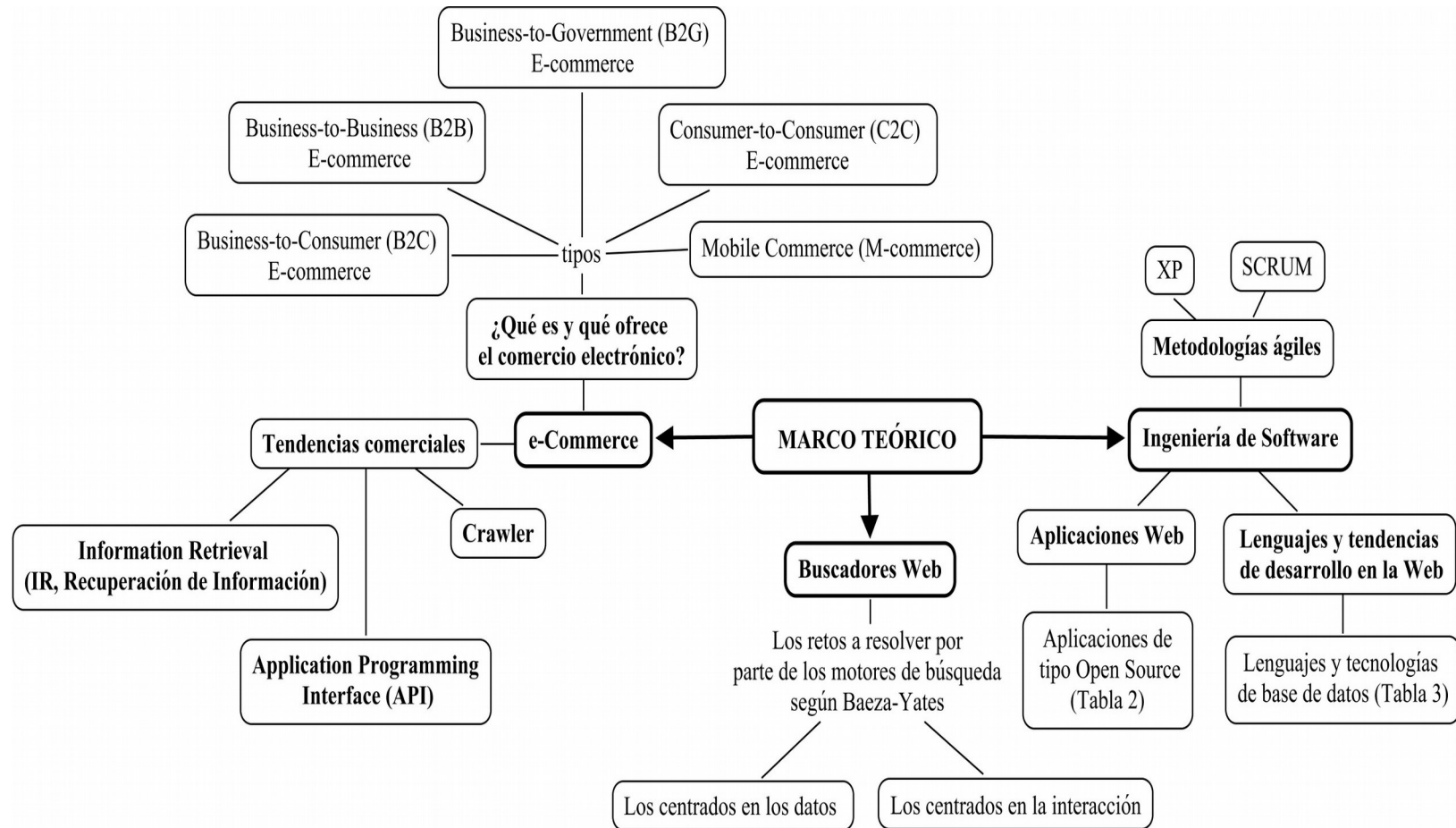


Figura 2. Estructura del marco teórico

**2.2.1 e-Commerce.** Según Arias Pou (2006), el comercio electrónico, es una actividad económica que se desarrolla por medios electrónicos y a la que se le aplican todas las formas del comercio tradicional, y además, las normas específicas del medio por el que se desarrolla. En el contexto actual, la anterior definición comúnmente se traduce como grandes y/o medianas aplicaciones Web que ofrecen catálogos de productos, promociones, formas de pago, clientes móviles, rastreo de pedidos, servicio de envío, etc., en otras palabras es un conjunto de software que ofrece servicios de venta y compra de productos junto a toda la infraestructura de hardware y de logística que necesita para su funcionamiento.

Existen 5 tipos de comercio electrónico (Andam, 2003):

- Business-to-Consumer (B2C) E-commerce. Consiste en comercio electrónico entre compañías y consumidores.
- Business-to-Business (B2B) E-commerce. Consiste en comercio electrónico únicamente entre compañías.
- Business-to-Government (B2G) E-commerce. Consiste en cualquier forma de comercio electrónico entre las compañías y el sector público.
- Consumer-to-Consumer (C2C) E-commerce. Consiste en comercio electrónico únicamente entre consumidores.
- Mobile Commerce (M-commerce). Consiste en cualquier forma de comercio electrónico que involucre transacciones desde dispositivos móviles.

Las plataformas de comercio electrónico son muy populares, cada vez es más fácil comprar productos desde tiendas virtuales y recibirlos en cualquier parte del mundo. El comercio electrónico es la mayor tendencia sobre la Web en nuestros días y ha beneficiado a millones de personas (Baeza-Yatez & Ribeiro-Neto, 2011).

**2.2.2 Tendencias comerciales.** En el ámbito de una plataforma de comercio electrónico, el autor define una tendencia comercial como el comportamiento que probablemente tendrá un artículo o una categoría de estos en un periodo determinado, tomando como base el análisis de las transacciones que los involucra y el análisis de un contexto específico. En esta investigación se trabajará en la identificación de tendencias de corto tiempo, esto es comportamientos en el rango de 1 a 3 semanas (Murphy, 1999).

**2.2.3 Information Retrieval (IR, recuperación de información).** El producto de software que se desarrollará en el contexto de la presente investigación se puede categorizar como un sistema de tipo IR (recuperación de información), este tipo de sistemas fue definido por Baeza-Yates y Ribeiro-Neto en la segunda edición de su libro *Modern Information Retrieval* como: “... una amplia área de las ciencias de la computación enfocada primordialmente en proveer a los usuarios fácil acceso a la información de su interés.”. También plantearon el problema fundamental al que se enfrentan estos sistemas: “El problema IR: la meta primaria de un sistema IR es recuperar todos los documentos que son relevantes como respuesta a una consulta de usuario...”. Con estas definiciones, queda clara la naturaleza de las aplicaciones IR: obtener información relevante para el usuario con la menor cantidad de ruido (información no relevante).

Detrás de un IR se manejan conceptos como índices, interfaces de usuario, filtros, rankings, etc. todos estos se unen para formar aplicaciones que responden de manera eficiente y rápida a consultas de usuario. En la actualidad se pueden observar gran cantidad de sistemas IR: software de gestión de bibliotecas, motores de búsqueda, sistemas de comercio electrónico, etc.

**2.2.4 Application Programming Interface (API).** En el diccionario de términos tecnológicos TechTerms se propone una definición de API, esta es: “una API es un conjunto de comandos, funciones y protocolos los cuales pueden ser usados por los programadores que construyen

software para un sistema operativo específico”. La anterior definición ciertamente es muy amplia, desde la experiencia del autor en el área de desarrollo de aplicaciones Web se puede entender una API de dos formas:

La primera forma como el conjunto de entidades internas y estructurales que conforman una aplicación junto a las definiciones de sus interfaces, por ejemplo: la API del lenguaje de programación JAVA que se encuentra en la URL <http://docs.oracle.com/javase/7/docs/api/>. En este caso se trata de un conjunto de métodos y clases con documentación, que retornan valores escalares, booleanos, etc. Normalmente explican el funcionamiento interno de una aplicación.

La segunda forma como el conjunto de funcionalidades que ofrece una aplicación para interactuar con otras, junto a las definiciones de sus interfaces. Por ejemplo: la API de Twitter que se encuentra en la URL <https://dev.twitter.com/docs/api/1.1>. En la actualidad muchas APIs de este tipo son Web, y están bajo arquitectura REST, se trata de interfaces que permiten obtener y/o modificar recursos a través de peticiones HTTP con una determinada semántica utilizando documentos en formato JSON o XML.

Las APIs están jugando un papel relevante en el ecosistema de las aplicaciones en la nube y las aplicaciones móviles, son estas interfaces de software las que han permitido que un entorno Web comparta información con aplicaciones de otras arquitecturas. La mayoría de las plataformas de comercio electrónico ofrecen APIs para la integración con software de terceros, por ejemplo Mercado Libre tiene la suya en <http://developers.mercadolibre.com>. Con una API de este tipo se podría desarrollar una aplicación móvil que permita comprar artículos desde smartphones o tablets.

**2.2.5 Crawler.** Un crawler, también conocido como Web Spider, Web robot, o simplemente un Bot, es un software para descargar páginas automáticamente de la Web (Baeza-Yatez & Ribe-

ro-Neto, 2011).

Ceri , Bozzon, Brambilla, Della, Fraternali & Quartoneroni. (2013) brindan una definición más técnica en su libro Web Information Retrieval, ellos definen al Web Cralwer como “el proceso automático que analiza recursos en la Web con el fin de extraer información y los hipervínculos que los interconectan.”

Baeza-Yatez y Ribeiro-Neto (2011) describen el funcionamiento del Crawler de la siguiente forma: “El crawler toma como entrada un conjunto de páginas de inicio o una semilla de páginas, que son descargadas, parseadas, y escaneadas para encontrar nuevos vínculos. Los vínculos que apuntan a páginas que no han sido aún descargadas son agregados a la cola central de URLs para una posterior descarga. Siguiendo a esto, el crawler selecciona una nueva página de la cola de descarga y el proceso es repetido hasta que un criterio de parada es encontrado. ”

Cómo aplicaciones reales del uso de crawlers se tienen los siguientes ejemplos:

- Buscadores Web generales.
- Buscadores Web verticales (temas específicos).
- Archivadores Web (archivadores de contenido).
- Caracterizadores Web (extracción de estadísticas).

En el conjunto de los buscadores Web verticales se encuentra un tipo especial de crawler llamado Shop-Bot que permite descargar información desde catálogos de tiendas en línea y provee una interfaz para la comparación de precios de una forma centralizada (Baeza-Yatez & Ribeiro-Neto, 2011).

Existe un tipo especial de software llamado Web Scrapping que tiene varias similitudes con el modo de funcionamiento de un crawler pero no se puede considerar como uno de estos, el Web Scrapping busca extraer datos concretos de un documento Web (objetivo seleccionado) sin utilizar

una semilla de páginas o sin utilizar escaneo iterativo. Dicho de otra forma, mientras que el crawler busca descargar páginas web, indexarlas y extraer de ellas hipervínculos para futuros paseos, el Web scraping tiene un objetivo seleccionado y busca procesar un documento Web para extraer datos específicos. En esta investigación se utilizara la técnica del Web Scraping para aquellos sitios de comercio electrónico seleccionado que no ofrezcan una API para desarrolladores.

**2.2.6 Buscadores Web.** Un buscador Web es un sistema que responde a consultas basadas en palabras con la extracción de resultados, los cuales apuntan a páginas Web (Ceri , Bozzon, Brambilla, Della, Fraternali & Quartoneroni).

Baeza-Yates & Ribeiro-Neto (2011) mencionan que hay básicamente dos formas principales de explorar la Web, la primera es realizar una consulta de usuario basada en palabras sobre un motor de búsqueda que indexe una porción de documentos Web, por ejemplo: el procesamiento que se hace al buscar un tema específico en Google. La segunda forma es navegar la Web en sí misma, esto es, seguir los hipervínculos de los recursos partiendo de uno específico, hasta llegar al más relevante según el contexto del usuario.

Los retos a resolver por parte de los motores de búsqueda según Baeza-Yates (2011) son de dos tipos: los retos centrados en los datos y los centrados en la interacción.

Desafíos centrados en los datos:

- Datos distribuidos.
- Alto porcentaje de datos volátiles.
- Gran volumen de datos.
- Datos redundantes y sin estructura.
- Calidad de los datos.



- Datos con formatos heterogéneos.

Desafíos centrados en la interacción:

- Interpretación de resultados.
- Expresividad de las consultas.

La mayoría de motores de búsqueda usan una arquitectura centralizada de crawler-indexer que inspecciona la Web enviando páginas nuevas o páginas actualizadas a un servidor central donde estas son indexadas (Baeza-Yatez & Ribeiro-Neto, 2011).

En las plataformas de comercio electrónico se ofrecen también buscadores internos especializados en la recuperación de productos, haciendo uso intensivo de las categorizaciones, ofertas, disponibilidad de artículos, etc.

**2.2.7 Ingeniería de Software. Aplicaciones Web.** Es un aplicación de software escrita en el contexto Web (se ejecuta en un navegador Web), normalmente cuenta con los siguientes componentes: HTML como lenguaje de marcado, CSS como lenguaje de presentación, JavaScript como lenguaje del lado del cliente y algún lenguaje del lado del servidor (PHP, Java, Python, etc.) para interacción con base de datos y lógica del negocio. Al igual que el software que se ejecuta en un entorno de escritorio, el software Web tiene especializaciones, una de estas justamente es el e-commerce, esto es, programas para el soporte y construcción de plataformas de comercio electrónico. En la actualidad existen varias aplicaciones de tipo Open Source para construir una tienda en línea, algunas de estas aplicaciones son:

**Tabla 2. Aplicaciones**

Aplicación	Sitio Web
PrestaShop	<a href="http://www.prestashop.com/">http://www.prestashop.com/</a>
Magento	<a href="http://magento.com/">http://magento.com/</a>
osCommerce	<a href="http://www.oscommerce.com/">http://www.oscommerce.com/</a>
Plugin de E-Commerce para Wordpress	<a href="http://getshopped.org/">http://getshopped.org/</a>

*Lenguajes y tendencias de desarrollo en la Web.* Como se puede vislumbrar en el apartado anterior, una aplicación Web se compone normalmente de tres capas: La capa del cliente, en la cual se define la interactividad con el usuario, la presentación y distribución de datos, todas estas en el contexto del navegador Web. La capa del servidor, en la cual se implementa toda la lógica del negocio. Finalmente está la capa de datos, en la cual se persiste todo el modelo de datos que soporta la aplicación. En cada una de estas capas se utilizan diferentes tecnologías, se selecciona una u otra de acuerdo a criterios como rendimiento, compatibilidad, madurez, etc.

Para el desarrollo de aplicaciones Web actualmente se han venido utilizando los siguientes lenguajes y tecnologías de base de datos.

**Tabla 3. Lenguajes y tecnologías de base de datos**

FrontEnd (client side)	BackEnd (server side)	Database
JavaScript, CSS, HTML5	C/C++, Java, Python, PHP, Go, ASP.NET, Ruby, Node.js, Scala, C#.	MySQL, Microsoft SQL Server, Oracle Database, MongoDB, Redis, Cassandra, MariaDB.

Fuente: Techcrunch, 2012.

A continuación, algunas tendencias en el área de desarrollo de aplicaciones Web, todas estas desde la perspectiva y experiencia del autor:

- *SPA (Simple Page Application):* Son aquellas aplicaciones Web que se ejecutan mayoritariamente sobre una sola página Web. David Pitt en *Modern Web Essentials*, agrega sobre el enfoque SPA: “El enfoque SPA le permite a las compañías proveer a sus usuarios aplicaciones ricas y responsivas a través del navegador Web”. Estas aplicaciones hacen uso intensivo del JavaScript, la idea principal es que el contenido y las acciones de la aplicación actualice solo secciones de páginas, en vez de páginas completas.

- *JavaScript del lado del servidor (Node.js)*: En su sitio Web se define a Node como “una plataforma construida sobre el entorno de ejecución Javascript de Chrome, para fácilmente construir rápidas y escalables aplicaciones de red. Node.js usa un modelo dirigido por eventos, de E/S no bloqueante que lo hace ligero y eficiente, perfecto para aplicaciones en tiempo real con uso intensivo de datos que corran a través de servicios distribuidos”. En términos del autor, Node es una plataforma que permite construir aplicaciones Web con JavaScript del lado del servidor. Recientemente Node se ha estado utilizando para aplicaciones de tipo “real-time” en contextos Web, además para la construcción de API's tipo REST.
- *MVC JavaScript Frameworks*: Son librerías Javascript que permiten estructurar el código del cliente (Front-End) bajo el patrón de arquitectura MVC o derivados. Ejemplo de estas librerías: AngularJS (<https://angularjs.org/>), EmberJS (<http://emberjs.com/>), BackboneJS (<http://backbonejs.org/>). Este tipo de frameworks están potenciando el desarrollo de aplicaciones tipo SPA.
- *Frameworks*: DocForge en su wiki define los frameworks como: “Un conjunto de código fuente o librerías que proveen funcionalidad común a una gran cantidad de aplicaciones”. Menciona también: “Mientras que una librería usualmente proveerá una pieza específica de funcionalidad, un framework ofrecerá un rango más amplio de funcionalidades, las cuáles son todas comúnmente usadas por un tipo de aplicación”. El autor define a un framework cómo un conjunto de piezas de software que permiten el desarrollo de aplicaciones de forma más rápida y ciertamente más seguras. Es común que los desarrollos actuales en el área Web se implementen bajo un framework con algún lenguaje del lado del servidor, por mencionar algunos tenemos: Symfony (<http://symfony.com/>), Django (<https://www.djangoproject.com/>), Ruby On Rails (<http://rubyonrails.org/>) Express (<http://expressjs.com/>).

- *Pre-procesadores CSS*: Un pre-procesador de CSS es un mecanismo que extiende el CSS común, esto es, un mecanismo que le agrega funcionalidades como variables, mixins, funciones, etc. a la capa de estilos de una aplicación Web. Entre los pre-procesadores más comunes se encuentra Less (<http://lesscss.org/>), Sass (<http://sass-lang.com/guide>) y Stylus (<http://learnboost.github.io/stylus/>).
- *Uso de metodologías ágiles*: Con el constante crecimiento de startups o emprendimientos a nivel de las tecnologías de la información y comunicación, el uso de metodologías ágiles de desarrollo ha venido tomando un papel importante, es común en empresas pequeñas y no tradicionales escuchar usos de SCRUM o XP ya que proponen el desarrollo de software sobre valores “prácticos”: pocos desarrolladores, más código menos trámites, etc. En el siguiente apartado se detallará la definición y algunos ejemplos de metodologías ágiles.
- *Bases de datos no SQL*: De acuerdo a la Wikipedia “Las bases de datos No SQL son una extensa clase de sistemas administradores de base de datos identificados por su no-adherencia al ampliamente usado modelo relacional de sistemas administradores de bases de datos; esto es, Las bases de datos No SQL no son primordialmente construidas sobre tablas, y como resultado, generalmente no usan SQL para la manipulación de datos.” Dan McCreary y Ann Kelly en Making Sense of NoSQL definen el movimiento No SQL como: “un conjunto de conceptos que permiten el procesamiento rápido y eficiente de datos, enfocado en el rendimiento, la confiabilidad y la agilidad”. Algunos ejemplos de bases de datos no relaciones son: MongoDB (<http://docs.mongodb.org/manual/>), Redis (<http://redis.io/>) y Cassandra (<http://planetcassandra.org/what-is-apache-cassandra/>). Últimamente, desde la experiencia del autor, se han estado desarrollando aplicaciones Web con bases de datos No SQL, porque éstas ofrecen más velocidad y flexibilidad en comparación con las bases de datos relacionales.

**2.2.8 Metodologías ágiles.** Las metodologías ágiles de desarrollo de software fundamentan su funcionamiento en el manifiesto ágil y en los principios detrás de este. Existen varios marcos y metodologías pertenecientes a esta vertiente, Blankendship, Bussa y Millett detallan seis metodologías ágiles en su libro *Pro Agile .NET Development with Scrum: Scrum, eXtreme Programming, Crystal, Dynamic Systems Development Method, Feature-Driven Development y Lean Software Development*. Para el desarrollo del producto de software de esta investigación se utilizarán algunas prácticas de dos de estas: SCRUM y XP.

Blankendship (2011) definen el marco de trabajo SCRUM (2014) cómo una metodología que consiste en una serie de sprints, frecuentemente de dos semanas de duración, en dónde cada uno de estos tiene como resultado un entregable funcional de software para el cliente. Los sprints son dirigidos por la pila del producto (lista de requerimientos), la cual es gestionada por el dueño del producto.

Blankendship (2011) también menciona sobre XP (2014) que es una metodología de fuerte interacción con el cliente, dónde los requerimientos se definen a través de historias de usuario, estas historias son estimadas por los desarrolladores y priorizadas por el cliente. Las historias se implementan en iteraciones pequeñas.

## **2.2.9 Métodos de tendencia lineal**

**2.2.9.1 Método gráfico (diagrama de dispersión).** Consiste en dibujar un gráfico de dispersión (nube de puntos), donde cada caso observado aparece representado como un punto en el plano cartesiano. Es una herramienta gráfica que permite identificar la posible relación entre dos variables, con el uso del coeficiente de correlación se mide que tanta correspondencia existe entre el conjunto variables observadas. El lector puede profundizar este método en las siguientes URLs:

[http://www.vitutor.net/2/16/diagrama\\_dispersion.html](http://www.vitutor.net/2/16/diagrama_dispersion.html)

[http://www.ub.edu/aplica\\_infor/spss/cap3-7.htm](http://www.ub.edu/aplica_infor/spss/cap3-7.htm)

**2.2.9.2 Regresión lineal simple.** Es una técnica estadística para estudiar la relación entre dos o más variables, se utiliza para predecir un rango amplio de eventos, básicamente consiste en hallar la mejor recta que se ajuste a la nube de puntos (diagrama de dispersión), las formulas para hallar estas rectas se definen a continuación:

$$y = \bar{y} + \frac{\sigma_{xy}}{\sigma_x^2}(x - \bar{x})$$

**Figura 3. Recta de regresión de Y sobre X**

$$x = \bar{x} + \frac{\sigma_{xy}}{\sigma_y^2}(y - \bar{y})$$

**Figura 4. Recta de regresión de X sobre Y**

Para determinar el grado de relación entre las variables, al igual que en el diagrama de dispersión se utiliza el coeficiente de correlación que se define como:

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

**Figura 5. Coeficiente de correlación**

El lector puede profundizar este método estadístico en la URL [http://www.vitutor.net/2/16/regresion\\_lineal.html](http://www.vitutor.net/2/16/regresion_lineal.html).

**2.2.9.3 Regresión lineal por mínimos cuadrados.** La universidad de las Américas de Puebla publicó un documento en la URL [http://catarina.udlap.mx/u\\_dl\\_a/tales/documentos/lad/mercado\\_g\\_ja/apendiceC.pdf](http://catarina.udlap.mx/u_dl_a/tales/documentos/lad/mercado_g_ja/apendiceC.pdf) que explica el método de los mínimos cuadrados como sigue:

Consiste en minimizar la suma de los cuadrados de las distintas verticales entre los valores de los datos y los de la regresión estimada, es decir, minimizar la suma de los residuos al cuadrado, teniendo como residuo la diferencia entre los datos observados y los valores del modelo. Dado

un conjunto de  $n$  parejas  $(x_i, y_i)$  de datos, se debe encontrar la ecuación de la recta  $y = ax + b$  que pasa lo más cerca posible de los puntos experimentales (de forma que éstos estén repartidos uniformemente alrededor de la recta). El método del ajuste por mínimos cuadrados permite obtener la pendiente  $a$  de la recta y la ordenada  $b$  en el origen, correspondientes a la recta  $y = ax + b$  que mejor se ajusta a los  $n$  datos  $(x_i, y_i)$ , es decir, permite establecer una relación funcional entre dos variables; donde  $x$  es la variable independiente e  $y$  es la variable dependiente.

La ecuación para determinar los coeficientes  $a$  y  $b$  son:

$$a = \frac{n \sum_{i=1}^n x_i y_i - \left( \sum_{i=1}^n x_i \right) \left( \sum_{i=1}^n y_i \right)}{n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2} \quad b = \frac{\sum_{i=1}^n y_i - a \sum_{i=1}^n x_i}{n}$$

### Figura 6. Cálculo de parámetros por mínimos cuadrados

Para medir el grado de relación entre las variables  $x$  e  $y$  se puede calcular el coeficiente de determinación o el coeficiente de correlación que se definen como:

$$r^2 = \frac{\sum (\hat{Y} - \bar{Y})^2}{\sum (Y^\circ - \hat{Y})^2 + \sum (\hat{Y} - \bar{Y})^2} = \frac{\sum (\hat{Y} - \bar{Y})^2}{\sum (Y^\circ - \bar{Y})^2}$$

### Figura 7. Coeficiente de determinación

$$\sqrt{r^2}$$

### Figura 8. Coeficiente de correlación

### 3. Marco Legal

El proyecto está enmarcado en normatividad de derechos de propiedad intelectual, registro de software, licencias de publicación de aplicaciones y términos de uso de páginas Web. A continuación se detallan alguna de estas.

Constitución Política de Colombia 1991. El artículo 61 establece la Protección a la Propiedad Intelectual que incluye el siguiente texto: El estado protegerá la propiedad intelectual por el tiempo y mediante las formalidades que establezca la ley.

Decreto 1360 del 23 de Junio de 1989. Por el cual se reglamenta la inscripción del soporte lógico (software) en el registro nacional de derechos de autor.

Art. 1. De conformidad con lo previsto en la ley 23 de 1982 sobre derechos de autor el soporte lógico (software) se considera como una creación propia del dominio literario.

Art. 2. El soporte lógico (software) comprende uno o varios de los siguientes elementos: el programa del computador, la descripción del programa y el material auxiliar.

Art. 3. Para efectos del artículo anterior se entiende por:

Programa de Computador: La expresión de un conjunto organizado de instrucciones, en el lenguaje natural o codificado, independiente del medio en el que encuentra almacenado, cuyo fin es el de hacer que una máquina capaz de procesar información, indique, realice u obtenga una información, una tarea, o un resultado específico.

Material Auxiliar: Todo material, distinto de un programa de computador o de una descripción del programa creado para facilitar su comprensión o aplicación, como por ejemplo descripción de problemas e instrucciones para el usuario.

Descripción del Programa: Una presentación completa del procesamiento idóneo, lo suficientemente detallada para determinar un conjunto de instrucciones que constituya el programa de



computador correspondiente.

Art. 7. La protección que otorga el Derecho de Autor al soporte lógico (software) no excluye otras formas de protección por el derecho común.

Art. 8. Este decreto rige a partir de su fecha de publicación.

Circular Nro. 6 del 15 de abril de 2002 de la Unidad Administrativa Dirección Nacional de Derechos de Autor, Derechos de autor en el ámbito universitario.

El derecho de autor es un reconocimiento que el Estado hace a los autores, a través de la Constitución y la Ley, respecto de sus obras literarias y artísticas, al entregarles instrumentos que les permiten reivindicar su condición de titulares sobre las mismas. Estos derechos surgen en favor del autor sin considerar el fin con cual fue creada la obra, siendo además irrelevante la calidad del creador, es decir, la ley no distingue si es un estudiante, un profesor o un investigador, así como tampoco es preciso establecer dónde tuvo lugar la creación o el tiempo que se haya utilizado, a efectos de esa misma protección. Así, los derechos de autor sobre una obra literaria o artística, como lo sería un trabajo de grado, son de la persona que la realizó, quien la elaboró imprimiendo todo su ingenio e inteligencia. Es su expresión la que queda plasmada en lo producido, siendo por lo tanto el titular de los derechos morales y patrimoniales de la creación. En consecuencia, si la obra es realizada por un estudiante, será él, a la luz de la legislación vigente en materia de derecho de autor, el titular de todas las prerrogativas y facultades que la misma concede... Es menester señalar que el artículo 6o de la Ley 23 de 1982 y 7o de la Decisión Andina 351 de 1993, consagran el principio universal de “la no protección de las ideas”. Por lo tanto, aun cuando el director realiza una valiosa labor de apoyo al aportar ideas, dicha contribución no está protegida por el derecho de autor”.

Licencia Creative Commons 4.0 Internacional. Este documento se publica bajo la licencia

Creative Commons Atribución – No Comercial – Compartir igual 4.0 Internacional, cuyas características son:

- El material puede ser distribuido, copiado y exhibido por terceros si se muestra en los créditos el autor de la obra.
- No se puede obtener ningún beneficio comercial.
- Las obras derivadas tienen que publicarse bajo la misma licencia.

En Creative Commons (2014), se puede consultar la definición completa de la licencia.

Licencia GNU Affero General Public License 3.0: El producto de software que se desarrollará en esta investigación se publicará bajo la licencia AGPLv3.0, cuyas características son:

- Respetar las cuatro libertades del software libre.
- Obliga a las aplicaciones Web a publicar su código.
- Es compatible con la GNU GPL v3.0.

En GNU, (2014) se puede consultar la definición completa de la licencia.

Ley estatutaria 1581 de 2012 (Reglamentada parcialmente por el Decreto Nacional 1377 de 2013). Por la cual se dictan disposiciones generales para la protección de datos personales.

La aplicación (Interfaz Web) desarrollada no almacenará información personal de usuarios, esto es, para su uso solo se solicitará login y password. A su vez el crawler no extraerá datos de usuarios registrados en los sitios de e-Commerce seleccionados, se pretende extraer datos de productos y ofertas, no de perfiles, por lo tanto, la ley de protección de datos no interfiere en el desarrollo de la investigación.

Con respecto al Web Scraping, no existe una ley que reglamente el uso de esta técnica, sin embargo el autor plantea las siguientes consideraciones:

- El Web Scraping automatiza algunas operaciones que el usuario realiza manualmente sobre

un sitio Web, agregándole el plus de poder “descargar” o extraer datos (públicos o accesibles a través de una cuenta activa) para su procesamiento local, es decir, por fuera del servidor desde dónde se extrajo. Desde este punto de vista de automatización el autor no considera delito inspeccionar una página Web a través de un software especializado, si este simplemente realiza las mismas operaciones que un usuario común y validado ejecuta.

- Siempre que los términos de uso de un sitio Web no prohiban la manipulación o transformación de datos, el Web Scraping es legal.

A partir de este análisis, un criterio definido para la selección de los sitios de e-Commerce a estudiar fue: si sus términos de uso permitían o no, que su DOM fuera leído por software automatizado de extracción de datos.

Las anteriores normas fueron escogidas porque cobijan las áreas de acción de la investigación, tanto en su fase de desarrollo como en su fase de publicación. En la fase de desarrollo entran en juego las normas relacionadas a la propiedad intelectual, derechos de autor, extracción de datos, protección de datos. Para la fase de despliegue se mencionaron las licencias de publicación tanto para el documento como para el producto de software, además del proceso de inscripción de software ante la oficina de derechos de autor.

## **4. Método de la Investigación**

Este capítulo comienza describiendo el enfoque metodológico del proyecto, luego, detalla la muestra utilizada en la selección de los sitios Web a analizar y especifica el propio método para llevar a cabo dicha selección, además de los criterios utilizados para la priorización de las plataformas. Posterior a la sección de las tiendas electrónicas, se define para cada una de estas la mejor estrategia de extracción de datos. Finalmente termina con la selección del método estadístico para la proyección de comportamiento y con la secuencia de pasos ejecutadas para la construcción del producto.

### **4.1 Enfoque metodológico**

Esta investigación es de tipo descriptiva, porque intenta detallar los pasos para generar una herramienta (o aproximación de esta) que permita identificar de forma automática tendencias comerciales a través del análisis de sitios de comercio electrónico. También puede ser vista como una investigación exploratoria, en el sentido de que la meta final es identificar tendencias comerciales a través de patrones o movimientos comerciales, y esto es un tema poco documentado / estudiado en el país (el autor hace esta afirmación con base en los estudios encontrados para la construcción del estado del arte).

### **4.2 Definición de población y muestra**

La población de la investigación estuvo conformada por todas las tiendas electrónicas registradas en la cámara colombiana de comercio electrónico en el año 2014 (CCCE). Para la selección de la muestra a trabajar, se tomó como referencia las catorce plataformas de comercio electrónico de propósito general que aparecieron en el reporte de abril de 2014 de la CCCE como sitios de uso frecuente por parte de los colombianos para compras por Internet. A continuación la relación de dichos sitios.

**Tabla 4. Muestra de tiendas electrónicas**

Código	Nombre	URL
ml	Mercado Libre Colombia	<a href="http://www.mercadolibre.com.co">http://www.mercadolibre.com.co</a>
mn	Tiendas Monomi	<a href="http://monomi.co/galeria">http://monomi.co/galeria</a>
olx	OLX Colombia	<a href="http://colombia.olx.com.co">http://colombia.olx.com.co</a>
ln	Linio Colombia	<a href="http://www.linio.com.co">http://www.linio.com.co</a>
ebay	Ebay	<a href="http://www.ebay.com">http://www.ebay.com</a>
amz	Amazon	<a href="http://www.amazon.com">http://www.amazon.com</a>
tmart	Tmart	<a href="http://www.tmart.com/">http://www.tmart.com/</a>
bg	BangGood	<a href="http://www.banggood.com/">http://www.banggood.com/</a>
df	Dafiti	<a href="http://www.dafiti.com.co">http://www.dafiti.com.co</a>
ex	Exito	<a href="http://www.exito.com/">http://www.exito.com/</a>
fl	Falabella Colombia	<a href="http://www.falabella.com.co/">http://www.falabella.com.co/</a>
al	Alibaba	<a href="http://www.alibaba.com/">http://www.alibaba.com/</a>
alk	Alkosto	<a href="http://www.alkosto.com/">http://www.alkosto.com/</a>
hm	HomeCenter	<a href="http://www.homecenter.com.co">http://www.homecenter.com.co</a>

### 4.3 Descripción general de los sitios

**4.3.1 Mercado libre Colombia.** Es una plataforma de comercio electrónica que ofrece a los usuarios registrados servicios como compra, venta, pagos y subastas. En Mercado Libre se ofrecen productos en general, tanto nuevos como usados, tiene presencia en varios países de latinoamérica, entre ellos Colombia.

**4.3.2 Tiendas Monomi.** Es una plataforma construida por la compañía Monoku (<http://monoku.com/>) que permite a pequeños y medianos empresarios construir y personalizar tiendas electrónicas sin conocimientos técnicos. Tiendas Monomi es un producto Colombiano, actualmente tiene presencia en México y Colombia.

**4.3.3 OLX Colombia.** Es un servicio Web que permite la publicación de anuncios gratuitos, en OLX se ofrece un producto nuevo o usado, pero no se vende a través del sitio, en cambio, se

contacta directamente al dueño del producto para finalizar la compra.

**4.3.4 Linio Colombia.** Es una tienda online, en Linio los usuarios regulares solo compran productos, no pueden ofertar o vender a través del sitio. La tienda cuenta con un catálogo extenso de productos, además ofrece servicios de entrega puerta a puerta y pagos electrónicos. Linio tiene presencia en Colombia, Chile, México, Panamá, Perú y Venezuela.

**4.3.5 eBay.** eBay es una plataforma Web internacional de compra y venta de productos, cuenta con un gran catálogo de productos, funciona de manera similar a Mercado Libre, tiene soporte multi-moneda y un servicio de entrega a varios países del mundo.

**4.3.6 Amazon.** Amazon es una de las plataformas de comercio electrónico más grande del mundo, cuenta con un inmenso catálogo de productos tanto nuevos como usados, además de heterogéneos, en Amazon los usuarios puede comprar desde un reloj hasta un servidor virtual. Tiene presencia global, cuenta con soporte para pago electrónico y entrega a casi cualquier lugar del mundo.

**4.3.7 Tmart.** Es una tienda en línea con un catálogo especializado en el que resaltan los productos electrónicos (celulares, linternas, ordenadores, accesorios, etc.), tiene cobertura para entregas en Colombia y cuenta con alrededor de 10 sitios internacionales (Estados Unidos, Rusia, Alemania, Italia, entre otros) con los que hacen presencia a nivel global.

**4.3.8 BangGood.** Es un sitio de comercio electrónico asiático, con soporte de ventas para varios países del globo, ofrece varios métodos de pago y dentro de su catálogo hay más de 70000 productos, es una compañía fuerte en la región asiática. BangGood se comporta como Linio, es una tienda en donde solo se pueden comprar productos.

**4.3.9 Dafiti.** Es un sitio de moda online, con un catálogo enfocado en zapatos, ropa, accesorios, productos de belleza y para el hogar. En Dafiti los usuarios solo compran productos, la tien-

da opera toda la logística de inventarios y entregas.

**4.3.10 Éxito.** Es la implementación Web de la tradicional cadena de tiendas Exito en Colombia, ofrece gran variedad de productos. La compañía se encarga de toda la logística necesaria para el manejo de inventarios y entregas. En la tienda Exito los usuarios solo pueden realizar compras, la compañía ofrece varias formas de pago, entre ellas su propia tarjeta de crédito.

**4.3.11 Falabella Colombia.** Es la implementación Web de la cadena de tiendas Falabella, ofrece gran variedad de productos, entre estos, construcción, mejoramiento del hogar, seguros, viajes. La compañía se encarga de toda la logística necesaria para el manejo de inventarios y entregas. Falabella ofrece varias formas de pago y cuenta con su propia entidad financiera (banco).

**4.3.12 Alibaba.** Es una de las plataformas de comercio electrónico mas fuertes en Asia, con soporte para 190 países, con millones de productos en su catálogo. Está al aire desde 1999 y es una tienda en la que los usuarios solo pueden realizar compras al por menor y de pago en línea, cuenta con su propio motor de comparación de precios.

**4.3.13 Alkosto.** Es el sitio Web de las tiendas Colombianas Alkosto, con un catálogo enfocado en electrodomésticos, llantas, motos, mercado y hogar. Alkosto ofrece a través su portal varias formas de pago, entre ellas todas las tarjetas de crédito, incluida la tarjeta Exito, además de todas las tarjetas débito.

**4.3.14 HomeCenter.** Es el sitio Web de las tiendas HomeCenter, cuenta con un catálogo especializado en todas las líneas del hogar, ofrece servicio de entrega a varias ciudades del país y varias opciones de pago.

#### **4.4 Selección de sitios Web**

**4.4.1 Criterios de selección para los Sitios Web de comercio electrónico a modelar.** Teniendo en cuenta los requerimientos básicos para el desarrollo del crawler, la tabla 5 presenta los

criterios utilizados para la sección de los sitios Web de comercio electrónico. La tabla presentada incluye una descripción y justificación de cada criterio.

**Tabla 5. Criterios de selección de tiendas electrónicas**

Criterio	Descripción	Justificación
Términos de uso del sitio Web.	Se refiere a los términos de uso (licencia de uso) de los sitios Web, tendrán más peso en este criterio aquellas plataformas que no prohíban la extracción de datos vía Scraping.	Permitirá o no aplicar la técnica Web Scraping.
Disponibilidad de una API.	Se refiere a la existencia de una API para desarrolladores.	Permitirá o no obtener datos de los sitios vía API.
Presencia en Colombia.	Se refiere a sitios que soporten pagos con moneda nacional (pesos) y que tengan alcance geográfico para la entrega de productos (que haya envíos a Colombia).	Permitirá filtrar sitios, en esta investigación se pretende modelar sitios donde existan Colombianos realizando transacciones.
Desarrollado sobre un Framework/CMS popular.	Se refiere a sitios que están contruidos sobre frameworks/cms especializados en comercio electrónico, por ejemplo, si está hecho sobre OSCo-merce ( <a href="http://www.oscommerce.com/">http://www.oscommerce.com/</a> )	Si el sitio Web está construido sobre un framework/cms, su modelamiento será más sencillo, ya que estudiando el framework se entenderá cómo funciona el sitio.
Localizado al idioma español.	Se refiere a sitios de comercio electrónico en idioma español.	Permitirá filtrar sitios localizados a Colombia.
Popularidad del sitio según cámara Colombiana de comercio electrónico.	Se refiere al número de visitas en promedio (mensual) que el sitio haya recibido. La cantidad de visitas será tomada de los boletines publicados por la cámara colombiana de comercio electrónico.	Permitirá determinar, en términos generales, la popularidad del sitio.



**4.4.2 Método de selección de sitios: matriz de selección.** El método de matriz de selección permite aplicar criterios a un conjunto de opciones para seleccionar un subconjunto de estas de acuerdo a la experiencia del investigador . En la URL <http://www.idea.edu.pe/herramientas/6taherramientadecalidadmatrizdeseleccion.doc>, se encuentra un ejemplo de uso la matriz de la selección, que de forma general, funciona como sigue:

- Asignar una escala cualquiera (1 a 5, 1 a 10, etc) al conjunto de criterios a utilizar, esta escala solo indica la importancia de los criterios (10 debe ser más importante que 5).
- Se compara cada criterio en una matriz de pares y se calcula el factor de ponderación para cada uno de ellos (total de la suma del criterio sobre el total general).
- Para cada criterio se compara cada uno de los sitios con respecto a los demás y se calcula de nuevo un criterio de ponderación.
- Finalmente se multiplican los criterios de ponderación por criterio-alternativa y al final se seleccionan los 4 resultados más altos, en esta investigación se seleccionarán 4 sitios de comercio electrónico para la recolección de la información.

**4.4.3 Proceso de selección de sitios de comercio electrónico.** Los criterios de selección se enumerarán como sigue:

(1): Términos de uso.

(2): Disponibilidad de API.

(3): Presencia en Colombia.

(4): Desarrollado sobre un CMS/Framework.

(5): Localizado al español.

(6): Popular según la cámara Colombiana de comercio electrónico.

La escala utilizada fue la siguiente:

10: Muy importante.

5: Importante.

1: Indiferente.

1/5: No es importante.

1/10: Poco importante.

A continuación la matriz de comparación de los criterios junto a sus factores de ponderación:

**Tabla 6. Matriz de comparación de criterios**

	(1)	(2)	(3)	(4)	(5)	(6)	Suma	Factor
(1)	x	1	10	5	5	5	26	0.35
(2)	1	x	1	5	5	5	17	0.23
(3)	1/10	1	x	5	1	5	12.1	0.16
(4)	1/5	1/5	1/5	x	5	5	10.6	0.14
(5)	1/5	1/5	1	1/5	x	5	6.6	0.09
(6)	1/5	1/5	1/5	1/5	1/5	x	1.0	0.01
						Total:	73.3	

Según el método de matriz de selección los criterios con bajo factor de ponderación pueden ser despreciados, por lo tanto, los factores (5) y (6) no serán tenidos en cuenta a la hora de comparar los sitios por cada factor.

En la siguiente tabla se resume el estudio por sitio con respecto a cada uno de los 4 criterios (del 1 al 4) con factor de ponderación más altos:

**Tabla 7. Evaluación de criterios por sitio**

ml	<p>Términos de uso: Prohibido explícitamente Web scraping.</p> <p>API: Si.</p> <p>Presencia en Colombia: Si.</p> <p>CMS / Framework: No.</p>
mn**	<p>Términos de uso: No prohíbe explícitamente Web scraping.</p> <p>API: No.</p> <p>Presencia en Colombia: Si.</p> <p>CMS / Framework: No.</p>
olx*	<p>Términos de uso: No prohíbe explícitamente Web scraping.</p> <p>API: No.</p> <p>Presencia en Colombia: Si.</p> <p>CMS / Framework: No.</p>
ln	<p>Términos de uso: No prohíbe explícitamente Web scraping.</p> <p>API: No.</p> <p>Presencia en Colombia: Si.</p> <p>CMS / Framework: No.</p>
ebay	<p>Términos de uso: Prohibido explícitamente Web scraping.</p> <p>API: Si.</p> <p>Presencia en Colombia: Si.</p> <p>CMS / Framework: No.</p>
amz	<p>Términos de uso: No prohíbe explícitamente Web scraping.</p> <p>API: Si.</p> <p>Presencia en Colombia: Si.</p>

	CMS / Framework: No.
tmart	<p>Términos de uso: No prohíbe explícitamente Web scraping.</p> <p>API: No.</p> <p>Presencia en Colombia: Si.</p> <p>CMS / Framework: No.</p>
bg	<p>Términos de uso: Prohibido explícitamente Web scraping.</p> <p>API: No.</p> <p>Presencia en Colombia: Si.</p> <p>CMS / Framework: No.</p>
df	<p>Términos de uso: No prohíbe explícitamente Web scraping.</p> <p>API: No.</p> <p>Presencia en Colombia: Si.</p> <p>CMS / Framework: No.</p>
ex	<p>Términos de uso: No prohíbe explícitamente Web scraping.</p> <p>API: No.</p> <p>Presencia en Colombia: Si.</p> <p>CMS / Framework: No.</p>
fl	<p>Términos de uso: No prohíbe explícitamente Web scraping.</p> <p>API: No.</p> <p>Presencia en Colombia: Si.</p> <p>CMS / Framework: No.</p>
al	<p>Términos de uso: Prohibido explícitamente Web scraping.</p> <p>API: No.</p> <p>Presencia en Colombia: Si.</p> <p>CMS / Framework: No.</p>
alk	<p>Términos de uso: Prohibido explícitamente Web scraping.</p> <p>API: No.</p>

	Presencia en Colombia: Si. CMS / Framework: No.
hm	Términos de uso: No prohíbe explícitamente Web scraping. API: No. Presencia en Colombia: Si. CMS / Framework: No.

\* OLX no se tuvo en cuenta por considerarse un sitio de clasificados para la venta de productos usados.

\*\* Tiendas Monomi no se tuvo en cuenta ya que ofrece tiendas individuales enfocadas a pequeñas y medianas empresas, es decir, Tiendas Monomi no se comporta como una gran tienda de comercio electrónico sino como una plataforma que permite crear tiendas especializadas en un mercado específico.

Con los resultados del estudio, se aplicó la calificación dentro de la escala escogida a cada uno de los 14 sitios con respecto a cada factor. A continuación la tabla de ponderación de cada uno de estos:

**Tabla 8. Ponderación para el factor términos de uso**

	ml	mn	olx	ln	ebay	amz	tmart	bg	df	ex	fl	al	alk	hm	sum	ftor
ml	x	1/5	5	1/5	1	1/5	1/5	1	1/5	1/5	1/5	1	1	1/5	10.6	0.030
mn	5	x	5	1	5	1	1	5	1	1	1	5	5	1	37	0.105
olx	1/5	1/5	x	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	2.6	0.0074
ln	5	1	5	x	5	1	1	5	1	1	1	5	5	1	37	0.105
ebay	1	1/5	5	1/5	x	1/5	1/5	1	1/5	1/5	1/5	1	1	1/5	10.6	0.030
amz	5	1	5	1	5	x	1	5	1	1	1	5	5	1	37	0.105
tmart	5	1	5	1	5	1	x	5	1	1	1	5	5	1	37	0.105
bg	1	1/5	5	1/5	1	1/5	1/5	x	1/5	1/5	1/5	1	1	1/5	10.6	0.030
dg	5	1	5	1	5	1	1	5	x	1	1	5	5	1	37	0.105
ex	5	1	5	1	5	1	1	5	1	x	1	5	5	1	37	0.105
fl	5	1	5	1	5	1	1	5	1	1	x	5	5	1	37	0.105
al	1	1/5	5	1/5	1	1/5	1/5	1	1/5	1/5	1/5	x	1	1/5	10.6	0.030

alk	1	1/5	5	1/5	1	1/5	1/5	1	1/5	1/5	1/5	1	x	1/5	10.6	0.030
hm	5	1	5	1	5	1	1	5	1	1	1	5	5	x	37	0.105
															351.6	

En la tabla anterior se puede observar la evaluación de cada sitio con respecto al criterio *términos de uso*, los sitios con mejor puntaje en este criterio fueron: Monomi, Linio, Amazon, Tmart y todos aquellos con puntaje 0.105.

**Tabla 9. Ponderación para el factor disponibilidad de una API**

	ml	mn	olx	ln	ebay	amz	tmart	bg	df	ex	fl	al	alk	hm	sum	ftor
ml	x	5	5	5	1	1	5	5	5	5	5	5	5	5	57.0	0.18
mn	1/5	x	5	1	1/5	1/5	1	1	1	1	1	1	1	1	14.6	0.046
olx	1/5	1/5	x	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	2.6	0.0081
ln	1/5	1	5	x	1/5	1/5	1	1	1	1	1	1	1	1	14.4	0.046
ebay	1	5	5	5	x	1	5	5	5	5	5	5	5	5	57.0	0.18
amz	1	5	5	5	1	x	5	5	5	5	5	5	5	5	57.0	0.18
tmart	1/5	1	5	1	1/5	1/5	x	1	1	1	1	1	1	1	14.6	0.046
bg	1/5	1	5	1	1/5	1/5	1	x	1	1	1	1	1	1	14.6	0.046
df	1/5	1	5	1	1/5	1/5	1	1	x	1	1	1	1	1	14.6	0.046
ex	1/5	1	5	1	1/5	1/5	1	1	1	x	1	1	1	1	14.6	0.046
fl	1/5	1	5	1	1/5	1/5	1	1	1	1	x	1	1	1	14.6	0.046
al	1/5	1	5	1	1/5	1/5	1	1	1	1	1	x	1	1	14.6	0.046
alk	1/5	1	5	1	1/5	1/5	1	1	1	1	1	1	x	1	14.6	0.046
hm	1/5	1	5	1	1/5	1/5	1	1	1	1	1	1	1	x	14.6	0.046
															319.4	

En la tabla anterior se puede observar la evaluación de cada sitio con respecto al criterio *dis-*

*ponibilidad de una API*, los sitios con mejor puntaje en este criterio fueron: Mercado Libre, eBay y Amazon.

**Tabla 10. Ponderación para el factor presencia en Colombia**

	ml	mn	olx	ln	ebay	amz	tmart	bg	df	ex	fl	al	alk	hm	sum	ftor
ml	x	1	1	1	1	1	1	1	1	1	1	1	1	1	13	0.071
mn	1	x	1	1	1	1	1	1	1	1	1	1	1	1	13	0.071
olx	1	1	x	1	1	1	1	1	1	1	1	1	1	1	13	0.071
ln	1	1	1	x	1	1	1	1	1	1	1	1	1	1	13	0.071
ebay	1	1	1	1	x	1	1	1	1	1	1	1	1	1	13	0.071
amz	1	1	1	1	1	x	1	1	1	1	1	1	1	1	13	0.071
tmart	1	1	1	1	1	1	x	1	1	1	1	1	1	1	13	0.071
bg	1	1	1	1	1	1	1	x	1	1	1	1	1	1	13	0.071
df	1	1	1	1	1	1	1	1	x	1	1	1	1	1	13	0.071
ex	1	1	1	1	1	1	1	1	1	x	1	1	1	1	13	0.071
fl	1	1	1	1	1	1	1	1	1	1	x	1	1	1	13	0.071
al	1	1	1	1	1	1	1	1	1	1	1	x	1	1	13	0.071
alk	1	1	1	1	1	1	1	1	1	1	1	1	x	1	13	0.071
hm	1	1	1	1	1	1	1	1	1	1	1	1	1	x	13	0.071
															182	

En la tabla anterior se puede observar la evaluación de cada sitio con respecto al criterio *presencia en Colombia*, en este criterio todos los sitios obtuvieron igual puntaje.

**Tabla 11. Ponderación para el factor desarrollado sobre un CMS/Framework**

	ml	mn	olx	ln	ebay	amz	tmart	bg	df	ex	fl	al	alk	hm	sum	ftor
ml	x	1	1	1	1	1	1	1	1	1	1	1	1	1	13	0.071
mn	1	x	1	1	1	1	1	1	1	1	1	1	1	1	13	0.071
olx	1	1	x	1	1	1	1	1	1	1	1	1	1	1	13	0.071
ln	1	1	1	x	1	1	1	1	1	1	1	1	1	1	13	0.071
ebay	1	1	1	1	x	1	1	1	1	1	1	1	1	1	13	0.071
amz	1	1	1	1	1	x	1	1	1	1	1	1	1	1	13	0.071
tmart	1	1	1	1	1	1	x	1	1	1	1	1	1	1	13	0.071
bg	1	1	1	1	1	1	1	x	1	1	1	1	1	1	13	0.071
df	1	1	1	1	1	1	1	1	x	1	1	1	1	1	13	0.071
ex	1	1	1	1	1	1	1	1	1	x	1	1	1	1	13	0.071
fl	1	1	1	1	1	1	1	1	1	1	x	1	1	1	13	0.071
al	1	1	1	1	1	1	1	1	1	1	1	x	1	1	13	0.071
alk	1	1	1	1	1	1	1	1	1	1	1	1	x	1	13	0.071
hm	1	1	1	1	1	1	1	1	1	1	1	1	1	x	13	0.071
															182	

En la tabla anterior se puede observar la evaluación de cada sitio con respecto al criterio *desarrollado sobre un CMS/Framework*, en este criterio todos los sitios obtuvieron igual puntaje.



**Tabla 12. Ponderación general**

Sitio/Factor	(1)	(2)	(3)	(4)	
ml	0.35 * 0.030	0.23 * 0.18	0.16 * 0.071	0.14 * 0.071	0.0732 **
mn	0.35 * 0.105	0.23 * 0.046	0.16 * 0.071	0.14 * 0.071	0.0685
olx	0.35 * 0.0074	0.23 * 0.0081	0.16 * 0.071	0.14 * 0.071	0.0257
ln	0.35 * 0.105	0.23 * 0.046	0.16 * 0.071	0.14 * 0.071	0.0685 **
ebay	0.35 * 0.030	0.23 * 0.18	0.16 * 0.071	0.14 * 0.071	0.0732 **
amz	0.35 * 0.105	0.23 * 0.18	0.16 * 0.071	0.14 * 0.071	0.0994 **
tmart	0.35 * 0.105	0.23 * 0.046	0.16 * 0.071	0.14 * 0.071	0.0685
bg	0.35 * 0.030	0.23 * 0.046	0.16 * 0.071	0.14 * 0.071	0.0423
df	0.35 * 0.105	0.23 * 0.046	0.16 * 0.071	0.14 * 0.071	0.0685
ex	0.35 * 0.105	0.23 * 0.046	0.16 * 0.071	0.14 * 0.071	0.0685
fl	0.35 * 0.105	0.23 * 0.046	0.16 * 0.071	0.14 * 0.071	0.0685
al	0.35 * 0.030	0.23 * 0.046	0.16 * 0.071	0.14 * 0.071	0.0423
alk	0.35 * 0.030	0.23 * 0.046	0.16 * 0.071	0.14*0.071	0.0423
hm	0.35 * 0.105	0.23 * 0.046	0.16 * 0.071	0.14*0.071	0.0685

\*\* Sitios con la más alta ponderación.

De acuerdo a la tabla anterior, los sitios seleccionados para el estudio de API's y DOM son (puntuajes más altos): Amazon, Ebay, Mercado Libre y Linio Colombia. Vale la pena mencionar que para la selección del cuarto y último sitio, varios finalizaron con igual puntaje (0,0685), se seleccionó Linio Colombia por que aparece con mejor puntuación en el reporte de abril de 2014 de la cámara Colombiana de comercio electrónico.

#### 4.5 Técnicas de Recolección de Datos

Para la construcción del producto de investigación no fue necesario el uso de instrumentos como entrevistas o encuestas, ya que los datos a recolectar se encontraban directamente en los sitios de comercio electrónico seleccionados.

Las consultas hechas a la Cámara de Comercio de Bogotá y a la Cámara Colombiana de Comercio Electrónico se realizaron directamente desde sus sitios Web, desde esta última se obtuvo un listado de compañías o servicios activos asociados al e-Commerce para el primer semestre del año 2014.

Para la extracción de datos desde los sitios de comercio electrónico se utilizó Web Scraping (método de extracción Web desde el HTML publicado) y consumo de APIs, de acuerdo a la siguiente relación (obtenida a partir del estudio de cada sitio):

**Tabla 13. Técnica de extracción por sitio seleccionado**

Sitio	Técnica
eBay	Consumo de API
Mercado Libre	Consumo de API
Amazon	Consumo de API
Linio Colombia	Web Scraping

Se menciona en este punto que el manejo de la API para los sitios eBay, Mercado Libre y Amazon es totalmente distinto, a continuación de detallan algunas características:

eBay: No cuenta con una API REST, sin embargo es posible obtener respuestas en formato JSON, se deben contar con keys de desarrollo para poder acceder a ellas.

Mercado Libre: Cuenta con una API REST, su manejo es sencillo y no requiere de registro ni llaves para hacer uso de ella.

Amazon: No cuenta con una API REST, no es posible obtener respuestas en JSON (solo XML), su manejo no es trivial y la documentación es ciertamente confusa.

Esta heterogeneidad en el diseño de las APIs puede ser consecuencia del tiempo de operación de la plataforma, por ejemplo, al ser Mercado Libre una tienda más joven que Amazon y eBay, su equipo de ingenieros basó su desarrollo en la arquitectura REST con el objetivo de crear

desde cero API de fácil manejo y con las últimas tendencias de diseño. A diferencia de Mercado Libre, Amazon, es una tienda que lleva más de una década de operación, cuando Amazon nació, el estándar de diseño a seguir era SOAP, en efecto transformar su API (antigua y compleja, pero funcional) a REST es un trabajo de alta complejidad técnica y de alto costo financiero, tal vez por esto aún no la ha ejecutado. eBay está en el medio de ambas situaciones, es una tienda más antigua que Mercado Libre pero más joven que Amazon, y su API refleja perfectamente este estado; no es REST (es SOAP) pero si permite obtener respuestas en JSON. Steve Francia escribió un artículo interesante sobre las diferencias entre una API SOAP y una REST, el lector puede acceder a el en la URL <http://spf13.com/post/soap-vmes-rest>.

#### **4.6 Método de Proyección de Comportamiento: Mínimos Cuadrados**

**4.6.1 Criterios de selección de métodos estadísticos.** Las variables precio, cantidad vendida y cantidad existente tienen un comportamiento lineal con respecto al tiempo, el lector puede observar por ejemplo que para un rango de tiempo, el precio de un producto en una tienda electrónica no sufre variaciones bruscas. Por simple inspección se vislumbra un comportamiento tendiente a la linealidad (creciente o decreciente) en observaciones que involucran al tiempo (con respecto al precio o la cantidad de productos), por esta razón solo se tuvieron en cuenta métodos estadísticos para el ajuste de líneas (métodos para estimaciones lineales). Una línea permite proyectar de forma puntual un conjunto de lecturas bidimensionales, además de hacer fácil su manipulación, varias hojas electrónicas de cálculo como EXCEL tiende a utilizar la líneas para predecir el comportamiento de variables.

**4.6.2 Selección del método estadístico.** Para realizar la proyección del comportamiento de los productos se seleccionó el método de los mínimos cuadrados (ver marco sección 2.2.9 del marco teórico) porque proporciona intervalos pequeños de error, es reproducible y permite obte-

ner la ecuación de una recta a partir de datos experimentales.

#### **4.7 Fases desarrolladas para la construcción del producto de software**

La secuencia de fases ejecutadas en esta investigación, partiendo desde la implementación (trabajo de campo) fueron:

Selección de sitios Web.

- Definición de criterios para selección de Sitios Web.
- Indagación digital sitios populares eCommerce.
- Consulta Cámara de comercio de Bogotá.
- Visita de sitios Webs y resumen General.
- Selección de sitios e Informe de avance.

Análisis DOM y APIs.

- Consulta documentación de APIs sitios seleccionados.
- Evaluación de utilización de APIs por sitio.
- Evaluación de utilización Scraping por sitio.
- Esquema funcionamiento sitios con API.
- Esquema funcionamiento sitios con Scraping.

Definición y selección de estrategias de extracción de datos.

- Definir estrategia de extracción por cada sitio.
- Documentar estrategia de extracción de datos por cada sitio.

Modelo de datos e implementación del repositorio.

- Definición del modelo de datos.
- Implementación del modelo de datos inicial del repositorio.

e) Implementación del crawler.

- Iteración de desarrollo 1.
- Iteración de desarrollo 2.
- Iteración de desarrollo 3.
- Iteración de desarrollo 4.
- Implementación del repositorio de datos.

f) Selección de métodos estadísticos.

- Definición de criterios para selección de métodos estadísticos.
- Consulta de métodos estadísticos.
- Selección de métodos estadísticos.
- Implementación de métodos estadísticos.

Implementación de la Interfaz Web.

- Iteración de desarrollo 1.
- Iteración de desarrollo 2.
- Iteración de desarrollo 3.
- Iteración de desarrollo 4.

#### 4.8 Algunos Aspectos de Diseño

Para la implementación de los scripts de extracción de datos (crawler) se aplicó el enfoque de tareas: un usuario administrador programa manualmente una tarea de extracción sobre una categoría específica de algún sitio registrado, esto significa que el crawler no extraerá datos de forma permanente, sino dentro de un periodo de tiempo fijado en el momento de registrar la tarea.

Lo anterior debido a que:

- Las plataformas pueden penalizar las fuentes desde donde se realizan lecturas si se hace uso excesivo de llamadas a su API, otras pueden exigir un pago al pasar un determinado número de solicitudes.
- El tiempo de escaneo puede crecer considerablemente, afectando el rendimiento del script.
- El servidor dónde se aloja la aplicación es de uso personal y no cuenta con los requerimientos técnicos para almacenar aplicaciones de software con alto uso de transferencia de datos.

Por otra parte, después de revisar los valores devueltos por las APIs se determinó que los campos precio, cantidad vendida y cantidad disponible son los más relevantes a la hora de hacer seguimiento sobre un producto.

Para la implementación de la aplicación se utilizaron las siguientes herramientas:

**Tabla 14. Herramientas de desarrollo utilizadas en la implementación de la aplicación**

Componente	Lenguaje/Tecnología	Librería / Descripción
Script de extracción de datos desde el HTML.	Python	Scrapy. <a href="http://scrapy.org/">http://scrapy.org/</a>
Script de extracción de datos vía API para eBay.	Python	S D K e B a y , <a href="https://github.com/timotheus/ebaysdk-python">https://github.com/timotheus/ebaysdk-python</a>
Script de extracción de datos vía API para Mercado Libre.	Python	SDK Mercado Libre, <a href="https://github.com/mercadolibre/python-sdk">https://github.com/mercadolibre/python-sdk</a>

Script de extracción de datos vía API para Amazon.	Python	Implementación de un tercero, <a href="https://github.com/yoavaviram/python-amazon-simple-product-api">https://github.com/yoavaviram/python-amazon-simple-product-api</a>
Base de datos relacional.	PostgreSQL	<a href="http://www.postgresql.org/">http://www.postgresql.org/</a>
Base de datos no relacional.	MongoDB	<a href="http://www.mongodb.org/">http://www.mongodb.org/</a>
App de visualización de datos.	Python	Django, <a href="https://djangoproject.com/">https://djangoproject.com/</a>
App de administración de entidades.	Python	Django, <a href="https://djangoproject.com/">https://djangoproject.com/</a>
Script para la proyección del comportamiento a través del método de los mínimos cuadrados.	Python	Numpy: <a href="http://www.numpy.org/">http://www.numpy.org/</a>

#### 4.9 Plan de Presentación de Resultados

Los entregables obtenidos en esta investigación serán detallados en el próximo capítulo, se comenzará definiendo el esquema básico de funcionamiento de cada sitio seleccionado de acuerdo a la técnica de extracción escogida, después se agregará el modelo final del repositorio de datos para culminar con detalles de diseño tanto de los scripts de recolección de datos como de proyección de comportamiento, además de la interfaz Web de visualización de resultados.

## 5. Resultados de la Investigación

Este capítulo detalla los entregables obtenidos como resultado de la ejecución de cada una de las fases planeadas para la construcción del producto. Inicialmente se describe el esquema básico de funcionamiento de los sitios seleccionados de acuerdo a la estrategia de extracción de datos aplicada a cada uno de estos, después se define el modelo del repositorio de datos y finalmente se mencionan aspectos de desarrollo de la aplicación.

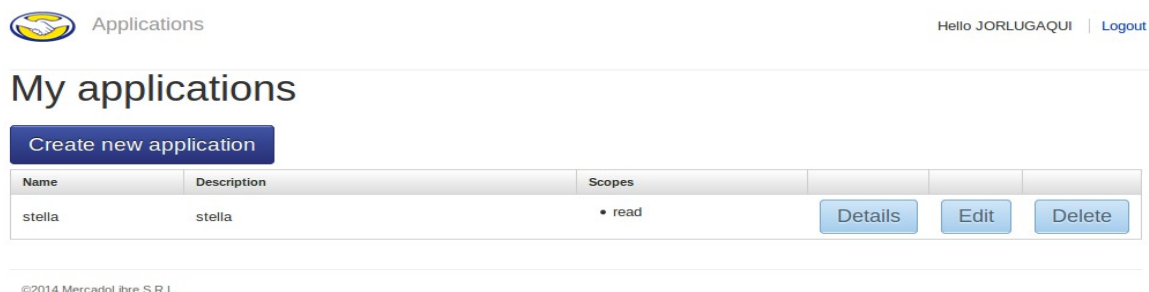
### 5.1 Análisis DOM y API de los Sitios Seleccionados

**5.1.1 Mercado libre.** Mercado libre cuenta con una API REST (<http://www.restapitutorial.com/>) para desarrolladores, para esta plataforma de comercio electrónico no se hizo necesario el uso de la técnica Web Scraping, la cual está prohibida en sus términos de uso (<http://developers.mercadolibre.com/terms-and-conditions/>).

Para interactuar con recursos privados en la API es requerido registrar una aplicación en el “Applications Manager” del sitio de desarrolladores, este panel de administración se encuentra en la URL <https://applications.mercadolibre.com.co/>, una vez se registre la app (para este ejemplo el nombre es stella) la plataforma asigna un `secret_key` y un `app_id`, datos que siempre se deben enviar en las peticiones que intenten acceder a recursos privados (datos de vendedores, compradores, etc).

Sin embargo la mayoría de consultas que esta investigación requirió son públicas y se pueden hacer directamente sobre el navegador, línea de comandos, SDK, etc. sin necesidad de tener registrada una app. El Applications manager luce como sigue:



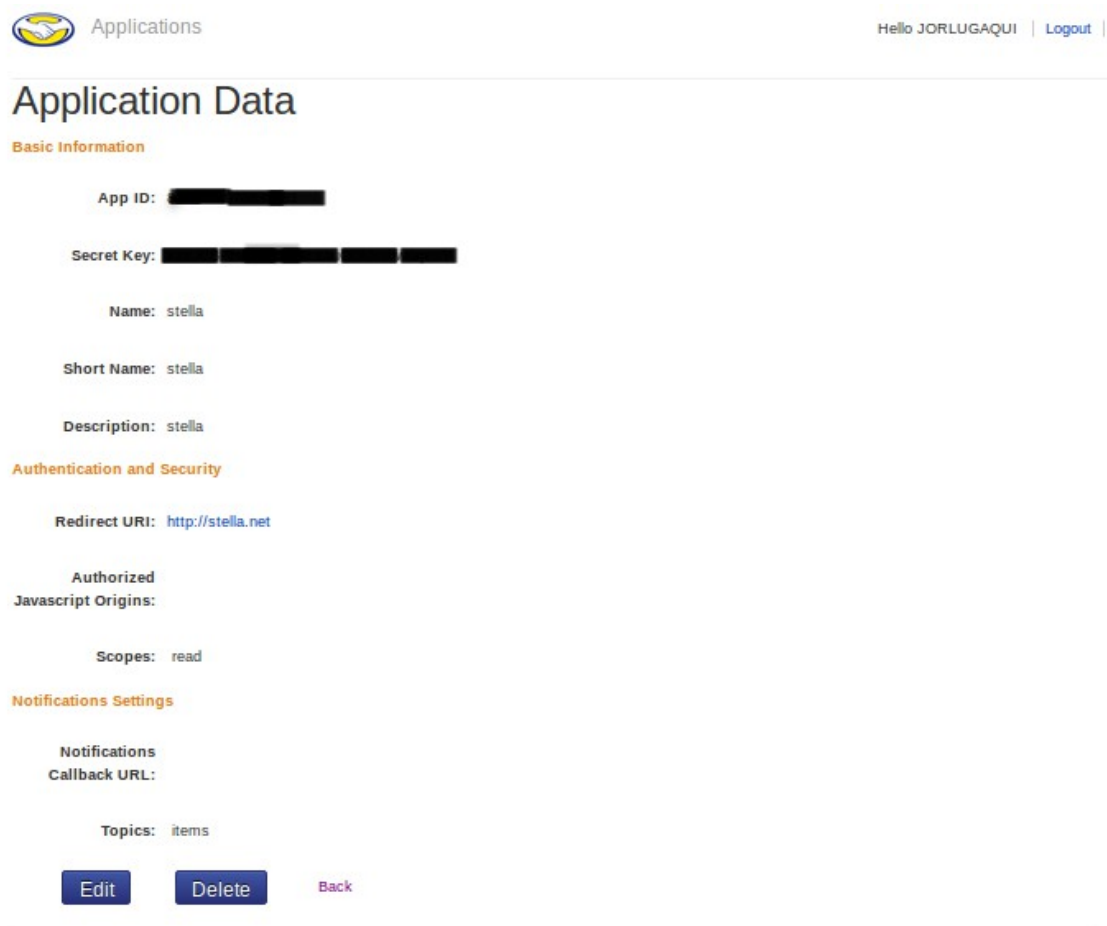


The screenshot shows the 'My applications' page in the Applications Manager. At the top left is the 'Applications' logo. At the top right, it says 'Hello JORLUGAQUI | Logout'. Below the header is a 'Create new application' button. A table lists the applications:

Name	Description	Scopes			
stella	stella	• read	Details	Edit	Delete

At the bottom left, there is a copyright notice: ©2014 MercadoLibre S.R.L.

**Figura 9. Interfaz principal del Applications Manager de Mercado Libre**



The screenshot shows the 'Application Data' page for the 'stella' application. At the top left is the 'Applications' logo. At the top right, it says 'Hello JORLUGAQUI | Logout'. Below the header is the title 'Application Data'. The page is divided into sections:

- Basic Information:**
  - App ID: [Redacted]
  - Secret Key: [Redacted]
  - Name: stella
  - Short Name: stella
  - Description: stella
- Authentication and Security:**
  - Redirect URI: <http://stella.net>
  - Authorized Javascript Origins:
  - Scopes: read
- Notifications Settings:**
  - Notifications Callback URL:
  - Topics: items

At the bottom, there are three buttons: 'Edit', 'Delete', and 'Back'.

**Figura 10. Detalle de la aplicación registrada**

La URL en la que la API responde es <https://api.mercadolibre.com>, este servicio Web espera peticiones a recursos públicos o privados con una estructura predefinida, a continuación algunos ejemplos:

```
$ curl https://api.mercadolibre.com/sites/MLA/search?q=ipod
```

```
$ curl https://api.mercadolibre.com/users/me?access_token=...
```

Como se puede observar, después del dominio principal es necesario colocar el nombre del recurso al que se requiere acceder (sites, users, etc.), después del recurso aparece la región (MLA), y finalmente la acción específica sobre el recurso (search?q=ipod, me?access\_token=...).

En esta investigación se realizaron peticiones sobre el recurso producto (búsqueda de productos por categoría o por texto), cuando un producto es encontrado en la base de datos, la API de Mercado Libre devuelve un documento JSON (ver anexo D). En este documento se puede apreciar datos asociados al vendedor, al producto, etc., dentro de los datos publicados por la API se encuentran tres de gran importancia para el trabajo a desarrollar, estos son: price (precio), available\_quantity (cantidad disponible), sold\_quantity (cantidad vendida).

Cuando desde la API se realiza una petición de consulta a la plataforma de Mercado Libre, el JSON obtenido tiene el siguiente formato:

```
{
  "site_id": "MLA",
  "paging": {},
  "results": [
  ],
  "secondary_results": [
  ],
  "related_results": [
  ],
  "sort": {},
  "available_sorts": [],
  "filters": [
  ],
  "available_filters": [
  ]
}
```

Dentro de la matriz de resultados (campo results) se encuentran los productos que coinciden con el criterio de búsqueda, cada uno de estos productos presenta el formato descrito en el JSON del párrafo anterior.

A continuación el esquema general de consultas para la API de Mercado Libre:

**Tabla 15. Esquema de consulta a la API de Mercado Libre**

Acción	URL
Realizar una búsqueda por consulta.	GET /sites/MLA/search?q=
Realizar una búsqueda por categoría.	GET /sites/MLA/search?category=
Realizar una búsqueda por el Id del vendedor.	GET /sites/MLA/search?seller_id=
Realizar una búsqueda por nombre de usuario.	GET /sites/MLA/search?nickname=
Realizar una búsqueda aplicando filtros.	GET /sites/MLA/search?special_filter=
Realizar una búsqueda con ordenamiento.	GET /sites/MLA/search?q=ipod&sort=sortId=

Algunos ejemplos para realizar búsquedas desde la API son:

Ejemplo de consulta por texto:

```
curl https://api.mercadolibre.com/sites/MLA/search?q=ipod nano
```

Ejemplo de consulta de categoría por país:

```
curl https://api.mercadolibre.com/sites/MLA/categories
```

Ejemplo de consulta subcategoría por país:

```
curl https://api.mercadolibre.com/categories/MLA1071
```

Ejemplo consulta de productos por categoría:

```
curl https://api.mercadolibre.com/sites/MLA/search?category=MLA3530
```

Ejemplo utilización de filtros:

```
curl https://api.mercadolibre.com/sites/MLA/search?q=ipod&price=700.0-1000.0
```

Todas las respuestas de la API están en formato JSON con codificación UTF-8 (<http://tools.ietf.org/html/rfc3629>). Finalmente, Mercado libre ofrece un SDK para el lenguaje Python (<https://github.com/mercadolibre/python-sdk>), este permite realizar consultas directamente a la plataforma, facilitando al programador operaciones como autenticación y tareas de transformación de datos. Para efectos de este proyecto, todas las consultas se deben hacer contra el sitio

Mercado Libre Colombia (MCO).

**5.1.2 eBay.** eBay cuenta con APIs segmentadas, estas permiten la consulta de datos directamente desde sus bases de datos, por lo cual, la inspección vía Web Scrapping no es necesaria, además esta práctica está prohibida en sus términos de uso (<http://developer.ebay.com/join/licenses/individual/api.pdf>).

Algunas de las APIs mencionadas son:

**Tabla 16. Listado parcial de APIs de eBay**

API	Documentación
Shopping API	<a href="http://go.developer.ebay.com/developers/ebay/documentation-tools#shoppingapi">http://go.developer.ebay.com/developers/ebay/documentation-tools#shoppingapi</a>
Trading API	<a href="http://go.developer.ebay.com/developers/ebay/documentation-tools#tradingapi">http://go.developer.ebay.com/developers/ebay/documentation-tools#tradingapi</a>
Finding API	<a href="http://go.developer.ebay.com/developers/ebay/documentation-tools#findingapi">http://go.developer.ebay.com/developers/ebay/documentation-tools#findingapi</a>
Feedback API	<a href="http://go.developer.ebay.com/developers/ebay/documentation-tools#feedback">http://go.developer.ebay.com/developers/ebay/documentation-tools#feedback</a>

Para consultar todas las APIs el lector se puede dirigir al centro de documentación para desarrolladores que se encuentra en la URL: <http://go.developer.ebay.com/developers/ebay/documentation-tools>

La URL de servicio (Service Endpoints) también está dividida según el entorno en el que se quiera ejecutar la petición y según el tipo de API. Los entornos que ofrece la plataforma para el caso del servicio de búsqueda (Finding) son:

- Producción: <http://svcs.ebay.com/services/search/FindingService/v1>
- Pruebas (Sandbox): <http://svcs.sandbox.ebay.com/services/search/FindingService/v1>

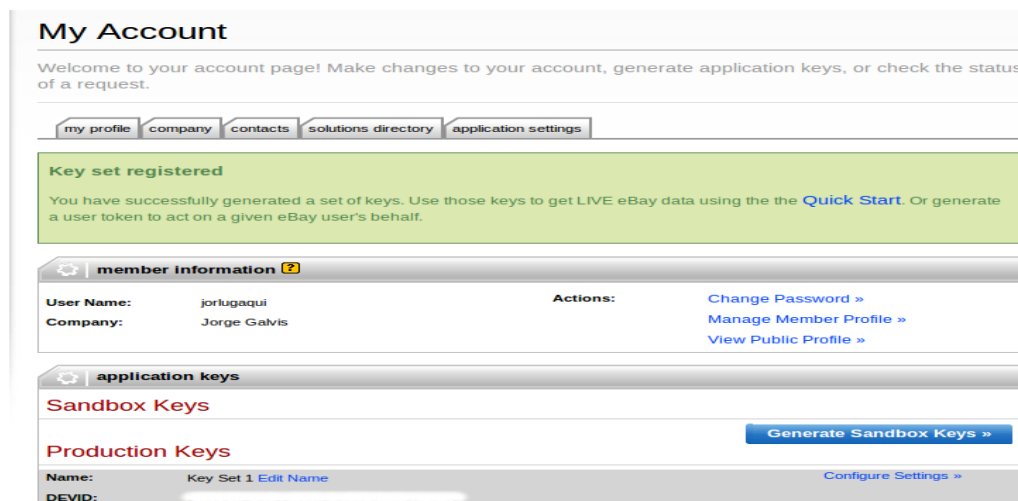
Para efectos de esta investigación se utilizó el entorno de producción, que refleja datos más actualizados de la plataforma. Para la consulta de productos en eBay se usó el siguiente procedi-

miento:

- Consultar la API de búsqueda (Finding API) para localizar productos por categoría.
- Con el ID del producto se usará la API de compra (Shopping API) que permite recuperar más datos asociados al producto (precio, cantidad vendida, cantidad en inventario, etc).
- De los datos obtenidos por la API de compra, registrar la cantidad disponible, la cantidad vendida y el precio.

La estructura JSON devuelta por la API de búsqueda para cada producto se puede observar en el anexo E.

Para realizar cualquier llamado (recursos públicos o privados) se debe enviar un APP ID que lo provee el centro para desarrolladores (<https://developer.ebay.com/join/default.aspx>) una vez el usuario programador se registre, para el caso del autor se tiene:



**Figura 11. Dashboard principal cuenta de desarrollador en eBay**

A continuación ejemplos de llamada a las APIs de compra y consulta que se utilizarán en esta investigación.

Ejemplo consulta de productos por categoría:

`http://svcs.ebay.com/services/search/FindingService/v1?OPERATION-NAME=findItemsByCategory&SERVICE-VERSION=1.0.0&SECURITY-APPNAME=AppId&RESPONSE-DATA-`

FORMAT=JSON&REST-PAYLOAD&categoryId=10181&paginationInput.entriesPerPage=2

Donde:

**Tabla 17. Detalle URL de servicio para la API de búsqueda de eBay**

Sección	Significado
<code>http://svcs.ebay.com/services/search/FindindingService/v1</code>	Es el EndPoint de la API de búsqueda.
<code>OPERATION-NAME</code>	Es la operación a ejecutar dentro de la llamada.
<code>SERVICE-VERSION=1.0.0</code>	Es la versión del servicio.
<code>SECURITY-APPNAME</code>	Es el App Id generado desde la cuenta de desarrollador.
<code>RESPONSE-DATA-FORMAT</code>	Es el formato con el que la API contestará la petición, en el ejemplo se usa JSON.
<code>REST-PAYLOAD</code>	Separador de parámetros de petición y datos de búsqueda.
<code>categoryId</code>	El Id de la categoría.
<code>paginationInput.entriesPerPage</code>	Cantidad de páginas en el resultado.

Ejemplo de consulta de producto por Id con detalles:

`http://open.api.ebay.com/shopping?`

`callname=GetSingleItem&responseencoding=JSON&appid=YourAppId&siteid=0&version=515&ItemID=291305408347&IncludeSelector=Details`

Donde:

**Tabla 18. Detalle URL de servicio para la API de compra de eBay**

Sección	Significado
<code>http://open.api.ebay.com/shopping?</code>	Es el EndPoint de la API de compra.
<code>callname</code>	Es la operación a ejecutar dentro de la llamada.
<code>responseencoding</code>	Es el formato con el que la API contestará la petición, en el ejemplo se usa JSON.
<code>appid</code>	Es el App Id generado desde la cuenta de desarrollador.
<code>siteid</code>	Es el sitio contra el cual se realiza la petición, 0 en esta petición significa Estados Unidos.

<code>version</code>	La versión del servicio.
<code>itemID</code>	El Id del producto.
<code>IncludeSelector=Details</code>	Significa mostrar más campos en la respuesta de la petición.

En este proyecto, se consultaron productos que tengan cobertura de entrega (shipping) en Colombia, con prioridad sobre el sitio de Estados Unidos (siteid=0). Finalmente, eBay desarrolló un SDK para Python que permite interactuar directamente con las diferentes APIs, este SDK está disponible en la URL <https://github.com/timotheus/ebaysdk-python>.

**5.1.3 Amazon.** Al igual que eBay, Amazon cuenta con un conjunto de APIs segmentadas por función u objetivo, ya que se dispone de una API para la consulta de productos no se realizó Web Scraping en esta plataforma, sin embargo, en la licencia de uso general del sitio no se encuentra explícitamente prohibido el uso de scripts para la extracción de datos (<http://www.amazon.com/gp/help/customer/display.html/104-5054883-7838319?ie=UTF8&Version=1&nodeId=508088&entries=0>). Algunas de las APIs de Amazon se encuentran en la URL <https://developer.amazon.com/public/apis>. Para consultar la información de productos se utilizó la `Product Advertising API` (<http://docs.aws.amazon.com/AWSECommerceService/latest/DG/Welcome.html>), que permite obtener datos como precio, nombre, entre otros. Para poder hacer uso de este servicio Web el usuario desarrollador debe cumplir los siguientes pasos:

Registrarse como `Product Advertising API Developer`, para obtener el `secret key` y el `access key`, necesarios para realizar peticiones que requieren autenticación.

The screenshot shows a three-step registration process: 1. SIGN IN, 2. ACCOUNT INFO (highlighted), and 3. SUCCESS. The user is identified as Jorge Galvis. The 'Contact Information' section includes fields for 'Full Name' (Jorge Luis Galvis Quintero) and 'Address'. A required field for '\* Website or Application Description' is present, with a note to describe the application. Below this is the 'Amazon.com Product Advertising API License Agreement' section, which includes a 'Printer Friendly Version' link and a note that the agreement was last updated on 2012-11-21.

**Figura 12. Proceso de registro en el Product Adversiting API Developer program**

Registrarse como Associate User, para obtener el tracking id, necesario en las peticiones que requieren autenticación.

The screenshot shows the 'Your Website Profile' step of the registration process. It includes a progress bar with four steps: 1. Your Account Information, 2. Your Website Profile (highlighted), 3. Your Identity Verification, and 4. Start Using Associates Central. The page title is 'Your Website Profile' with a help icon. A note states '\* All fields required unless otherwise noted'. The 'Website Subject & Content' section contains three main questions: 'What is the name of your website?' (with 'stella' entered), 'What is the URL of the website(s) you will use to send traffic to Amazon?' (with 'stella.net' entered and a link to add more URLs), and 'What is your website(s) about?' (with an empty text area). A note at the bottom explains that the name is used for the Associates ID and that the URL should be the top 5 by traffic volume.

**Figura 13. Registro como Associate User**



Aceptar los términos de uso de la API (<https://affiliate-program.amazon.com/gp/advertising/api/detail/agreement.html>).

El Endpoint de la API se encuentra en la URL: <http://webservices.amazon.com/onca/xml>, todas las respuestas del servicio son devueltas en formato XML (XML 2015). Para extraer los datos desde la plataforma de Amazon se realizó el siguiente procedimiento:

Se consultó una lista de productos a un nodo específico a través del operación ItemSearch de la API.

Con cada ítem seleccionado se consultó su detalle a través de la operación ItemLookup. No se encontró el mecanismo para obtener la cantidad de ítems vendidos, o la cantidad de ítems disponibles. Después de una rigurosa consulta a los servicios web ofrecidos y a los datos publicados en su sitio Web, el autor concluye que es altamente probable que la plataforma no publique dichos datos por estrategia comercial. En efecto en Amazon solo se monitorearon los precios de los productos.

El siguiente es el esquema general que se debe utilizar para hacer una petición relacionada a la obtención de información de un producto:

`http://webservices.amazon.com/onca/xml?`

```
Service=AWSECommerceService&
AWSAccessKeyId=[AWS Access Key ID]&
Operation=ItemLookup&
ItemId=B000080E6I
&Timestamp=[YYYY-MM-DDThh:mm:ssZ]
  &Signature=[Request Signature]
```

Donde:

**Tabla 19. Detalle URL de servicio para la API de consulta de Amazon**

Sección	Significado
<code>http://webservices.amazon.com/onca/xml</code>	Es el Endpoint del servicio.
<code>Service</code>	Es el nombre del servicio.

AWSAccessKeyId	Es el hash generado por Producto Advertising API Developer Site.
Operation	Es la operación que se quiere ejecutar.
ItemId	Es el id del producto.
Timestamp	Es la fecha y hora en la que se ejecuta la consulta.
Signature	Es un hash que encapsula el método de autenticación.

Como ya se mencionó, todas las respuestas del servicio Web son devueltas en XML, en el anexo F se puede apreciar el esquema general que utiliza la plataforma. En el documento XML habrán tanto nodos como campos encontrados en la base de datos, estos dependen del índice sobre el que se consulte, la operación que se quiera ejecutar y los grupos de respuesta que se utilicen. A continuación ejemplos de llamada a la API.

Ejemplo de consulta de un producto:

```
http://webservices.amazon.com/onca/xml?
Service=AWSECommerceService&Operation=ItemLookup&SubscriptionId=[ ]&Associate-
Tag=[ ]&Version=2011-08-01&ItemId=B000080E6I&IdType=ASIN&Condition=All&Respon-
seGroup=ItemAttributes
```

En este ejemplo se consultó el ítem con id B000080E6I con la operación ItemLookup sobre todos los nodos (grandes categorías) de Amazon, además en la consulta se especificó que los datos devueltos deben pertenecer al grupo de ItemAttributes. El XML devuelto contiene nodos de interés como precios y título del producto:

```
...
<ListPrice>
  <Amount>44999</Amount>
  <CurrencyCode>USD</CurrencyCode>
  <FormattedPrice>$449.99</FormattedPrice>
</ListPrice>
...
<Title>Canon PowerShot S400 4MP Digital Camera w/ 3x Optical Zoom</Title>
...
<Price>
  <Amount>7499</Amount>
  <CurrencyCode>USD</CurrencyCode>
  <FormattedPrice>$74.99</FormattedPrice>
```

```
</Price>
...
```

Ejemplo de consulta de productos por segmentos:

```
http://webservices.amazon.com/onca/xml?
Service=AWSECommerceService&Operation=ItemSearch&SubscriptionId=[ ]&Associate-
Tag=[ ]&Version=2011-08-01&SearchIndex=Electronics&Condition=New&Keywords=asi-
mov&ResponseGroup=ItemAttributes
```

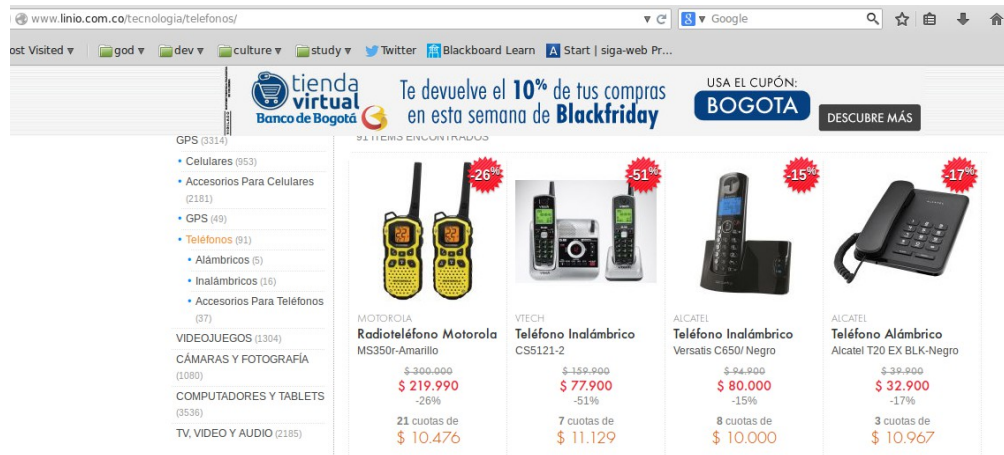
En esta consulta se utilizó la operación “ItemSearch” para buscar productos nuevos sobre el nodo “Electronics” con palabras claves “asimov”. Dentro del XML devuelto se encuentra el dato que interesa recuperar, el Id del producto:

```
...
<Item>
  <ASIN>B002C54F0M</ASIN>
</Item>
...
```

No existe un SDK oficial para el lenguaje Python en Amazon, sin embargo el usuario yoavaviram ha publicado en github una implementación que encapsula las solicitudes a la Product Advertising API desde este lenguaje, la librería se puede encontrar en: <https://github.com/yoavaviram/python-amazon-simple-product-api>.

**5.1.4 Linio Colombia.** En la plataforma Linio se utilizó Web Scraping para extraer datos, lo anterior por dos razones: La primera es que no cuenta con una API pública para desarrolladores, la segunda es que sus términos de uso no prohíben explícitamente esta práctica.

En Linio Colombia, las categorías pueden ser extraídas desde la URL, ya que presentan el siguiente patrón: <http://www.linio.com.co/categoría/subcategoría1/subcategoría2/subcategoríaN/>. Esto significa por ejemplo que, en la URL <http://www.linio.com.co/hogar/camas/> hogar es la categoría principal y camas es una subcategoría de esta.



**Figura 14. Categoría tecnología/teléfonos en Linio Colombia**

Algunos ejemplos de categorías en Linio Colombia son:

<http://www.linio.com.co/tecnologia/televisores/> categoría tecnología, subcategoría televisores.

<http://www.linio.com.co/tecnologia/televisores/ao/> categoría tecnología, subcategoría televisores, marca AO.

En el sitio Web de Linio todo el bloque de categorías está encerrado en el tag HTML `<ul class="nav__menu">`, con esta premisa, si se hace un escaneo al contenido de dicha etiqueta se pueden obtener todas las categorías activas en la plataforma. La estructura general del DOM que abarca el menú de categorías se puede consultar en el anexo G. Al hacer clic sobre una categoría se despliega la siguiente estructura DOM (se repite por producto):

```
<ul class="product-items" id="catalog-items">
<li id="id-item" class="product-item" itemscope="" itemtype="http://schema.org/Product" data-bad
badge="badgeDivFalse">
<a id="order:id-item" href="url-to-product" title="product-title">
<ul>
<li class="product-item-img" style="position: relative;">
<span class="lazyImage loaded">
... # HTML que se puede despreciar en el escaneo.
</span>
```

```

</li>
<li class="product-item-brand">Brand Product</li>
<li class="product-item-name"><em itemprop="name">Product Title</em></li>
<li class="product-item-price">
<span class="product-itm-price-old">$ 149.900</span>
<span class="product-itm-price-new" itemprop="price">$ 89.900</span>
<span class="product-itm-discount-percent">-40<sup>%</sup></span>
... # HTML que se puede despreciar en el escaneo.
</li>
</ul>
</a>
</li>
... # HTML que se puede despreciar en el escaneo.
</ul>

```

Si se escanea el DOM a partir del tag `<ul class="product-items" id="catalog-items">` se podrán obtener todos los productos asociados a la categoría seleccionada, y dentro de cada ítem listado se podrá acceder a los posibles datos a monitorear.

El modelo de publicación de Linio Colombia está basado en los esquemas que se encuentran en la URL <http://schema.org/Product>. En Linio solo se monitorizó el dato precio, ya que es el más adecuado para aplicarle métodos de predicción. A diferencia de las API's de Mercado Libre, o eBay, Linio Colombia en su HTML de publicación no cuenta con datos como cantidad en almacén y cantidad vendida. La URL para el detalle de un producto presenta un formato distinto a la de categorías, el esquema es el siguiente:

<http://www.linio.com.co/slug-del-producto>.



**Figura 15. Producto específico en Linio Colombia**

Se compone del dominio principal más una cadena de texto válida para direcciones Web (slug), algunos ejemplos de productos en Linio Colombia son:

- <http://www.linio.com.co/Televisor-Ao-39-Pulgadas-Led-AOLED39-Negro-331662.html>
- <http://www.linio.com.co/Licuadora-12-Velocidades-Hamilton-Beach---Negro-133408.html>
- <http://www.linio.com.co/Moto-Electrica-Gogreen-Hormiga-Plus---Azul-153563.html>

Todas las páginas a escanear están dentro del contexto de Linio Colombia, esto es, al extraer datos del dominio [linio.com.co](http://www.linio.com.co) se están obteniendo datos localizados para Colombia.

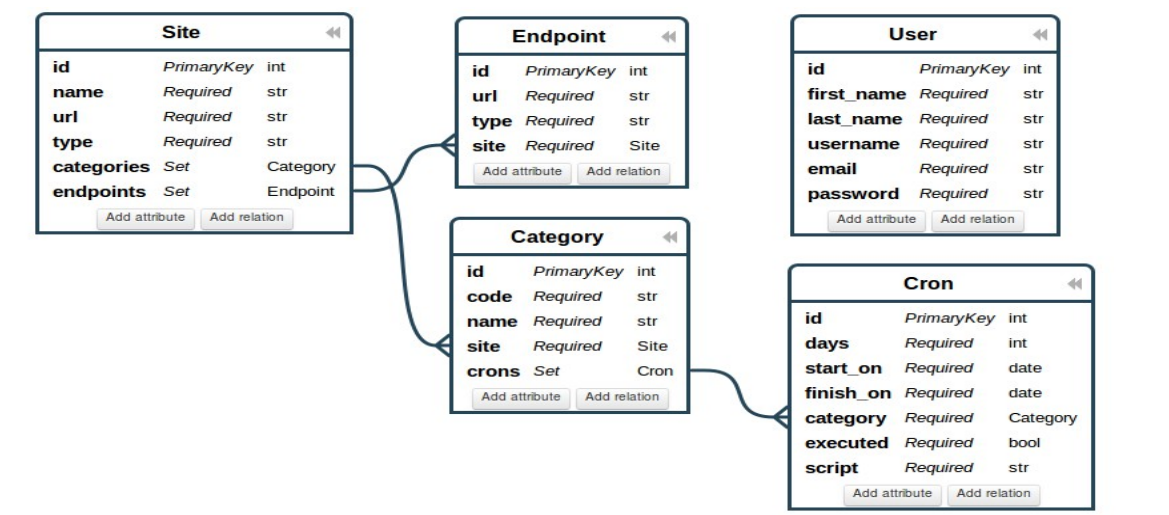
## 5.2 Modelo del Repositorio de Datos

Para almacenar los datos a extraer de las plataformas de comercio electrónico seleccionadas se hace uso de dos bases de datos, una de ellas relacional (PostgreSQL) y la otra no relacional (MongoDB).

En la base de datos relacional se almacena la información básica de los sitios, junto a las categorías que ofrece, además de las tareas programadas de extracción (crons) y la información bá-

sica de los productos.

Estas entidades (Sitios, Categorías, Tareas y Productos) son estáticas en el sentido de que su estructura no cambiará, además de que no estarán involucradas en procesos intensivos de escritura/lectura, por esto una base de datos relacional como PostgreSQL es recomendada para modelar esta sección del repositorio. A continuación el modelo de datos del componente relacional del repositorio:



**Figura 16. Modelo relacional del repositorio para las entidades sitios, categorías, tareas y productos**

En la base de datos no relacional se almacenan las lecturas diarias de los productos asociados en una categoría, en otras palabras, se almacenan las lecturas de las tareas registradas en la base de datos relacional, esta base de datos nos permite tener documentos con estructura flexible, lo que la hace muy conveniente, debido a que los datos a extraer no son los mismos en cada sitio de comercio electrónico seleccionado. A continuación el esquema de los documentos JSON que se tiene en MongoDB:

```

{
  "category_id":23,
  "category_name":"Electrónica",
  "cron_id":12,
  "site_id":3,
  "site_name":"Mercado Libre",
  "product_id":"MB983",
  "product_name":"Cámara Fotográfica",
  "history": [
    {
      "date":"2014-07-01",
      "sold_quantity":12,
      "available_quantity":2,
      "price":23.98
    }
  ],
  "future": [
    {
      "date":"2014-09-01",
      "sold_quantity":"X",
      "available_quantity":"X",
      "price":"X"
    }
  ]
}

```

**Figura 17. Esquema JSON para el almacenamiento de lecturas**

Como se puede apreciar, hay un conjunto de campos con información básica (categoría, sitio, producto) y un campo “history” con la información histórica de los productos escaneados. Sobre este campo se aplicó el método estadístico de los mínimos cuadrados para determinar el comportamiento de cada producto en el futuro por el mismo periodo de tiempo en el que se realizaron las lecturas. La velocidad de lectura sobre la base de datos no relacional es alta en comparación con la relacional, esto ayuda a que los reportes gráficos de los productos sean más rápidos de generar. Cada sitio de comercio electrónico agregado a la aplicación cuenta con su propia colección de almacenamiento de lecturas (como una tabla si se compara con la base de datos relacional), esto permite que los reportes sean más fáciles de implementar ya que los datos por sitio estarán separados.

Las categorías de los sitios seleccionados se cargaron desde Scripts Python, estas pueden ser



extraídas desde la API de cada plataforma o desde la exploración manual del sitio. Las tareas programadas de extracción se registran a través del administrador de la aplicación.

En el anexo H se encuentra el SQL necesario para construir el modelo relacional, y en el anexo I, un ejemplo de instrucción JavaScript que se debe ejecutar en el cliente de MongoDB para crear un documento JSON (componente no relacional).

### **5.3 Aspectos de Diseño de la Aplicación**

**5.3.1 Aspectos visuales de la interfaz Web de consulta.** Para la implementación de la interfaz de consulta se utilizó el framework bootstrap (<http://getbootstrap.com/>), ya que permite construir aplicaciones de forma más rápida, evitando invertir tiempo excesivo en la capa de presentación. Se creó un esquema sencillo y básico con respecto a las tendencias de diseño actual, es un esquema plano y bicolor (blanco y negro) pero funcional, que evita confusiones a los usuarios, además es un diseño que permite a los usuarios concentrarse en los datos y no en los detalles visuales de la aplicación, al ser una interfaz Web básica, está abierta a nuevas implementaciones.

A continuación la distribución de elementos de la pantalla de categorías:

Inicio	Sitios	Categorías	Tareas	Productos	Salir
--------	--------	------------	--------	-----------	-------

### Categorías

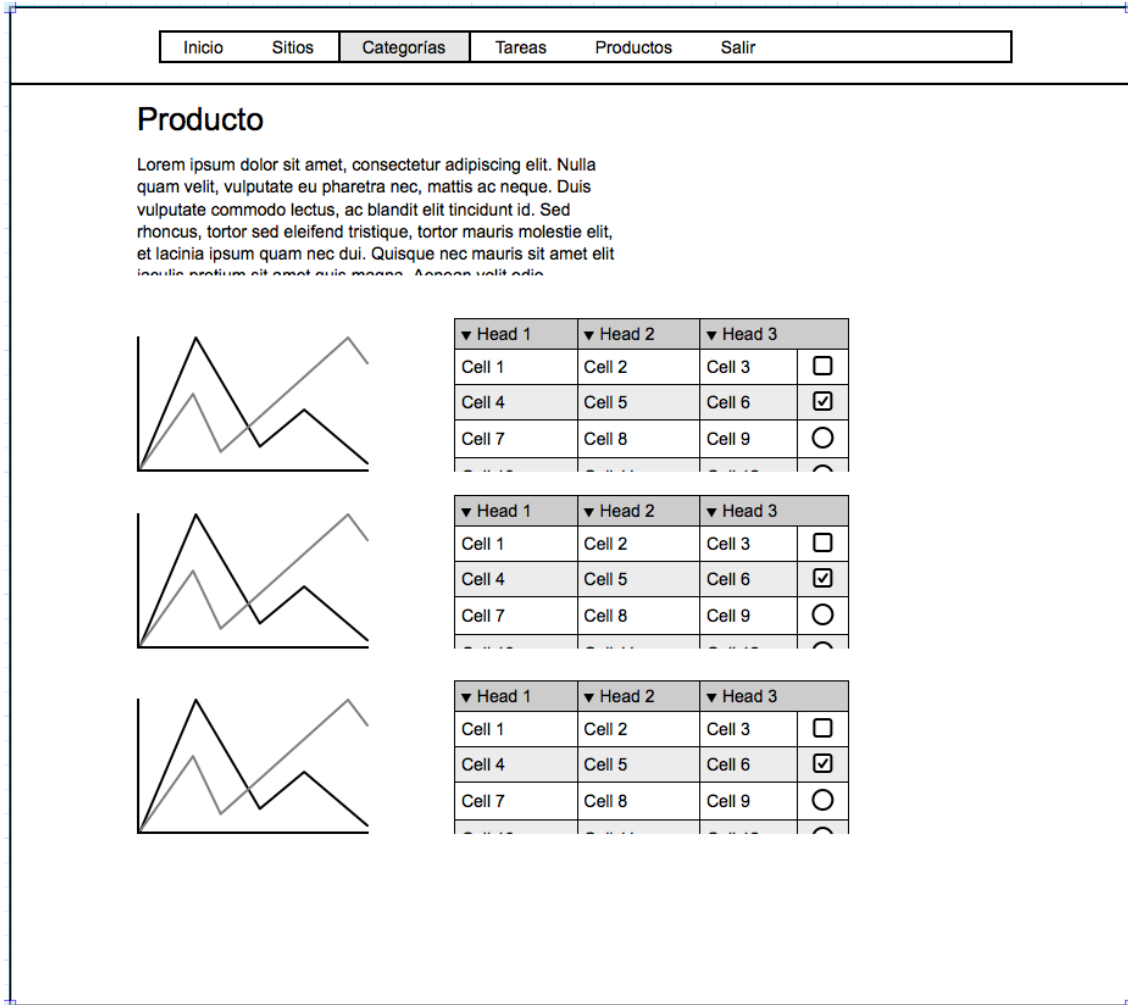
Código  Nombre

▼ Código	▼ Nombre	▼ Sitio	▼ Tareas
Cell 1	Cell 2	Cell 3	Ver
Cell 4	Cell 5	Cell 6	Ver
Cell 7	Cell 8	Cell 9	Ver
Cell 10	Cell 11	Cell 12	Ver

<< Página 2 >>

**Figura 18. Interfaz de categorías de la aplicación de consulta**

Se puede observar un menú horizontal en la parte superior de la aplicación, una zona de filtros y un listado paginado con enlaces que llevan a otras acciones. Todas las demás pantallas de la aplicación (sitios, tareas, productos) de consulta tienen la misma distribución, a excepción de la pantalla de estadísticas de productos (figura 20), que ofrece un espacio para la visualización de gráficas y tablas.



**Figura 19. Interfaz de estadística de productos**

La aplicación de consulta se encuentra desplegada en la URL <http://172.245.180.180:8000/>.

**5.3.2 Aspectos visuales de la interfaz Web de administración.** Para la interfaz de administración se utilizó el módulo administrador que ofrece el framework Django, este viene equipado con su propia capa de presentación, al igual que la interfaz de presentación de resultados, tiene un esquema plano y funcional, con una paleta de colores que involucran al verde aceituna, blanco y azul. En la siguiente figura se puede observar el esquema de distribución del módulo de gestión de categorías.

### Administración de Django

Home
Level
Level
Level
Last

Escoja la categoría a modificar
Añadir Categoría

Acción:   Seleccionados 0 de 10

		▼ Categorías
<input type="checkbox"/>		Cell 1
<input type="checkbox"/>		Cell 2
<input type="checkbox"/>		Cell 3
<input type="checkbox"/>		Cell 4
<input type="checkbox"/>		Cell 5
<input type="checkbox"/>		Cell 6

10 Categorías

**Figura 20. Interfaz de administración de las categorías**

Los demás módulos (sitios y tareas), tienen el mismo esquema de distribución, solo varían los campos que se despliegan en los formularios de registro y en sus respectivos listados. A continuación el esquema del formulario para edición de categorías.

Administración de Django

Home > Level > Level > Level > Last

Modificar Categoría Histórico

Code:

Nombre:

Sitio:

**Figura 21. Formulario de registro y edición de categorías**

Todos los demás formularios de la interfaz administrativa tienen el mismo patrón de diseño. La aplicación de consulta se encuentra desplegada en la URL <http://172.245.180.180:8000/admin>.

**5.3.3 Modelo de clases de la aplicación.** Como se mencionó antes, la aplicación fue desarrollada en el framework Django, este framework ofrece un conjunto de clases (librerías) que se encargan del trabajo repetitivo a la hora de implementar aplicaciones Web, tareas como conexiones a la base de datos, manejo de logs, entre otras, ya están cubiertas por el framework. El desarrollador que utilice Django debe preocuparse por la lógica del negocio que la aplicación debe manejar en vez de invertir tiempo en tareas triviales y repetitivas. A continuación se muestra un conjunto de diagramas de clases conceptual que muestran las relaciones existentes tanto de las clases escritas por el autor como de las clases que ya provee la herramienta, para profundizar en dichas clases el autor recomienda la revisión directa del código fuente.

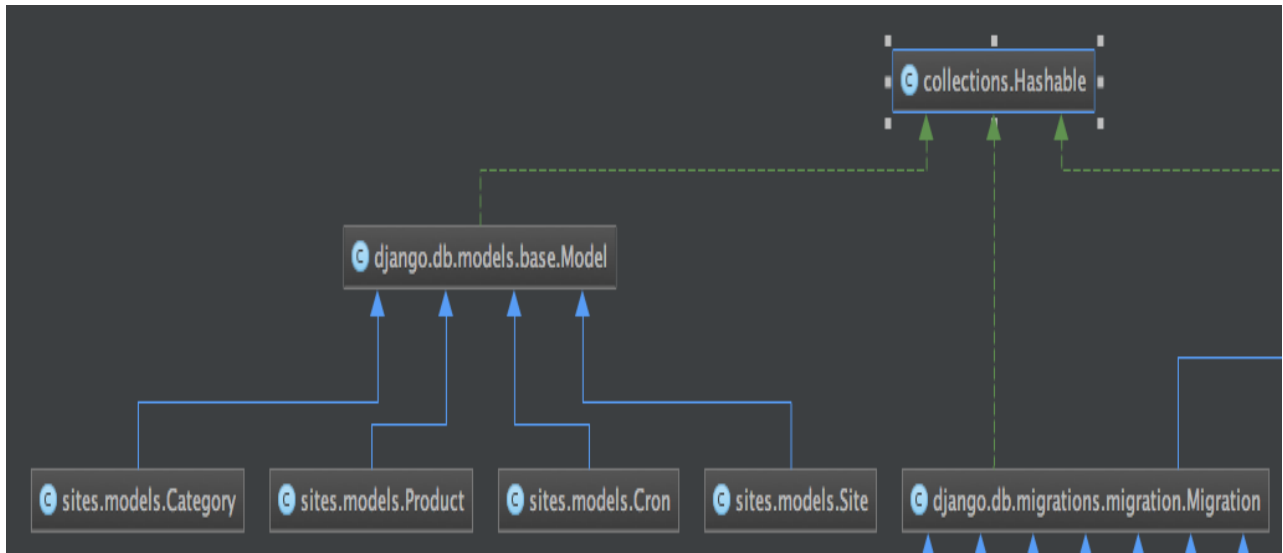


Figura 22. Herencia entre los modelos de la aplicación

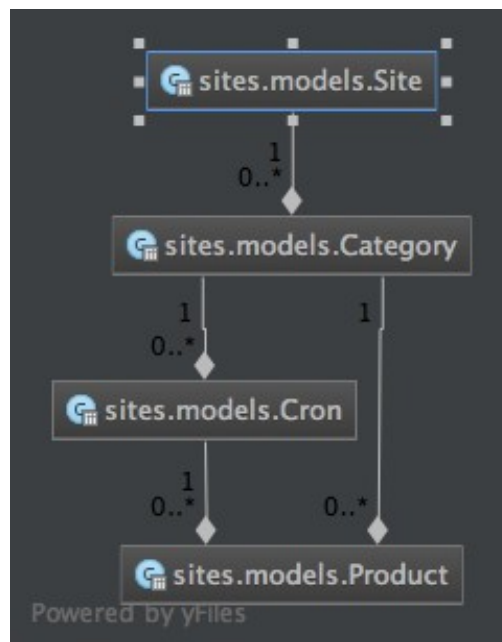
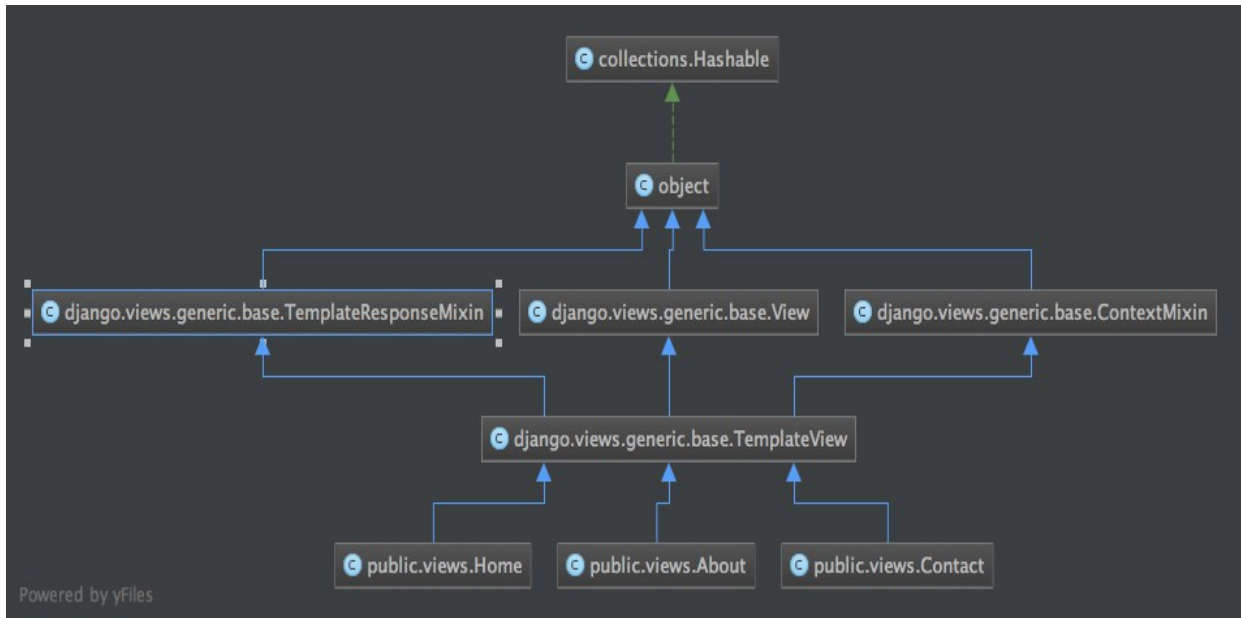
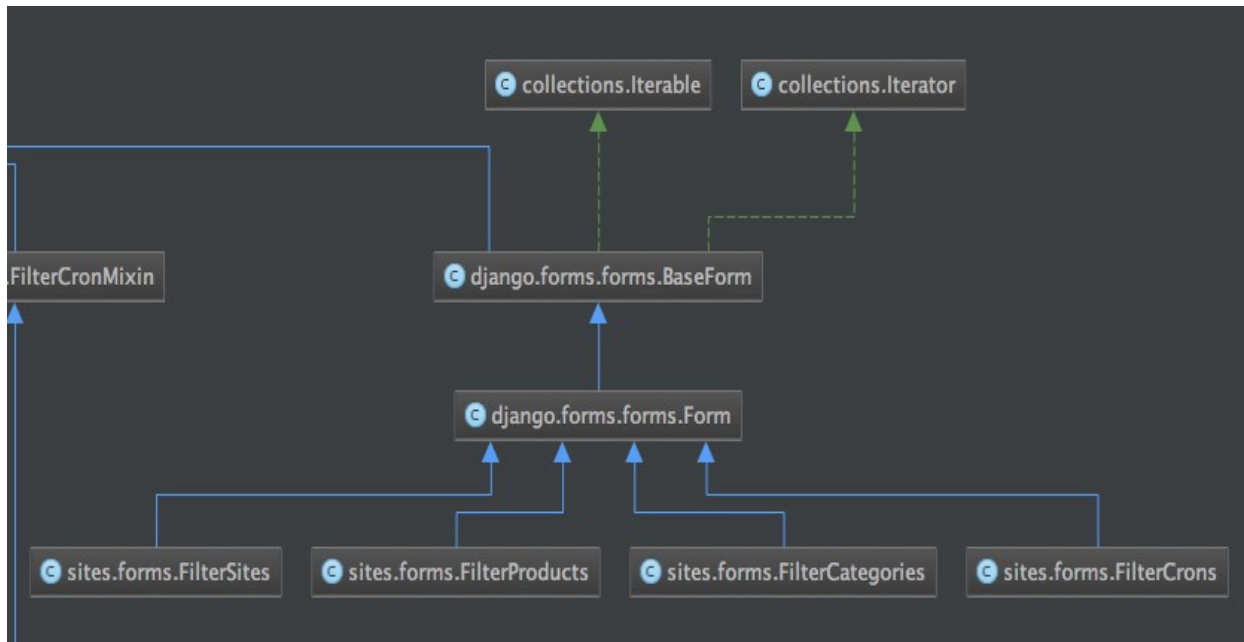


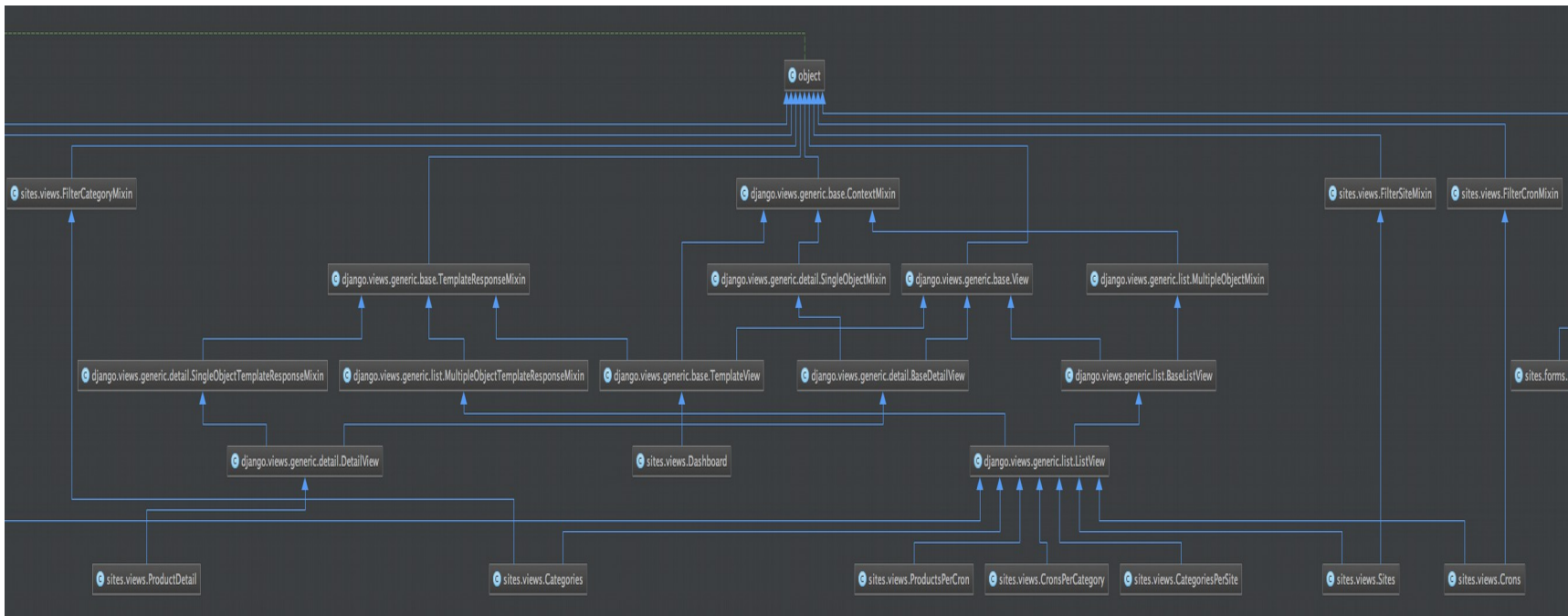
Figura 23. Relaciones entre los modelos de la aplicación



**Figura 24. Clases de las vistas públicas de la aplicación de consulta**



**Figura 25. Clases de los formularios de la aplicación de consulta**



**Figura 26. Clases de las vistas privadas de la aplicación de consulta**



**5.3.4 Consideraciones de despliegue y código fuente.** Actualmente la aplicación se encuentra desplegada sobre un servidor virtual con las siguientes características:

- Ubuntu Server 14.04 de 64 bits.
- PostgreSQL Server 9.3.
- MongoDB 2.4.9.
- Supervisor 3.0.
- Python 2.7.6.
- Django 1.7.1.
- Gunicorn 19.1.1
- Nginx 1.4.6.

El código fuente de la aplicación está alojado en la cuenta de github del autor, que se encuentra en la URL <https://github.com/jorlugaqui/stella>, la aplicación se publicó bajo la licencia GNU AFFERO GENERAL PUBLIC LICENSE (<http://www.gnu.org/licenses/agpl-3.0.html>) que garantiza la realización de obras derivadas (a nivel de software) sin el consentimiento del autor.

Si el lector desea desplegar la aplicación sobre un servidor propio puede seguir la guía publicada en la URL <https://www.digitalocean.com/community/tutorials/how-to-install-and-configure-django-with-postgres-nginx-and-gunicorn>.

**5.3.5 Ejemplos de categorías y productos.** En este apartado se describen algunos ejemplos de categorías y productos con base en el sitio Mercado Libre ([www.mercadolibre.com.co](http://www.mercadolibre.com.co)), esto debido a la facilidad de uso de su API. Para obtener todas la categorías y subcategorías disponibles en Mercado Libre Colombia se debe ejecutar la siguiente instrucción en la línea de comandos del sistema operativo (consola de windows, terminal en unix):

```
$curl https://api.mercadolibre.com/sites/MCO/categories
```

Como requisito a la ejecución de esta instrucción, se debe instalar previamente el paquete curl (<http://curl.haxx.se/>) en el sistema operativo. Es posible consumir el recurso directamente en el navegador Web escribiendo solo la URL:

```
https://api.mercadolibre.com/sites/MCO/categories
```

La diferencia radica en que con curl es posible descargar las categorías al sistema de ficheros local, mientras que en el navegador Web solo se pueden visualizar. A continuación una relación de las principales categorías de Mercado Libre Colombia.

**Tabla 20. Principales categorías de Mercado Libre**

Código	Nombre	URL
MCO1747	Accesorios para Vehículos	<a href="http://home.mercadolibre.com.co/vehiculos-accesorios/">http://home.mercadolibre.com.co/vehiculos-accesorios/</a>
MCO1071	Animales y Mascotas	<a href="http://home.mercadolibre.com.co/animales-y-mascotas/">http://home.mercadolibre.com.co/animales-y-mascotas/</a>
MCO1384	Bebés	<a href="http://home.mercadolibre.com.co/bebes/">http://home.mercadolibre.com.co/bebes/</a>
MCO1743	Carros, Motos y Otros	<a href="http://www.mercadolibre.com.co/vehiculos/">http://www.mercadolibre.com.co/vehiculos/</a>
MCO1051	Celulares y Teléfonos	<a href="http://home.mercadolibre.com.co/telefonos/">http://home.mercadolibre.com.co/telefonos/</a>
MCO1798	Coleccionables y Hobbies	<a href="http://home.mercadolibre.com.co/coleccionables/">http://home.mercadolibre.com.co/coleccionables/</a>
MCO1648	Computación	<a href="http://home.mercadolibre.com.co/computacion/">http://home.mercadolibre.com.co/computacion/</a>
MCO1144	Consolas y Videojuegos	<a href="http://home.mercadolibre.com.co/consolas/">http://home.mercadolibre.com.co/consolas/</a>
MCO1039	Cámaras y Accesorios	<a href="http://home.mercadolibre.com.co/fotografia/">http://home.mercadolibre.com.co/fotografia/</a>
MCO5726	Electrodomésticos	<a href="http://home.mercadolibre.com.co/electrodomesticos/">http://home.mercadolibre.com.co/electrodomesticos/</a>

Para complementar el listado de categorías, se agrega también una relación (tabla 21) de algunos productos que se encuentran registrados en la plataforma.





**Tabla 21. Listado de algunos productos de Mercado Libre**

Código	Categoría	Nombre	URL
MCO413968804	Carros, Motos y Otros	Yamaha 2012 2012	<a href="http://nautica.mercadolibre.com.co/MCO-413968804-yamaha-2012-2012-_JM">http://nautica.mercadolibre.com.co/MCO-413968804-yamaha-2012-2012-_JM</a>
MCO413963342	Carros, Motos y Otros	Harley Davidson Ironhead 1000 Custom	<a href="http://moto.mercadolibre.com.co/MCO-413963342-harley-davidson-ironhead-1000-custom-_JM">http://moto.mercadolibre.com.co/MCO-413963342-harley-davidson-ironhead-1000-custom-_JM</a>
MCO413923040	Carros, Motos y Otros	Ktm Duke 390 251 Cc - 500 Cc	<a href="http://moto.mercadolibre.com.co/MCO-413923040-ktm-duke-390-251-cc-500-cc-_JM">http://moto.mercadolibre.com.co/MCO-413923040-ktm-duke-390-251-cc-500-cc-_JM</a>
MCO413932454	Carros, Motos y Otros	Avion Ultraliviano Quik Silver Sprin Ii	<a href="http://vehiculo.mercadolibre.com.co/MCO-413932454-avion-ultraliviano-quik-silver-sprin-ii-_JM">http://vehiculo.mercadolibre.com.co/MCO-413932454-avion-ultraliviano-quik-silver-sprin-ii-_JM</a>
MCO413990711	Carros, Motos y Otros	Vendo Bomba Hidraulica Caterpillar Para Excavadora 320d	<a href="http://vehiculo.mercadolibre.com.co/MCO-413990711-vendo-bomba-hidraulica-caterpillar-para-excavadora-320d-_JM">http://vehiculo.mercadolibre.com.co/MCO-413990711-vendo-bomba-hidraulica-caterpillar-para-excavadora-320d-_JM</a>
MCO413668494	Animales y Mascotas	Maquina Peluqueria Canina Profesional Genuina Wahl Obsequios	<a href="http://articulo.mercadolibre.com.co/MCO-413668494-maquina-peluqueria-canina-profesional-genuina-wahl-obsequios-_JM">http://articulo.mercadolibre.com.co/MCO-413668494-maquina-peluqueria-canina-profesional-genuina-wahl-obsequios-_JM</a>
MCO413933554	Animales y Mascotas	Repelente Ultrasonico Pulgas Perro Garrapatas Gato Collar	<a href="http://articulo.mercadolibre.com.co/MCO-413933554-repelente">http://articulo.mercadolibre.com.co/MCO-413933554-repelente</a>

			ultrasonico-pulgas-pe-ro-garrapatas-gato-co-llar-_JM
MCO413512113	Animales y Mascotas	Sticky Buddy Quita Motas Y Pelo De Perro Y Gato + Furminator	<a href="http://articulo.mercadolibre.com.co/MCO-413512113-sticky-buddy-quita-motas-y-pelo-de-perro-y-gato-furminator-_JM">http://articulo.mercadolibre.com.co/MCO-413512113-sticky-buddy-quita-motas-y-pelo-de-perro-y-gato-furminator-_JM</a>
MCO413748129	Animales y Mascotas	Sticky Buddy Quita Motas Y Pelo De Perro Y Gato + Obsequios	<a href="http://articulo.mercadolibre.com.co/MCO-413748129-sticky-buddy-quita-motas-y-pelo-de-perro-y-gato-obsequios-_JM">http://articulo.mercadolibre.com.co/MCO-413748129-sticky-buddy-quita-motas-y-pelo-de-perro-y-gato-obsequios-_JM</a>
MCO413695958	Animales y Mascotas	Maquina Peluqueria Canina Wahl 100% Genuina Americana	<a href="http://articulo.mercadolibre.com.co/MCO-413695958-maquina-peluqueria-canina-wahl-100-genuina-americana-_JM">http://articulo.mercadolibre.com.co/MCO-413695958-maquina-peluqueria-canina-wahl-100-genuina-americana-_JM</a>

**5.3.6 Descripción y uso de la herramienta.** En los anexos J y K se encuentran los manuales de usuario tanto para la interfaz de consulta como para la interfaz de administración, en este apartado se describen algunos flujos básicos que el usuario puede realizar en la interfaz de consulta.

**5.3.6.1 Listar sitios.** Para listar los sitios registrados el usuario autenticado puede hacer click en el menú sitios, sobre la franja horizontal superior de la aplicación, el módulo contiene un formulario de búsqueda que permite realizar filtros por el nombre del sitio, además cuenta con paginación de 5 elementos y con un enlace que permite ver las categorías de cada sitio ordenadas por relevancia.

Nombre	URL	Logo	Categorías
Amazon	http://www.amazon.com/		<a href="#">Ver</a>
Linio Colombia	http://www.linio.com.co/		<a href="#">Ver</a>
Mercado Libre	http://www.mercadolibre.com.co/		<a href="#">Ver</a>
eBay	http://www.ebay.com/		<a href="#">Ver</a>






**Figura 27. Módulo de sitios**

**5.3.6.2 Listar categorías.** Para listar las categorías registradas el usuario autenticado puede hacer click en el menú categorías, luego click en el submenú listar categorías. El módulo contiene un formulario de búsqueda que permite realizar filtros por código, nombre o relevancia de la categoría y por nombre del sitio asociado, además cuenta con paginación de 5 elementos y con un enlace que permite ver las tareas programadas hechas sobre dicha categoría.

Sitios Categorías Tareas Productos Salir

## Categorías

Código:  Nombre:  Sitio:  Relevancia:  Buscar

Código	Nombre	Sitio	Tareas
MCO1747	Accesorios para Vehículos		<a href="#">Ver</a>
MCO1071	Animales y Mascotas <span style="background-color: black; color: white; padding: 2px;">Categoría relevante</span>		<a href="#">Ver</a>
MCO1384	Bebés		<a href="#">Ver</a>
MCO1743	Carros, Motos y Otros		<a href="#">Ver</a>
MCO1051	Celulares y Teléfonos		<a href="#">Ver</a>




**Figura 28. Módulo de categorías**

**5.3.6.3 Listar tareas programadas.** Para listar las tareas programadas (crons) el usuario autenticado puede hacer click en el menú tareas, sobre la franja horizontal superior de la aplicación, el módulo contiene un formulario de búsqueda que permite realizar filtros por la categoría asociada y por el campo finalizado, además cuenta con paginación de 5 elementos y con un enlace que permite ver los productos involucrados en la tarea de extracción. Cuando una nueva tarea es registrada a través de la interfaz de administración, la aplicación automáticamente comienza las labores de extracción de datos y de proyección de comportamiento sobre los productos de la categoría asociada al cron.

Sitios Categorías Tareas Productos Salir

## Tareas

Categoría:  ¿Finalizado?:

Días	Inicio	Fin	¿Finalizado?	Código Categoría	Sitio	Productos
12	18 de Enero de 2015	30 de Diciembre de 2014	Si	MCO1747		<a href="#">Ver</a>
15	17 de Enero de 2015	31 de Enero de 2015	No	MCO1071		<a href="#">Ver</a>
20	17 de Enero de 2015	5 de Febrero de 2015	No	MCO1743		<a href="#">Ver</a>

**Figura 29. Módulo de tareas programadas**






**5.3.6.4 Listar productos.** Para listar los productos el usuario autenticado puede hacer click en el menú productos, sobre la franja horizontal superior de la aplicación, el módulo contiene un formulario de búsqueda que permite realizar filtros por código o nombre del producto, por categoría asociada y por número de tarea programada.

Además cuenta con paginación de 5 elementos y con un enlace que permite ver la proyección de su comportamiento, calculada por la aplicación con respecto a las variables precio, cantidad disponible y cantidad vendida.

Sitios Categorías Tareas Productos Salir

## Productos

Código:  Código Nombre:  Nombre Categoría:  Categoría Tarea #:  Tarea #

Código	Nombre	Categoría	Sitio	Tarea #	Estadística
MCO413411483	Adiestrador Ahuyentador Perros, Mod. Radarcán Sc-35 Original	MCO1071		104	<a href="#">Ver</a>
MCO413973461	Akt 2011 126 Cc - 250 Cc	MCO1743		105	<a href="#">Ver</a>
MCO413570757	American Pitbull Terrier Cachorros, Apbt Alta Genética.	MCO1071		104	<a href="#">Ver</a>
MCO413439799	Arenera Gato I Espectacular,nueva, Bonita Y Barata!	MCO1071		104	<a href="#">Ver</a>
MCO413375167	Arenera Sanitaria Gigante Para Gatos , (nueva)	MCO1071		104	<a href="#">Ver</a>

Página 1 of 29 »

Figura 30. Módulo de productos

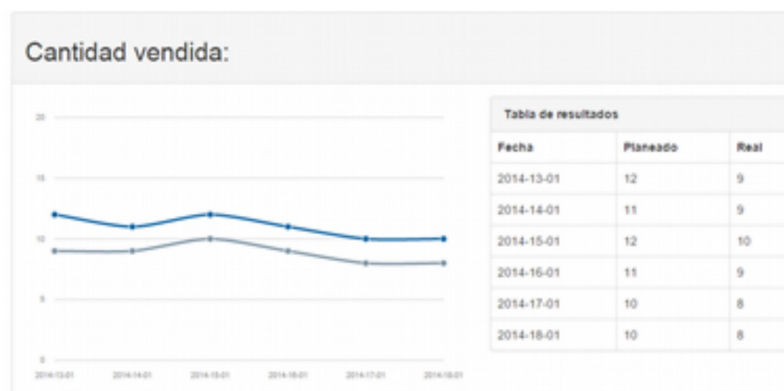


Figura 31. Página de estadísticas



#### **5.4 Descripción de resultados para un producto específico.**

Para comprobar los resultados de la aplicación se programó una tarea de extracción al producto *Ikea Developmental Soft Toy 14pc Vegetable Set Brand New* (<http://www.ebay.ie/itm/Ikea-Developmental-Soft-Toy-14pc-Vegetable-Set-Brand-New-/321647917600>) identificado con el código 321647917600 en la plataforma eBay, por un periodo de 5 días, comprendidos entre el 23 de enero de 2015 y el 27 de enero del mismo año. Este producto pertenece a la categoría *Baby* con código 2893, que se encuentra en la url <http://www.ebay.com/chp/baby>. Después de que la tarea programada descargara el histórico de datos, la aplicación calculó las ecuaciones de proyección con respecto a las variables precio, cantidad disponible y cantidad vendida, con estas ecuaciones se predijo el comportamiento del producto en relación a las mismas tres variables, esta predicción cubre el mismo número de días hacia el futuro que duró la recolección, esto es, los siguientes 5 días a partir de la fecha de finalización de la extracción (del 28 de enero al 2 de febrero del 2015). Los resultados obtenidos por la aplicación se pueden observar en las siguientes gráficas:

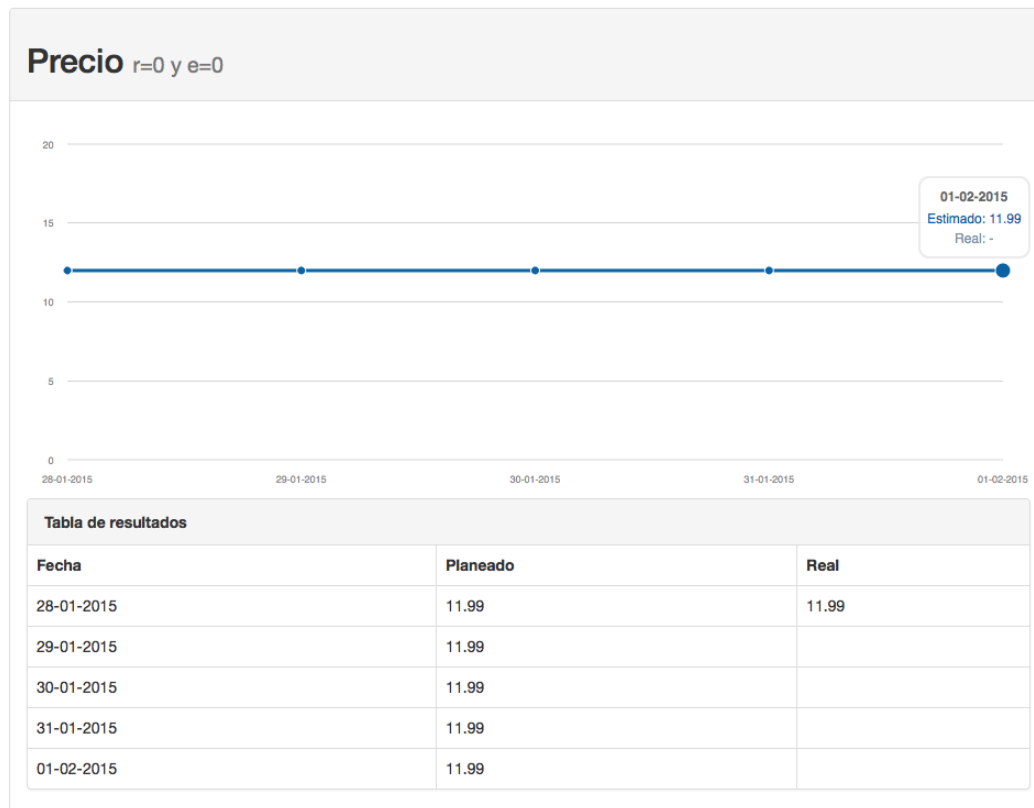


eBay

Categoría: Baby

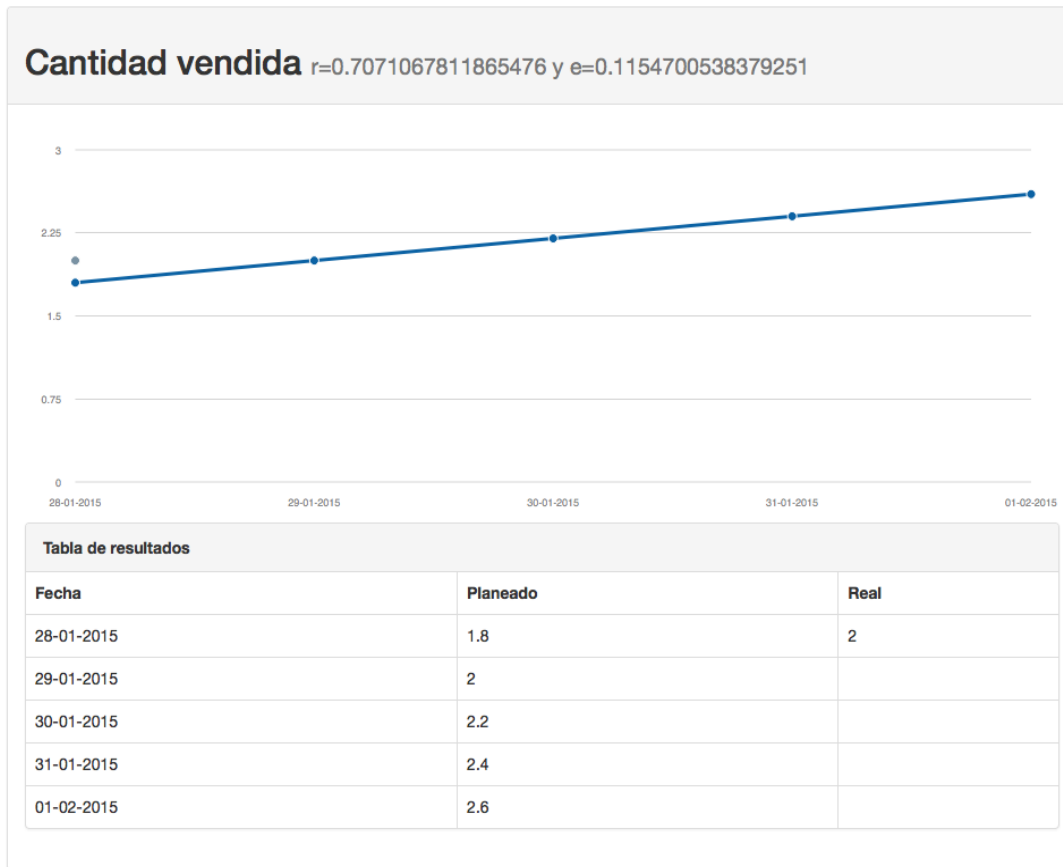
Tarea #: 109

## Ikea Developmental Soft Toy 14pc Vegetable Set Brand New



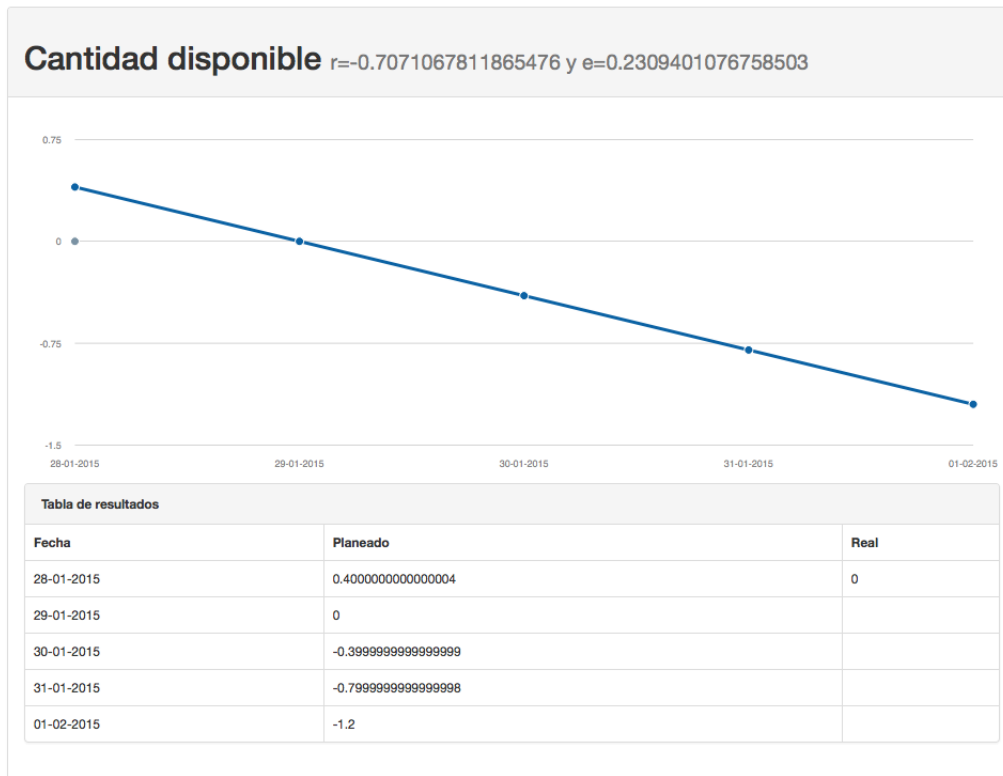
**Figura 32. Proyección de precios para el producto 321647917600 en eBay**

Se puede apreciar que la aplicación proyectó un comportamiento estable para el precio del producto, manteniéndose en 11.99 dolares. También se puede observar la lectura del valor real del producto en el primer día de predicción, este valor se ajusta perfectamente al calculado por el software.



**Figura 33. Proyección de ventas para el producto 321647917600 en eBay**

En la figura 33 se observa una pendiente positiva con respecto al comportamiento de las ventas, esto significa que la aplicación proyectó un aumento de ventas de acuerdo al comportamiento de los históricos. Según los resultados, se espera que haya entre 2 y 3 ventas aproximadamente.



**Figura 34. Proyección de inventario para el producto 321647917600 en eBay**

La figura 34 demuestra que para el segundo día de predicción se espera el agotamiento del inventario del producto. Cada proyección fue calculada con su correspondiente coeficiente de Pearson y su porcentaje de error.

En las figuras anteriores el lector pudo observar la tendencia que el producto tendrá durante el periodo de tiempo de predicción, sin embargo la aplicación también cuenta con un reporte que permite explícitamente establecer la tendencia de cada una de las tres variables evaluadas, a continuación las tendencias obtenidas para este producto:



**eBay**

Categoría: Baby

Tarea #: 109

Volver

**Comportamiento de precios del 28 de Enero de 2015 a 2 de Febrero de 2015**

Código	Nombre	Tendencia
201133571094	1Pairs Animal Baby Infant Kids Foot Sock Rattles Toys Developmental Soft	▬▬
321647917600	Ikea Developmental Soft Toy 14pc Vegetable Set Brand New	▬▬

**Figura 35. Tendencia de precio para el producto 321647917600 en eBay**



**eBay**

Categoría: Baby

Tarea #: 109

Volver

**Comportamiento de ventas del 28 de Enero de 2015 a 2 de Febrero de 2015**

Código	Nombre	Tendencia
201133571094	1Pairs Animal Baby Infant Kids Foot Sock Rattles Toys Developmental Soft	↗
321647917600	Ikea Developmental Soft Toy 14pc Vegetable Set Brand New	↗

**Figura 36. Tendencia de venta para el producto 321647917600 en eBay**



**eBay**

Categoría: Baby

Tarea #: 109

Volver

**Comportamiento de inventarios del 28 de Enero de 2015 a 2 de Febrero de 2015**

Código	Nombre	Tendencia
201133571094	1Pairs Animal Baby Infant Kids Foot Sock Rattles Toys Developmental Soft	=
321647917600	Ikea Developmental Soft Toy 14pc Vegetable Set Brand New	↘

**Figura 37. Tendencia de inventario para el producto 321647917600 en eBay**

En los anteriores reportes se listan los productos de la categoría a la que se le realizó el seguimiento (Baby), dentro de estos, se encuentra el producto de interés para esta descripción (321647917600). Se puede observar al frente de cada ítem una imagen indicando si la tendencia con respecto a cada una de las variables es a la alta, a la baja, o a mantenerse estable.

## 6. Conclusiones

El desarrollo de la investigación permitió conocer el esquema de funcionamiento de cuatro tiendas electrónicas con presencia en Colombia (Mercado Libre, Linio, eBay y Amazon), tanto a nivel de sitio Web como a nivel de APIs, además se construyó una aplicación que permite hacer seguimiento al comportamiento de los productos registrados en dichos sitios con respecto a las variables precio, cantidad vendida y cantidad disponible.

A través de gráficas, los usuarios de la aplicación pueden comparar el comportamiento proyectado con el comportamiento real del producto, como complemento, también pueden diferenciar, para la mayoría de los sitios agregados, cuáles son las categorías con más movimiento. Las personas que realicen compras o ventas sobre los sitios analizados pueden tomar decisiones anticipadas con base en las proyecciones calculadas por los scripts de la aplicación.

Desde los resultados obtenidos se concluye que:

Es posible hacer seguimiento de productos de forma automática desde tiendas electrónicas, pero este seguimiento está sujeto a variables como: términos de uso de los sitios Web, herramientas de consulta que ofrezca la tienda y calidad de publicación de ofertas a nivel DOM/HTML.

Cada plataforma de comercio electrónico tiene una estructura de funcionamiento distinta, que se ve reflejada en la forma como publican sus ofertas, tanto a nivel de HTML como a nivel de sus respectivas APIs.

Las plataformas de comercio electrónico eBay, Amazon y Mercado Libre ofrecen APIs que permiten la consulta de datos asociados a los productos que ofertan en sus Sitios Web, estas APIs son heterogéneas, dependiendo de la plataforma varía su forma de funcionamiento (campos mínimos para realizar una petición, campos devueltos, etc.).

La API más fácil de usar es la de Mercado Libre, la API más difícil de utilizar es la de Ama-

zon.

Linio Colombia es una plataforma que cuenta con un DOM de publicación de productos claro y de fácil inspección, aunque no ofrece una API de consulta.

En Colombia no existen bases de datos públicas asociadas a las plataformas de comercio electrónico que operan en el país, solo a través de la inspección HTML y del uso de APIs se logra la extracción automática de datos.

La mayoría de tiendas electrónicas prohíben el uso del Web Scraping sobre sus sitios Web.

El método de los mínimos cuadrados es una herramienta estadística fundamental para realizar proyecciones de comportamiento de una variable con respecto al tiempo, en el contexto de esta investigación permitió estimar variables como precio, cantidad vendida y cantidad disponible.

El diseño de scripts para la extracción de datos se debe realizar con la premisa de no sobrecargar con peticiones a los sitios objetivos, ya que usos indebidos pueden generar tráfico innecesario y penalizaciones a los dominios desde donde estos se ejecutan.

La técnica del Web Scraping es ideal para la extracción de datos desde sitios Web, pero siempre que exista una API en el sitio objetivo es mejor hacer uso de esta.

Python es un lenguaje de programación con suficiente madurez, ya que cuenta con librerías que implementan el método de los mínimos cuadrados, la técnica del Web Scraping, facilidades para la configuración de tareas programadas y frameworks de desarrollo.

Una combinación entre tecnologías de bases de datos relaciones y no relacionales permite construir repositorios de datos más flexibles y de fácil configuración en ambientes dinámicos de desarrollo de software.

Establecer tendencias comerciales sin el acceso directo a las bases de datos donde se almace-



nan las transacciones es una tarea compleja, sin embargo, desde los datos publicados por las tiendas electrónicas en sus sitios Web o APIs, se pueden establecer comportamientos que evidencien estas tendencias.

## 7. Recomendaciones y Trabajos Futuros

### Recomendaciones

Para el correcto funcionamiento de la aplicación desarrollada, se propone las siguientes recomendaciones:

Instalar la aplicación en un servidor GNU/Linux, de preferencia Debian (<https://www.debian.org/>) o Ubuntu (<http://www.ubuntu.com/>), con una RAM mínima de 2 GB y un disco duro mínimo de 15 GB, con uno o más procesadores por encima de los 1.2 GHz y un ancho de banda mínimo de 512MB mensuales.

Utilizar gunicorn (<http://gunicorn.org/>) como servidor de aplicaciones, nginx (<http://nginx.org/en/>) como servidor Web y PostgreSQL (<http://www.postgresql.org/>) como servidor de base de datos.

Programar las tareas de recolección de datos por un periodo superior a 10 días y no mayor a 30, de modo que el método de los mínimos cuadrados tenga suficiente lecturas sobre las cuales pueda operar y a su vez no se genere tráfico innecesario en el sitio objetivo.

No programar más de un tarea de extracción de datos sobre la misma categoría en el mismo periodo de tiempo.

Preferiblemente agregar sitios Web que cuenten con APIs de consulta.

Implementar un script de extracción por cada nuevo sitio agregado.

Subir al repositorio del proyecto (<https://github.com/jorlugaqui/stella>) las mejoras que el lector haya implementado.

### Trabajos futuros

Tanto el código fuente de la aplicación como la documentación del proyecto cuentan con licencias amigables que permiten el desarrollo de obras derivadas sin mayor problema, desde la

experiencia adquirida, las siguientes propuestas pueden ser aplicadas para darle continuidad al trabajo desarrollado:

Agregar contexto al tiempo de lecturas: significa que la aplicación sea capaz de determinar cuándo se está en temporada alta de ventas por razón de alguna fecha especial (día de las madres, navidad, etc.) y aprovechando este conocimiento pueda establecer tendencias sobre los productos de forma directa.

Establecer notificaciones por categorías: significa que la aplicación de forma asíncrona y automática genere notificaciones cuando pueda detectar variaciones importantes en el comportamiento de las categorías.

Agregar más plataformas de comercio electrónico: consiste en agregar más sitios de comercio electrónico a los cuatro ya estudiados.

Aplicar más métodos estadísticos a las lecturas obtenidas: consiste en profundizar en más métodos estadísticos distintos a los mínimos cuadrados con el fin de establecer mejores conclusiones sobre los datos leídos.

Mejoras en la capa de presentación: significa mejorar la capa de presentación de la aplicación Web.

Volver totalmente administrable la agregación de plataformas: consiste en que, con solo agregar la URL de la plataforma de comercio electrónico y una definición de extracción de datos, la aplicación pueda de forma automática obtener las categorías y las lecturas de los productos.

Contacto con las plataformas agregadas: consiste en establecer contacto con las plataformas agregadas en la aplicación, para tener acceso directo a la información de transacciones vía consulta de base de datos, de modo que las lecturas no se limiten solo a la información publicada en sus sitios Web o APIs.

## Bibliografía

- Adam, Z. (2003). *e-Commerce and e-Business. E-primers. e-ASEAN Task Force, UNDP-APDIP. Modern Information Retrieval: Pearson Education.*
- Aponte, F. (2011). *Prototipo para la detección automática de perfiles de navegación de usuarios de una empresa aplicando técnicas de minería de datos y algoritmos de aprendizaje. Universidad Autónoma de Bucaramanga: Bucaramanga.*
- Arias, M. (2006). *Manual Práctico del Comercio Electrónico. Madrid: La Ley.*
- Baeza, R., & Ribeiro, B. (2007). *Modern Information Retrieval: Pearson Education.*
- Baeza, R. (2011). *Modern Information Retrieval, the concepts and technology behind search. (2 edición). Inglaterra: Pearson Education.*
- Baeza, R. (2006). *The intention behind web queries. In Crestani, F., Ferragina, P., and Sanderson, M., editors, String Processing and Information Retrieval, volume 4209 of Lecture Notes in Computer Science, chapter 9, pages 98–109. Springer Berlin Heidelberg, Berlin: Heidelberg.*
- Blankenship, J. (2011). *Pro Agile .NET Development with Scrum. New York: Apress.*
- Bootstrap. (2015). *Página Web. Recuperado de <http://getbootstrap.com/>.*
- Caicedo, J. (2008). *Aplicación de técnicas de web mining para la búsqueda de patrones de comportamiento en clientes dentro de la web en la Universidad de Nariño. Universidad Antonio Nariño: San Juan de Pasto.*
- Calderón, L. (2011). *Unsupervised Identification of the User's Query Intent in Web Search. Universitat Pompeu Fabra: Barcelona.*
- Cámara Colombiana de Comercio Electrónico (2014). *Página Web. Recuperado de: <http://cc-ce.org.co/>*

- Ceri, A. (2013). Web Information Retrieval. Springer. Recuperado de <http://it-ebooks.info/read/2816/>
- comScore (2014). Página Web. Recuperado de: <http://www.comscore.com/>
- Copeland, Rick. (2012). MongoDB Applied Design Patterns; Practical Use Cases with the Leading NoSQL Database. O'Really Media.
- Dan, Shen., & Ruvini, Jean., & Somaiya, Manas., & Sundaresan, Neel. (2011). Item Categorization in the e-Commerce Domain. eBay Research Labs. San José: USA.
- Da-sheng, W., & Qin-fen, Y., & Li-juan, L. (2009). An Efficient Text Classification Algorithm in E-commerce Application. School of Information Engineering, Zhejiang Forestry College, Lin'an, Zhejiang: China.
- Debian (2015). Página Web. Recuperado de <https://www.debian.org/>
- Diagrama de Dispersión (2015). Página Web. Recuperado de [http://www.ub.edu/aplica\\_infor/spss/cap3-7.htm](http://www.ub.edu/aplica_infor/spss/cap3-7.htm)
- Django (2015). Página Web. Recuperado de <https://djangoproject.com/>
- DOM (2015). Página Web. Recuperado de [https://developer.mozilla.org/en-US/docs/Web/API/Document\\_Object\\_Model](https://developer.mozilla.org/en-US/docs/Web/API/Document_Object_Model)
- Documentación API Mercado Libre. (2014). Página Web. Recuperado de <http://developers.mercadolibre.com>
- Documentación API eBay. (2014). Página Web. <http://go.developer.ebay.com/developers/ebay/documentation-tools>
- Documentación API Amazon. (2014). Página Web. <http://docs.aws.amazon.com/AWSECommerceService/latest/DG/Welcome.html>
- DuckDuck. (2104). Motor de búsqueda. Recuperado de: <https://duckduckgo.com/>

- Dumitru, H., Gibiec, M., Hariri, N., Huang, J. C., Mobasher, B., Herrera, C. C., and Mirakhorli, M. (2011). On-demand feature recommendations derived from mining public product descriptions. In Proceeding of the 33rd international conference on Software engineering, ICSE New York: USA
- eBay. (2014). Página Web. Recuperado de: <http://www.ebay.com/>
- E-commerce Application. School of Information Engineering, Zhejiang Forestry College, Lin'an, Zhejiang.
- Flórez, E. (2002). Prototipo para la detección automática de perfiles de usuarios. Universidad Autónoma de Bucaramanga, Bucaramanga.
- Gginx (2015). Página Web. Recuperado de: <https://www.nginx.org/en/>
- Gunicorn (2015). Página Web. Recuperado de: <https://www.gunicorn.org/>
- Hafernik, Carolyn., & Jansen, Bernard. (2013). Understanding the specificity of web search queries. The Pennsylvania State University: USA.
- HTML. (2015). Página Web. Recuperado de <https://developer.mozilla.org/en-US/docs/Web/HTML>.
- Jansen, B. (2006). How are we searching the world wide web?: a comparison of nine search engine transaction logs. Inf. Process.
- JSON. (2015). Página Web. Recuperado de <http://www.json.org/>.
- Kumar, Sanjay., & Rizvi, SAM. (2011). Information Extraction using Web Usage Mining, Web Scrapping and Semantic Annotation. University School of Information Technology, GGS Indraprastha University, New Delhi: New Delhi
- Laudon, K., & Traver. C. (2011). E-Commerce Business, Technology, Society. Boston: Prentice Hall.

- Laudon, Kennet C. Guercio Traver, Carol. (2011). E-commerce 2011: Business, Technology, Society. Pearson, Pretince Hall.
- Lewandowski, Dirk. (2011). The Influence of Commercial Intent of Search Results on Their Perceived Relevance. Hamburg: University of Applied Sciences.
- Linio Colombia. (2014). Página Web. Recuperado de: <http://www.linio.com.co/>
- Linoff, G., & Berry, M. (2011). Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management. Wiley.
- Long, Bo., & Bian, J., & Dong, Anlei., & Chang, Yi. (2012). Enhancing Product Search by Best-Selling Prediction in eCommerce. Yahoo! Labs: USA.
- Manning, Christoper D., & Raghavan, Prabhakar., & Schütze, Hinrich. (2009). An Introduction to Information Retrieval. (Online edition) .Cambridge UP.
- Martínez, Ciro. (2012). Estadística y muestreo. Bogotá: Ecoe Ediciones.
- McCreary, D. Kelly, Ann. (2013). Making Sense of NoSQL. USA: Manning Publications.
- McKinney, Wes. (2012). Python for Data Analysis; Agile Tools for Real World Data. O'Really Media.
- Mejía Orozco, W. (2011). Prototipo computacional para identificar patrones de comportamiento de usuarios de servicio a domicilio aplicando minería de datos. Universidad Autónoma de Bucaramanga. Valledupar.
- Mercado Libre. (2014). (Página Web). Recuperado de <http://www.mercadolibre.com.co/>
- Mínimos Cuadrados. (2015). Página Web. Recuperado de [http://catarina.udlap.mx/u\\_dl\\_a/tales/documentos/lad/mercado\\_g\\_ja/apendiceC.pdf](http://catarina.udlap.mx/u_dl_a/tales/documentos/lad/mercado_g_ja/apendiceC.pdf)
- MongoDB. (2015). Página Web. Recuperado de <http://www.mongodb.org/>
- Monomi. (2014). Página Web. Recuperado de: <http://monomi.co/>

- Murphy, J. (1999). Technical analysis of the financial markets. New York: Prentice Hall.
- Na, Li., & Geng, Peng., & Hang, Chen., & Jiaying, Bao. (2013) A Prediction Study on E-commerce Orders Based on Site Search Data. University of Chinese Academy of Sciences, UCAS. Beijing, China.
- Newsha Ghoreishi, Seyyedeh., & Sun, Aixin. (2013). Predicting Event-Relatedness of Popular Queries. School of Computer Engineering, Nanyang Technological University. Singapore.
- Numpy. (2015). Página Web. Recuperado de <http://www.numpy.org/>
- OLX. (2014). Página Web. Recuperado de [www.olx.com](http://www.olx.com)
- Ouyang, L., & Dong, R., & Zou, B. (2009). Information Extraction Based on Table Area Locating for E-Commerce Websites. Central: South University Changsha
- Parikh, N. & Sriram, P. & Al Hasan, M. (2013). On segmentation of eCommerce Queries. Ebay Inc. San Jose, CA. Indiana University – Purdue University, Indianapolis, IN. USA.
- Pitt, David. (2014). Modern Web Essentials, using JavaScript and HTML5. Kansas: Keyhole Software.
- PostgreSQL (2015). Página Web. Recuperado de <https://www.postgresql.org/>.
- Python Library for Amazon Product Advertising API. (2014). Página Web. Recuperado de <https://github.com/yoavaviram/python-amazon-simple-product-api>.
- Rakuten. (2014). Página Web. Recuperado de <http://www.rakuten.co.jp/> <http://global.rakuten.com/en/>.
- REST Tutorial (2015). Página Web. Recuperado de <http://www.restapitutorial.com/>.
- REST Vs SOAP (2015). Página Web. Recuperado de <http://spf13.com/post/soap-vs-rest>.



- Regresión Lineal Simple (2015). Página Web. Recuperado de [http://www.vitutor.net/2/16/regresion\\_lineal.html](http://www.vitutor.net/2/16/regresion_lineal.html)
- Schema of Product. (2014). Página Web. Recuperado de <http://schema.org/Product>
- Schmarzo, B. (2013). *Big Data; Understanding How Data Powers Big Business*. Wiley.
- Schrenk, M. (2012). *Webbots, Spiders, and Screen Scrapers; A Guide to Developing Internet Agents with PHP/CURL*. Starch Press.
- Scrapy. (2015). Página Web. Recuperado de <http://scrapy.org/>
- SDK Python eBay. (2015). Página Web. Recuperado de <https://github.com/timotheus/ebaysdk-python>.
- SDK Python Mercado Libre. (2015). Página Web. Recuperado de <http://github.com/mercadolibre/python-sdk>
- Shen, Zegian., & Sundaresan, Neel. (2011). *eBay: an E-commerce marketplace as a complex network*. eBay Research Labs. San Jose: CA, USA.
- Sieg, A., Mobasher, B., and Burke, R. (2007). Representing context in web search with ontological user profiles. In Kokinov, B., Richardson, D. C., Roth-Berghofer, T. R., Vieu L.
- Spink, A., Jansen, B. J., Wolfram, D., and Saracevic, T. (2002). *From E-Sex to E-Commerce: Web search changes*. Computer..
- Stella (2015). Página Web. Recuperado de <https://github.com/jorlugaqui/stella>
- Taobao. (2014). Página Web. Recuperado de [http://www.taobao.com/market/global/index\\_new.php](http://www.taobao.com/market/global/index_new.php)
- Ubuntu (2015). Página Web. Recuperado de <https://www.ubuntu.com/>
- UTF-8. (2015). Página Web. Recuperado de <http://tools.ietf.org/html/rfc3629>
- Usability. (2014). Pagina web. Recuperado de: *Usability Best Practices For eCommerce Web-*

sites: <http://spyrestudios.com/usability-best-practices-for-ecommerce-websites/>

Xiaojing, Liu., & Liwei, Zhang., & Weiqing, Wang. (2012). The Mechanism Analysis of the Impact of Ecommerce to the Changing of Economic Growth Mode. ChangChun University of Chinese Medicine. Inner Mongolia Academy of Social Science Institute of Urban Development. Shijiazhuang City Vocational College: China

XML. (2015). Página Web. Recuperado de <http://www.w3schools.com/xml/>.

Zhou, Y., Mobasher, B. (2006). Web User Segmentation Based on a Mixture of Factor Analyzers. K. Bauknecht: Berlin.

## Anexos

### Anexo A. Términos de uso de los sitios tomados como muestra

Términos de uso de los sitios tomados como muestra

Sitio	URLs términos y condiciones
Mercado Libre	<a href="http://ayuda.mercadolibre.com.co/seguro_terminos">http://ayuda.mercadolibre.com.co/seguro_terminos</a> <a href="http://developers.mercadolibre.com/terms-and-conditions/">http://developers.mercadolibre.com/terms-and-conditions/</a>
Tiendas Monomi	<a href="http://monomi.co/terminos">http://monomi.co/terminos</a>
OLX Colombia	No aplica
Linio Colombia	<a href="http://www.linio.com.co/terminos/">http://www.linio.com.co/terminos/</a>
eBay	<a href="http://pages.ebay.com/help/policies/everyone-ov.html">http://pages.ebay.com/help/policies/everyone-ov.html</a> <a href="http://developer.ebay.com/join/licenses/individual/api.pdf">http://developer.ebay.com/join/licenses/individual/api.pdf</a> <a href="http://pages.ebay.com/co/es-co/help/policies/overview.html">http://pages.ebay.com/co/es-co/help/policies/overview.html</a>
Amazon	<a href="https://developer.amazonservices.com/gp/mws/api.html/187-2611253-5758501?ie=UTF8&amp;group=products&amp;section=products&amp;version=latest">https://developer.amazonservices.com/gp/mws/api.html/187-2611253-5758501?</a> <a href="https://affiliate-program.amazon.com/gp/advertising/api/detail/main.html">ie=UTF8&amp;group=products&amp;section=products&amp;version=latest</a> <a href="https://affiliate-program.amazon.com/gp/advertising/api/detail/main.html">https://affiliate-program.amazon.com/gp/advertising/api/detail/main.html</a> <a href="http://www.amazon.com/gp/help/customer/display.html/ref=footer_cou?ie=UTF8&amp;nodeId=508088">http://www.amazon.com/gp/help/customer/display.html/ref=footer_cou?ie=UTF8&amp;nodeId=508088</a>
Tsmart	<a href="http://www.tmart.com/HelpCenter/Terms-of-Use.html">http://www.tmart.com/HelpCenter/Terms-of-Use.html</a>
Banggood	<a href="http://www.banggood.com/Terms-And-Conditions_hl42">http://www.banggood.com/Terms-And-Conditions_hl42</a>
Dafiti	<a href="http://www.dafiti.com.co/terminos-y-condiciones/">http://www.dafiti.com.co/terminos-y-condiciones/</a>
Exito	<a href="http://www.exito.com/about/content.jsp?id=terminos">http://www.exito.com/about/content.jsp?id=terminos</a>
Falabella	<a href="http://www.falabella.com.co/falabella-co/category/cat111251/Terminos-y-condiciones">http://www.falabella.com.co/falabella-co/category/cat111251/Terminos-y-condiciones</a>
Alibaba	<a href="http://www.alibaba.com/help/safety_security/policies_rules/others/002.html">http://www.alibaba.com/help/safety_security/policies_rules/others/002.html</a>
Alkosto	<a href="http://www.alkosto.com/terminos">http://www.alkosto.com/terminos</a>
Home Center	<a href="http://portal.homecenter.com.co/terminosycondiciones">http://portal.homecenter.com.co/terminosycondiciones</a>

### Anexo B. Lista de APIs utilizadas por plataforma

Lista de APIs utilizadas por plataforma

Sitio	API	URL
Mercado Libre	General	<a href="https://api.mercadolibre.com/">https://api.mercadolibre.com/...</a>
eBay	Finding API	<a href="http://svcs.ebay.com/services/search/FindingService/v1?...">http://svcs.ebay.com/services/search/FindingService/v1?...</a>
	Shopping API	<a href="http://open.api.ebay.com/shopping?...">http://open.api.ebay.com/shopping?...</a>
Amazon	Adversiting API	<a href="http://webservices.amazon.com/onca/xml?...">http://webservices.amazon.com/onca/xml?...</a>

**Anexo C. Esquema de licenciamiento de la documentación y de la aplicación Web**

Esquema de licenciamiento de la documentación y de la aplicación Web

Producto	Licencia
Aplicación Web Stella (Crawlers + interfaz Web)	GNU AFFERO GENERAL PUBLIC LICENSE Version 3, Disponible en <a href="http://www.gnu.org/licenses/agpl-3.0.html">http://www.gnu.org/licenses/agpl-3.0.html</a>
Documento de investigación	Creative Commons Attribution-NonComercial-NoDerivates 4.0 International, Disponible en: <a href="http://creativecommons.org/licenses/by-nc-nd/4.0/">http://creativecommons.org/licenses/by-nc-nd/4.0/</a>

### Anexo D. Documento JSON devuelto por la API de Mercado Libre

```

{
  "id": "MLA533657947",
  "site_id": "MLA",
  "title": "Apple Ipod Touch 32gb 5ta Generacion Video Hd Retina Ios7",
  "subtitle": null,
  "seller": {
    "id": 7055243,
    "power_seller_status": "platinum",
    "car_dealer": false,
    "real_estate_agency": false
  },
  "price": 4849,
  "currency_id": "ARS",
  "available_quantity": 809,
  "sold_quantity": 190,
  "buying_mode": "buy_it_now",
  "listing_type_id": "gold",
  "stop_time": "2015-01-18T15:16:38.000Z",
  "condition": "new",
  "permalink": "http://articulo.mercadolibre.com.ar/MLA-533657947-apple-ipod-touch-32gb-5ta-generacion-video-hd-retina-ios7- JM",
  "thumbnail": "http://mla-s1-p.mlstatic.com/19052-MLA20164236472_092014-I.jpg",
  "accepts_mercadopago": true,
  "installments": {
    "quantity": 12,
    "amount": 618.21,
    "currency_id": "ARS"
  },
  "address": {
    "state_id": "AR-C",
    "state_name": "Capital Federal",
    "city_id": "TUxBQ1BBTDI1MTVa",
    "city_name": "Palermo"
  },
  "shipping": {
    "free_shipping": true,
    "mode": "me2"
  },
  "seller_address": {
    "id": 99717158,
    "comment": "",
    "address_line": "",
    "zip_code": "",
    "country": {
      "id": "AR",
      "name": "Argentina"
    },
    "state": {
      "id": "AR-C",
      "name": "Capital Federal"
    },
    "city": {
      "id": "TUxBQ1BBTDI1MTVa",
      "name": "Palermo"
    },
    "latitude": -34.59081,
    "longitude": -58.414223
  },
  "attributes": [
  ],
  "original_price": null,
  "category_id": "MLA43718"
}

```

## Anexo E. Documento JSON devuelto por la API de eBay

```
{
  "itemId": [
    "261670479937"
  ],
  "title": [
    "Harry Potter and the Order of the Phoenix "
  ],
  "globalId": [
    "EBAY-US"
  ],
  "primaryCategory": [
    {
      "categoryId": [
        "617"
      ],
      "categoryName": [
        "DVDs & Blu-ray Discs"
      ]
    }
  ],
  "galleryURL": [
    "http://thumbs2.ebaystatic.com/m/mAfb0gOTp7rxLX8nYO5rcag/140.jpg"
  ],
  "viewItemURL": [
    "http://www.ebay.com/itm/Harry-Potter-and-Order-Phoenix-/261670479937?pt=US_DVD_HD_DVD_Blu_ray"
  ],
  "productId": [
    {
      "@type": "ReferenceID",
      "__value__": "109102746"
    }
  ],
  "paymentMethod": [
    "PayPal"
  ],
  "autoPay": [
    "false"
  ],
  "postalCode": [
    "49337"
  ],
  "location": [
    "Newaygo,MI,USA"
  ],
  "country": [
    "US"
  ],
  "shippingInfo": [
    {
      "shippingType": [
        "Calculated"
      ],
      "shipToLocations": [
        "US",
        "CA",
        "GB",
        "AU",
        "MX"
      ],
      "expeditedShipping": [
        1. "false"
      ],
      "oneDayShippingAvailable": [
        2. "false"
      ],
      "handlingTime": [

```

```

        3. "3"
      ]
    },
    "sellingStatus":[
      {
        "currentPrice":[
          {
            "@currencyId":"USD",
            "__value__":"3.5"
          }
        ],
        "convertedCurrentPrice":[
          {
            "@currencyId":"USD",
            "__value__":"3.5"
          }
        ],
        "sellingState":[
          "Active"
        ],
        "timeLeft":[
          "P2DT11H30M51S"
        ]
      }
    ],
    "listingInfo":[
      {
        "bestOfferEnabled":[
          "true"
        ],
        "buyItNowAvailable":[
          "false"
        ],
        "startTime":[
          "2014-11-22T01:17:44.000Z"
        ],
        "endTime":[
          "2014-11-24T13:17:29.000Z"
        ],
        "listingType":[
          "FixedPrice"
        ],
        "gift":[
          "false"
        ]
      }
    ],
    "returnsAccepted":[
      "true"
    ],
    "condition":[
      {
        "conditionId":[
          "4000"
        ],
        "conditionDisplayName":[
          "Very Good"
        ]
      }
    ],
    "isMultiVariationListing":[
      "false"
    ],
    "topRatedListing":[
      "false"
    ]
  }
}

```



## Anexo F. Documento XML devuelto por la API de Amazon

```

<BrowseNode>
  <BrowseNodeId>163357</BrowseNodeId>
  <Name>Comedy</Name>
  <Children>
    <BrowseNode>
      <BrowseNodeId>599826</BrowseNodeId>
      <Name>Boxed Sets</Name>
    </BrowseNode>
    <BrowseNode>
      <BrowseNodeId>538712</BrowseNodeId>
      <Name>African American Comedy</Name>
    </BrowseNode>
    <BrowseNode>
      <BrowseNodeId>163358</BrowseNodeId>
      <Name>Animation</Name>
    </BrowseNode>
    <BrowseNode>
      <BrowseNodeId>720556</BrowseNodeId>
      <Name>Black Comedy</Name>
    </BrowseNode>
    <BrowseNode>
      <BrowseNodeId>291102</BrowseNodeId>
      <Name>British</Name>
    </BrowseNode>
  </Children>
</Item>
  <Ancestors>
    <BrowseNode>
      <BrowseNodeId>549726</BrowseNodeId>
      <Name>Performing Arts</Name>
      <Ancestors>
        <BrowseNode>
          <BrowseNodeId>5</BrowseNodeId>
          <Name>Entertainment</Name>
          <Ancestors>
            <BrowseNode>
              <BrowseNodeId>1000</BrowseNodeId>
              <Name>Subjects</Name>
              <Ancestors>
                <BrowseNode>
                  <BrowseNodeId>1000</BrowseNodeId>
                  <Name>Books</Name>
                </BrowseNode>
              </Ancestors>
            </BrowseNode>
          </Ancestors>
        </BrowseNode>
      </Ancestors>
    </BrowseNode>
  </Ancestors>

```

... El documento XML continua, según campos encontrados en la base de datos.

## Anexo G. Estructura general del DOM de categorías en Linio Colombia

```

<ul class="nav__menu">
... # HTML que se puede despreciar en el escaneo.
<li class="nav__element" id="id-category-1" data-cat-id="...">
  <a href="main-url" class="nav__element-link">Category Name 1</a>
  <div class="nav__element-container">
    <div class="row">
      <div class="col-3">
        <a href="url-to-sub-category-1" class="subnav__title">SubCategory Name 1</a>
        <a href="url-to-sub-category-2" class="subnav__title">SubCategory Name 2</a>
      </div>
      <div class="col-3">
        <a href="url-to-sub-category-3" class="subnav__title">SubCategory Name 3</a>
        <a href="url-to-sub-category-4" class="subnav__title">SubCategory Name 4</a>
      </div>
    </div>
  </div>
</li>
<li class="nav__element" id="id-category-2" data-cat-id="...">
  <a href="main-url" class="nav__element-link">Category Name 2</a>
  <div class="nav__element-container">
    <div class="row">
      <div class="col-3">
        <a href="url-to-sub-category-1" class="subnav__title">SubCategory Name 1</a>
        <a href="url-to-sub-category-2" class="subnav__title">SubCategory Name 2</a>
      </div>
      <div class="col-3">
        <a href="url-to-sub-category-3" class="subnav__title">SubCategory Name 3</a>
        <a href="url-to-sub-category-4" class="subnav__title">SubCategory Name 4</a>
      </div>
    </div>
  </div>
</li>
... # HTML que se puede despreciar en el escaneo.
</ul>

```

## Anexo H. SQL generado a partir del modelo relacional

```

CREATE TABLE "site" (
  "id" SERIAL PRIMARY KEY,
  "name" VARCHAR(100) NOT NULL,
  "url" VARCHAR(255) NOT NULL,
  "type" VARCHAR(10) NOT NULL
);

CREATE TABLE "category" (
  "id" SERIAL PRIMARY KEY,
  "code" VARCHAR(10) NOT NULL,
  "name" VARCHAR(255) NOT NULL,
  "site" INTEGER NOT NULL
);

CREATE INDEX "idx_category__site" ON "category" ("site");

ALTER TABLE "category" ADD CONSTRAINT "fk_category__site" FOREIGN KEY
("site") REFERENCES "site" ("id");

CREATE TABLE "cron" (
  "id" SERIAL PRIMARY KEY,
  "days" INTEGER NOT NULL,
  "start_on" DATE NOT NULL,
  "finish_on" DATE NOT NULL,
  "category" INTEGER NOT NULL,
  "executed" BOOLEAN NOT NULL,
  "script" TEXT NOT NULL
);

CREATE INDEX "idx_cron__category" ON "cron" ("category");

ALTER TABLE "cron" ADD CONSTRAINT "fk_cron__category" FOREIGN KEY ("cat-
egory") REFERENCES "category" ("id");

CREATE TABLE "endpoint" (
  "id" SERIAL PRIMARY KEY,
  "url" VARCHAR(255) NOT NULL,
  "type" TEXT NOT NULL,
  "site" INTEGER NOT NULL
);

CREATE INDEX "idx_endpoint__site" ON "endpoint" ("site");

ALTER TABLE "endpoint" ADD CONSTRAINT "fk_endpoint__site" FOREIGN KEY
("site") REFERENCES "site" ("id");

CREATE TABLE "user" (
  "id" SERIAL PRIMARY KEY,
  "first_name" VARCHAR(35) NOT NULL,
  "last_name" VARCHAR(35) NOT NULL,
  "username" VARCHAR(32) NOT NULL,
  "email" VARCHAR(75) NOT NULL,
  "password" VARCHAR(32) NOT NULL
)

```

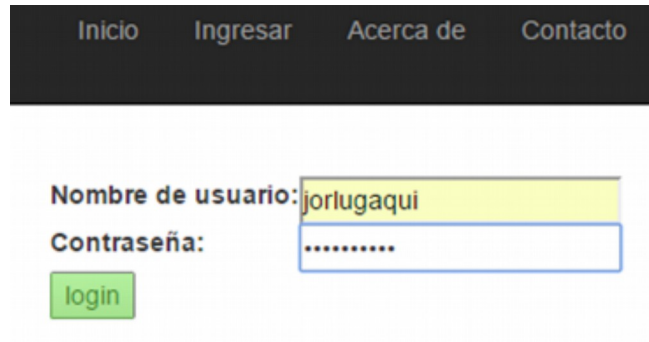
## Anexo I. Instrucción JavaScript de ejemplo para crear un documento en la base de datos no relacional

```
db.mcl.insert({
  "category_id":23,
  "category_name":"Electrónica",
  "cron_id":12,
  "site_id":3,
  "site_name":"Mercado Libre",
  "product_id":"MB983",
  "product_name":"Cámara Fotográfica",
  "history":[
    {
      "date":"2014-07-01",
      "sold_quantity":12,
      "available_quantity":2,
      "price":23.98
    },
    {
      "date":"2014-08-01",
      "sold_quantity":11,
      "available_quantity":1,
      "price":23.98
    }
  ]
})
```

## Anexo J. Manual de usuario de la interfaz de consulta

### Ingreso a la aplicación de consulta

Al ingresar a la página el usuario debe autenticarse, para esto, debe diligenciar los campos solicitados (Nombre de usuario y Contraseña), luego dar click en el botón “login”.



The screenshot shows a login form with a dark navigation bar at the top containing the links: Inicio, Ingresar, Acerca de, and Contacto. Below the navigation bar, the form has two input fields: 'Nombre de usuario:' with the value 'jorlugaqui' and 'Contraseña:' with a masked password '.....'. A green 'login' button is positioned below the password field.

### Ingreso a la aplicación

Luego, será dirigido a la pantalla de inicio

### Dashboard de usuario autenticado



Y encontrará los siguientes módulos:

- Sitios
- Categorías
- Tareas

- Productos

## Sitios

Nombre	URL	Tipo	Categorías
Amazon	http://www.amazon.com/	api	<a href="#">Ver</a>
Linio Colombia	http://www.linio.com.co/	scraping	<a href="#">Ver</a>
Mercado Libre	http://www.mercadolibre.com.co/	api	<a href="#">Ver</a>
eBay	http://www.ebay.com/	api	<a href="#">Ver</a>

### Página de sitios

En este módulo podrá observar los sitios registrados en la app, y podrá:

#### Consultar los sitios.

Aquí seleccionará el sitio que está buscando, luego podrá ver las categorías de dicho sitio dando click en “ver”, de la columna categorías del sitio.

Nombre	URL	Tipo	Categorías
Amazon	http://www.amazon.com/	api	<a href="#">Ver</a>
Linio Colombia	http://www.linio.com.co/	scraping	<a href="#">Ver</a>
Mercado Libre	http://www.mercadolibre.com.co/	api	<a href="#">Ver</a>
eBay	http://www.ebay.com/	api	<a href="#">Ver</a>

### Filtros en la página de sitios

Al dar click en “ver”, de la columna categorías, será enviado a la categoría del sitio, en este ejemplo “Mercado Libre”, le mostrará las categorías con todos sus datos (código, nombre, sitio, un enlace de tareas).

### Categorías para el sitio Mercado Libre

Código	Nombre	Sitio	Tareas
LkvU	Exercitationem atque culpa id minima possimus totam ea rerum quia velit veritatis, iusto esse nobis porro eum, id porro est odio non ullam vero consequatur iusto at. Quis quae in, laboriosam sint inventore eum hic est perferendis	Mercado Libre	<a href="#">Ver</a>
drRJ ILI	Pariatur distinctio earum incidun	Mercado Libre	<a href="#">Ver</a>

### Página categorías por sitio

### Categorías

Inicio
Sitios
**Categorías**
Tareas
Productos
Salir

## Categorías

Código:  Nombre:  Sitio:

Código	Nombre	Sitio	Tareas
epKkDrUT	Accusantium hic delectus id facere iste libero quia asperiores perferendis vel? Mollitia quidem suscipi, neque ratione nihil fugit at, voluptatum commodi molestiae tempore ex iusto vel, modi amet blanditis assumenda expedita natus delenti ipsa re	Linio Colombia	<a href="#">Ver</a>
Sb7kgnTOsq	Ad iste minima qui doloribus dicta eligendi libero fugiat amet, libero porro fugiat incidunt tempora perferendis necessitatibus ab officis dolore, voluptatum mollitia illum volupta	eBay	<a href="#">Ver</a>
pTRsaeH	Animi iusto tempore sunt quas minus a rem facilis, repudiandae numquam aliquam fug	eBay	<a href="#">Ver</a>
tNSGrRQrq	Dignissimos repellat eius beatae, fugiat quasi quos cupiditate possimus	Amazon	<a href="#">Ver</a>
LkvU	Exercitationem atque culpa id minima possimus totam ea rerum quia velit veritatis, iusto esse nobis porro eum, id porro est odio non ullam vero consequatur iusto at. Quis quae in, laboriosam sint inventore eum hic est perferendis	Mercado Libre	<a href="#">Ver</a>

Página 1 of 2
>

### Filtros en página de categorías

En este módulo podrá observar las categorías de los sitios (comercio electrónico) y podrá:

## Consultar las Categorías.

Aquí buscará la categoría por código, nombre o sitio, cabe resaltar que en este módulo encontrará la paginación correspondiente a la búsqueda de la categoría.

Una vez ubique la categoría deseada, dará click en “ver” de la columna tareas.

**Categorías**

Código:  Nombre:  Sitio:

Código	Nombre	Sitio	Tareas
epKAKDrUt	Accusantium hic delectus id facere iste libero quia asperiores perferendis vel? Mollitia quidem suscipit, neque ratione nihil fugit at, voluptatum commodi molestiae tempore ex iusto vel, modi amet blanditis assumenda expedita natus deleniti ipsa re	Linio Colombia	<a href="#">Ver</a>
SbTkgntOsq	Ad iste minima qui doloribus dicta eligendi libero fugiat amet, libero porro fugiat incurrunt tempora perferendis necessitatibus ab officis dolore, voluptatum mollitia illum volupta	eBay	<a href="#">Ver</a>
pTRsaeH	Animi iusto tempore sunt quas minus a rem facilis, repudiandae numquam aliquam fug	eBay	<a href="#">Ver</a>
bNSGrRQiq	Dignissimos repellat eius beatae, fugiat quasi quos cupiditate possimus	Amazon	<a href="#">Ver</a>
Lkvtu	Exercitationem atque culpa id minima possimus totam ea rerum quia velit veritatis, iusto esse nobis porro eum, id porro est odio non ullam vero consequatur iusto at. Quis quae in, laboriosam sint inventore eum hic est perferendis	Mercado Libre	<a href="#">Ver</a>

Página 1 of 2 »

### Página de categorías y paginación

Al dar click en “ver” de la columna Tareas, será enviado a la(s) tarea(s) programadas (seguimiento de productos) asignadas a dicha categoría, en este ejemplo se mostrará la tarea de la categoría “Animi iusto..” (aquí se observan todos los datos de esta tarea).

**Tareas para la categoría Animi iusto tempore sunt quas minus a rem facilis, repudiandae numquam aliquam fug**

Días	Inicio	Fin	¿Terminado?	Código Categoría	Productos
34	30 de Mayo de 2015	25 de Septiembre de 2010	False	pTRsaeH	<a href="#">Ver</a>

### Página tareas por categoría



## Tareas

Días	Inicio	Fin	¿Terminado?	Código Categoría	Productos
34	30 de Mayo de 2015	25 de Septiembre de 2010	No	pTRsaeH	<a href="#">Ver</a>
76	19 de Marzo de 2015	5 de Julio de 2013	No	OTtzDtwm	<a href="#">Ver</a>
93	12 de Enero de 2015	24 de Febrero de 2013	No	SbTkgnTOsq	<a href="#">Ver</a>
23	16 de Julio de 2014	22 de Agosto de 2014	No	epKAKDrUt	<a href="#">Ver</a>
77	7 de Abril de 2014	10 de Abril de 2015	No	vC	<a href="#">Ver</a>

Página 1 of 2 »

### Página de tareas programadas

En este módulo se encontrarán todas las tareas asignadas de extracción para cada categoría y aquí se podrán buscar las deseadas.

## Consultar tareas

Días	Inicio	Fin	¿Terminado?	Código Categoría	Productos
34	30 de Mayo de 2015	25 de Septiembre de 2010	No	pTRsaeH	<a href="#">Ver</a>
76	19 de Marzo de 2015	5 de Julio de 2013	No	OTtzDtwm	<a href="#">Ver</a>
93	12 de Enero de 2015	24 de Febrero de 2013	No	SbTkgnTOsq	<a href="#">Ver</a>
23	16 de Julio de 2014	22 de Agosto de 2014	No	epKAKDrUt	<a href="#">Ver</a>
77	7 de Abril de 2014	10 de Abril de 2015	No	vC	<a href="#">Ver</a>

Página 1 of 2 »

### Filtro de tareas programadas

Si se requiere consultar una tarea, se podrá efectuar por medio del campo “categoría” y/o “ejecutado”, Para ver los productos involucrados en la tarea de recolección de datos podrá hacer click en la columna ver.

Abrirá una tabla con los productos involucrados en esa tarea (Productos se explicará en el siguiente ejemplo).

Productos para Cron #9 ejecutado en 2014-04-07

Código	Nombre	Categoría	Cron #	Estadisti
sEppfRdCY	Alias dolorem natus quia facere temporibus quam minus repudiandae cum laborum illum, porro architecto eius cum, delectus labore repudiandae ex nemo aliquam rem. Nobis illu	OTtzDtwm	9	<a href="#">Ver</a>
Ftcs	Impedit inventore aut unde voluptas ut nihil veritatis illum dolores. Rem porro magni, eum placeat maiores expedita quibusdam, explicabo architecto exercitationem perspicatis e	Lkvtu	9	<a href="#">Ver</a>

### Página producto por tarea programada

## Productos

Inicio Sitios Categorías Tareas **Productos** Salir

### Productos

Código:  Nombre:  Categoría:  Cron #:  ¿Ejecutado?:

Código	Nombre	Categoría	Cron #	Estadistic
sEppfRdCY	Alias dolorem natus quia facere temporibus quam minus repudiandae cum laborum illum, porro architecto eius cum, delectus labore repudiandae ex nemo aliquam rem. Nobis illu	OTtzDtwm	9	<a href="#">Ver</a>
zZZw	Asperiores placeat odio consequuntur exercitationem, voluptatibus mollitia necessitatibus a cumque s	OTtzDtwm	5	<a href="#">Ver</a>
SbYoWsPl	Dolore laudantium provident distinctio iste quibusdam in maiores sint perferendis hic natus, sequi	SbTkgnTOsq	8	<a href="#">Ver</a>
zZZvfr	Exercitationem, voluptatibus mollitia necessitatibus a cumque s	OTtzDtwm	10	<a href="#">Ver</a>
TzGK	Fuga porro sunt nihil iste dolore molestias doloremque, accusamus voluptate ea sint nulla ad rem, illo eaque volupta	qdtV	6	<a href="#">Ver</a>

Página 1 of 2 >

### Página de productos

En este módulo se encontrarán todos los productos, de igual forma podrá:

## Consultar productos y ver estadísticas

Para realizar la búsqueda puede hacerlo por medio de cualquier casilla de búsqueda. En este ejemplo se realizará por el “cron” (tarea programada), luego dar click en “buscar”.

Productos

Código:  Nombre:  Categoría:  Cron #: 5 ¿Ejecutado?

Código	Nombre	Categoría	Cron #	Estadística
sEppfRdCY	Alias dolore natus quia facere temporibus quam minus repudiandae cum laborum illum, porro architecto eius cum, delectus labore repudiandae ex nemo aliquam rem. Nobis illu	OTtzDtwm	9	<a href="#">Ver</a>
zZZw	Asperiores placeat odio consequuntur exercitationem, voluptatibus mollitia necessitatibus a cumque s	OTtzDtwm	5	<a href="#">Ver</a>
SbYoWvPI	Dolore laudantium provident distinctio iste quibusdam in maiores sint perferendis hic natus, sequi	SbTkgntOsq	8	<a href="#">Ver</a>
zZZvtr	Exercitationem, voluptatibus mollitia necessitatibus a cumque s	OTtzDtwm	10	<a href="#">Ver</a>
TzGK	Fuga porro sunt nihil iste dolore molestias doloremque, accusamus voluptate ea sint nulla ad rem, illo eaque volupta	qdtV	6	<a href="#">Ver</a>

Página 1 of 2 >

### Página filtro de búsqueda para productos

Luego de dar click en “buscar” arrojará el resultado con los respectivos datos.

Productos

Código:  Nombre:  Categoría:  Cron #: 5 ¿Ejecutado?

Código	Nombre	Categoría	Cron #	Estadística
zZZw	Asperiores placeat odio consequuntur exercitationem, voluptatibus mollitia necessitatibus a cumque s	OTtzDtwm	5	<a href="#">Ver</a>

### Página filtro de búsqueda para productos

Para ver las estadísticas del producto seleccionado, se dará click en “ver” de la columna “estadística”. Luego de haber dado click, aparecerán los datos del producto junto con los porcentajes del nivel de confianza y error, las gráficas de la cantidad vendida, cantidad disponible y el precio. Cada gráfica tiene unos puntos que determinan la fecha, lo que se calculó y lo que sucedió realmente.

## Linio Colombia

Categoría: Temporibus vitae atque laboriosam ex sint at assumenda paria

Tarea #: 5

## Asperiores placeat odio consequuntur exercitationem, voluptatibu mollitia necessitatibus a cumque s

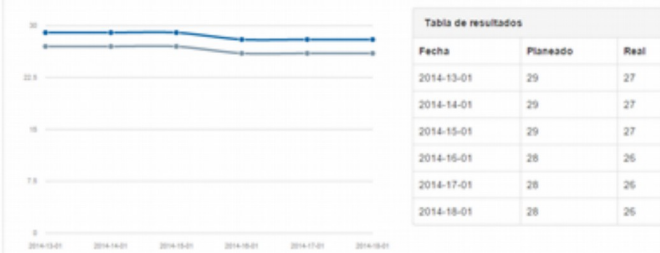
Nivel de confianza: 34%

Error: 23%

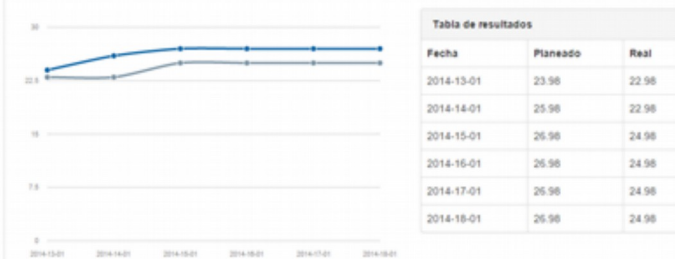
### Cantidad vendida:



### Cantidad disponible:



### Precio:



## Anexo K. Manual de usuario de la interfaz de administración

### Ingreso a la aplicación de administración

Para ir al sitio de administración de Stella debe ingresar al siguiente link <http://172.245.180.180:8000/admin>, una vez en la página, deberá hacer el respectivo “login”. Ingrese el nombre de usuario y contraseña, luego debe dar clic en el botón “iniciar sesión”.

Autenticación app de administración

Al iniciar sesión, cargará el menú del sitio administrativo.

Autenticación y autorización	
Grupos	+ Añadir / Modificar
Usuarios	+ Añadir / Modificar
Sitios	
Categories	+ Añadir / Modificar
Crons	+ Añadir / Modificar
Products	+ Añadir / Modificar
Sites	+ Añadir / Modificar

**Acciones recientes**

**Mis acciones**

- zZZvv Asperiores placeat odio consequuntur exercitationem, voluptatibus mollitia necessitatibus a cumque s OTzDtwm Product
- zZZvvtr Exercitationem, voluptatibus mollitia necessitatibus a cumque s OTzDtwm Product

Dashboard app de administración

Aquí encontrará los siguientes módulos:

#### Sitios

- Agregar sitios

- Modificar y eliminar sitios

#### Categorías

- Agregar categorías
- Modificar y eliminar categorías

#### Tareas

- Agregar tareas
- Modificar y eliminar tareas

### Agregar sitios

Para agregar un sitio, debe dar clic en “Añadir”.

Sites administracion		
Sites		
Categories	<a href="#">+ Añadir</a>	<a href="#">✎ Modificar</a>
Crons	<a href="#">+ Añadir</a>	<a href="#">✎ Modificar</a>
Products	<a href="#">+ Añadir</a>	<a href="#">✎ Modificar</a>
Sites	<a href="#">+ Añadir</a>	<a href="#">✎ Modificar</a>

Opción añadir sitio en panel de administración

Aparecerá un pequeño formulario donde indicará los datos del nuevo sitio.

Añadir site

Name:	<input type="text" value="Example#4"/>
Url:	<input type="text" value="www.example#4.com"/>
Type:	<input type="text" value="Web Scraping"/>
<input type="button" value="Grabar y añadir otro"/> <input type="button" value="Grabar y continuar editando"/> <input type="button" value="Grabar"/>	

Formulario de registro de sitios

Una vez diligenciados los campos del nuevo sitio, procederá a dar clic en el botón “Guardar”, aparecerá el siguiente mensaje.

✔ Se añadió con éxito el site "Example#4 http://www.example/#4.com".

Mensaje de confirmación al agregar un sitio

## Modificar y eliminar sitios

Para modificar los sitios dará clic en “Modificar”.

### Sites administracion

Sites	
Categories	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>
Crons	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>
Products	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>
Sites	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>

Opción modificar sitio en panel de administración

Aparecerá un listado con los sitios registrados, aquí seleccionará el que desea modificar.

### Escoja site a modificar

[Añadir site](#)

Acción:   1 de 5 seleccionado

<input type="checkbox"/>	Site
<input checked="" type="checkbox"/>	Example#4 <a href="http://www.example/#4.com">http://www.example/#4.com</a>
<input type="checkbox"/>	Linio Colombia <a href="http://www.linio.com.co/">http://www.linio.com.co/</a>
<input type="checkbox"/>	Amazon <a href="http://www.amazon.com/">http://www.amazon.com/</a>
<input type="checkbox"/>	eBay <a href="http://www.ebay.com/">http://www.ebay.com/</a>
<input type="checkbox"/>	Mercado Libre <a href="http://www.mercadolibre.com.co/">http://www.mercadolibre.com.co/</a>

5 sites

Lista de sitios registrados

Al hacer clic sobre un sitio aparecerá el formulario de edición, modificará los campos deseados y luego dará clic en el botón “Guardar”.

**Modificar site** Historico

Name:

Url: Actualmente: <http://www.example/#4.com>  
 Cambiar:

Type:

[✖ Eliminar](#)

### Formulario de edición de sitios

Si desea eliminar el sitio, simplemente realizará los pasos vistos anteriormente y dará clic en el botón “eliminar”.

### Agregar categorías

Para agregar una categoría, debe dar clic en “Añadir”.

**Sites administracion**

Sites		
Categories	<a href="#">+ Añadir</a>	<a href="#">✎ Modificar</a>
Crons	<a href="#">+ Añadir</a>	<a href="#">✎ Modificar</a>
Products	<a href="#">+ Añadir</a>	<a href="#">✎ Modificar</a>
Sites	<a href="#">+ Añadir</a>	<a href="#">✎ Modificar</a>

### Opción añadir categoría en panel de administración

Aparecerá un pequeño formulario donde indicará los datos de la nueva categoría.

### Añadir category

Code:

Name:

Site:  [+](#)

### Formulario de registro de categorías

Una vez diligenciados los campos de la nueva categoría, procederá a dar clic en el botón “Guardar”, aparecerá el siguiente mensaje.



✔ Se añadió con éxito el category "o melhor site Amazon".

Mensaje de confirmación al agregar una categoría

## Modificar y eliminar categorías

Para modificar las categorías debe dar clic en “Modificar”.

## Sites administracion

Sites	
Categories	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>
Crons	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>
Products	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>
Sites	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>

Opción modificar categoría en panel de administración

Aparecerá un listado con las categorías, aquí seleccionará la que desea modificar.

Escoja category a modificar Añadir category +

Acción:  1 de 11 seleccionado

Category
<input checked="" type="checkbox"/> o melhor site Amazon
<input type="checkbox"/> Incidunt totam quod voluptates repellendus iste corrupti beatae, laborum ipsam eum cum quam assumenda facere, eveniet officis eos dicta nost eBay
<input type="checkbox"/> Labore suscipit voluptatibus, deserunt itaque laboriosam iusto quo rem eveniet quos? Culpa dolores asperiores, aspernatur quo debitis accusamus quam itaque ipsum qui exercitationem eBay
<input type="checkbox"/> Ad iste minima qui doloribus dicta eligendi libero fugiat amet, libero porro fugiat incidunt tempora perferendis necessitatibus ab officis dolore, voluptatum mollitia illum volupta eBay
<input type="checkbox"/> Animi iusto tempore sunt quas minus a rem facilis, repudiandae numquam aliquam fug eBay
<input type="checkbox"/> Accusantium hic delectus id facere iste libero quia asperiores perferendis vel? Mollitia quidem suscipit, neque ratione nihil fugit at, voluptatum commodi molestiae tempore ex iusto vel, modi amet blanditiis assumenda expedita natus deleniti ipsa re Linio Colombia
<input type="checkbox"/> Pariatur distinctio earum incidun Mercado Libre
<input type="checkbox"/> Dignissimos repellat eius beatae, fugiat quasi quos cupiditate possimus Amazon
<input type="checkbox"/> Exercitationem atque culpa id minima possimus totam ea rerum quia velit veritatis, iusto esse nobis porro eum, id porro est odio non ullam vero consequatur iusto at. Quis quae in, laboriosam sint inventore eum hic est perferendis Mercado Libre
<input type="checkbox"/> Temporibus vitae atque laboriosam ex sint at assumenda paria Linio Colombia
<input type="checkbox"/> Qui molestiae totam iure, itaque enim aliqu eBay

11 categories

Lista de categorías registradas

Al hacer clic sobre un sitio aparecerá el formulario de edición, modificará los campos deseados y luego dará clic en el botón “Guardar”.

Modificar category Histórico

Code:

Name:

Site:

✖ Eliminar

## Formulario de edición de categorías

Si desea eliminar la categoría, simplemente realizará los pasos vistos anteriormente y dará clic en el botón “eliminar”.

## Agregar tareas

Para agregar una tarea, debe dar clic en “Añadir”.

## Sites administracion

Sites	
Categories	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>
Crons	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>
Products	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>
Sites	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>

### Opción añadir una tarea en panel de administración

Aparecerá un pequeño formulario donde indicará los datos del nueva tarea.

#### Añadir cron

Days:	<input type="text" value="9"/>
Start on:	<input type="text" value="05/01/2015"/> Hoy
Nota: Usted esta a 5 horas detrás de la hora del servidor.	
Finish on:	<input type="text" value="06/01/2015"/> Hoy
Nota: Usted esta a 5 horas detrás de la hora del servidor.	
Script:	<input type="text" value="ejemplo#"/>
<input type="checkbox"/> Executed	
Category:	<input type="text" value="Incidunt totam quod voluptates repellendus iste corrupti beatae, laborum ipsam eum cum quam assumenda facere, eveniet officiis eos dicta nost eBay"/>
<input type="button" value="Grabar y añadir otro"/> <input type="button" value="Grabar y continuar editando"/> <input type="button" value="Grabar"/>	

### Formulario de registro de tareas

Una vez diligenciados los campos de la nueva tarea, procederá a dar clic en el botón “Guardar”, aparecerá el siguiente mensaje.

✔ Se añadió con éxito el cron "Incidunt totam quod voluptates repellendus iste corrupti beatae, laborum ipsam eum cum quam assumenda facere, eveniet officiis eos dicta nost eBay 9 2015-01-05 qdTV".

### Mensaje de confirmación al agregar una tarea

## Modificar y eliminar tareas

Para modificar las tareas dará clic en “Modificar”.

### Sites administracion

Sites	
Categories	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>
Crons	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>
Products	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>
Sites	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>

Opción modificar tarea en panel de administración

Aparecerá un listado con las tareas, aquí seleccionará el que desea modificar.

Escoja cron a modificar Añadir cron +

Acción:  1 de 11 seleccionado

<input type="checkbox"/>	Cron
<input checked="" type="checkbox"/>	Incidunt totam quod voluptates repellendus iste corrupti beatae, laborum ipsam eum cum quam assumenda facere, eveniet officiis eos dicta nost eBay 9 2015-01-05 qdtv
<input type="checkbox"/>	Exercitationem atque culpa id minima possimus totam ea rerum quia velit veritatis, iusto esse nobis porro eum, id porro est odio non ullam vero consequatur iusto at. Quis quae in, laboriosam sint inventore eum hic est perferendis Mercado Libre 31 2014-04-03 Lkvtu
<input type="checkbox"/>	Labore suscipit voluptatibus, deserunt itaque laboriosam iusto quo rem eveniet quos? Culpa dolores asperiores, aspernatur quo debitis accusamus quam itaque ipsum qui exercitationem eBay 77 2014-04-07 vC
<input type="checkbox"/>	Animi iusto tempore sunt quas minus a rem facilis, repudiandae numquam aliquam fug eBay 34 2015-05-30 pTRsEh
<input type="checkbox"/>	Exercitationem atque culpa id minima possimus totam ea rerum quia velit veritatis, iusto esse nobis porro eum, id porro est odio non ullam vero consequatur iusto at. Quis quae in, laboriosam sint inventore eum hic est perferendis Mercado Libre 120 2013-08-28 Lkvtu
<input type="checkbox"/>	Accusantium hic delectus id facere iste libero quia asperiores perferendis vel? Mollitia quidem suscipit, neque ratione nihil fugit at, voluptatum commodi molestiae tempore ex iusto vel, modi amet blanditiis assumenda expedita natus deleniti ipsa re Linio Colombia 23 2014-07-16 epKakDrUt
<input type="checkbox"/>	Temporibus vitae atque laboriosam ex sint at assumenda paria Linio Colombia 83 2013-09-29 OTzDtwm
<input type="checkbox"/>	Temporibus vitae atque laboriosam ex sint at assumenda paria Linio Colombia 76 2015-03-19 OTzDtwm
<input type="checkbox"/>	Accusantium hic delectus id facere iste libero quia asperiores perferendis vel? Mollitia quidem suscipit, neque ratione nihil fugit at, voluptatum commodi molestiae tempore ex iusto vel, modi amet blanditiis assumenda expedita natus deleniti ipsa re Linio Colombia 35 2013-05-21 epKakDrUt
<input type="checkbox"/>	Incidunt totam quod voluptates repellendus iste corrupti beatae, laborum ipsam eum cum quam assumenda facere, eveniet officiis eos dicta nost eBay 17 2012-03-06 qdtv
<input type="checkbox"/>	Ad iste minima qui doloribus dicta eligendi libero fugiat amet, libero porro fugiat incidunt tempora perferendis necessitatibus ab officiis dolore, voluptatum mollitia illum volupta eBay 93 2015-01-12 SbtKgnT0sq

11 crons

Lista de tareas registradas

Al hacer clic sobre un sitio aparecerá el formulario de edición, en dónde modificará los campos deseados y luego dará clic en el botón “Guardar”.

Modificar cron Eliminar

Days:

Start on:  Hoy |

Nota: Usted esta a 5 horas detrás de la hora del servidor.

Finish on:  Hoy |

Nota: Usted esta a 5 horas detrás de la hora del servidor.

Script:

Executed

Category:

Eliminar Grabar y añadir otro Grabar y continuar editando Grabar

Formulario de edición de tareas

Si desea eliminar una tarea, simplemente realizará los pasos vistos anteriormente y dará clic en el botón “eliminar”.