



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE INGENIERÍA

**Sistema de análisis de grandes
volúmenes de datos data bajo
metodología ágil**

TESIS

Que para obtener el título de
Ingeniero en Computación

P R E S E N T A

López Araiza Guevara Rodrigo

DIRECTOR DE TESIS:

Dr. Daniel Trejo Medina



Ciudad Universitaria, Cd. Mx., 2017



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

RESUMEN

Existen diversas aplicaciones del denominado *Big Data* que trabajan en distintas áreas de conocimiento, algunas de ellas presentando análisis de información y datos sobre el comportamiento de su entorno. Una de éstas áreas es la mercadotecnia que se encarga de formar y mantener relaciones entre un negocio y sus consumidores.

Se investigó, desarrolló y probó una aplicación de *software* con metodologías ágiles para el análisis de datos estructurados y no estructurados en grandes volúmenes para determinar escenarios aplicados a mercadotecnia digital, misma que tiene aplicación al control de campañas o tasas de abandono de navegación.

La solución desarrollada es capaz de analizar las estadísticas *web* básicas que permiten obtener indicadores sobre el resultado de la exposición de un negocio en internet a través de un sitio que funcione como canal de comunicación con sus consumidores.

Durante el desarrollo se identificaron áreas de oportunidad y brechas entre los conocimientos adquiridos en la Facultad de Ingeniería de la Universidad Nacional Autónoma de México y las habilidades básicas requeridas para laborar en la industria de desarrollo de soluciones de *software* de manera adecuada.

ABSTRACT

Nowadays there are a lot of applications of the so called Big Data that work on different areas, some of them present some analysis on information and data about the behavior of their environment. One of these areas is marketing, which focuses on creating and maintaining relationships between a business and its consumers.

We researched, developed and tested a software solution following agile methodologies to obtain the analysis of structured and unstructured data on large volume sets to determine scenarios related to digital marketing, which has an application on the control of campaigns and navigation abandonment rates.

The developed solution is capable of getting the basic web analytics that provide indicators about the exposure of a business on the internet through a web site that works as a communication channel with its customers.

During the development we identified areas of opportunity and gaps between the knowledge acquired on the Facultad de Ingeniería y of the Universidad Nacional Autónoma de México and the basic skill set required to work in the software development industry.

AGRADECIMIENTOS

Este trabajo está dedicado a mis padres y a mi hermano Mauricio quienes además de creer en mí me han apoyado y empujado a lograr los objetivos que me propongo.

Agradezco también a mis amigos que jamás permiten que me dé por vencido cuando las circunstancias son adversas.

De mis profesores aprecio el conocimiento que impartieron, su paciencia y esfuerzo por impulsarme a desarrollar mis habilidades y a crecer como persona, especialmente al Dr. Daniel Trejo Medina que sin su ayuda este trabajo no habría sido posible.

Adicionalmente, reconozco el apoyo de la empresa que colaboró con la identificación del caso de estudio.

ÍNDICE

RESUMEN	III
ABSTRACT	IV
AGRADECIMIENTOS	V
ÍNDICE	VII
ÍNDICE DE ILUSTRACIONES	IX
ÍNDICE DE TABLAS	X
INTRODUCCIÓN	1
PROBLEMA A RESOLVER	3
TIPO DE ESTUDIO (PROPUESTA METODOLÓGICA)	3
OBJETIVO GENERAL	4
OBJETIVOS PARTICULARES	4
JUSTIFICACIÓN	4
HIPÓTESIS	5
MARCO DE REFERENCIA	6
MATERIAS RELEVANTES	6
<i>Computación para ingenieros</i>	6
<i>Programación avanzada y métodos numéricos</i>	7
<i>Algoritmos y estructura de datos</i>	9
<i>Probabilidad y estadística</i>	10
<i>Redacción y exposición de temas de Ingeniería</i>	12
<i>Sistemas operativos</i>	14
<i>Ingeniería de software</i>	15
<i>Lenguajes formales y autómatas</i>	17
<i>Lenguajes de programación</i>	19
<i>Ética profesional</i>	21

<i>Bases de datos</i>	22
<i>Administración de proyectos de software</i>	24
<i>Redes de datos</i>	25
<i>Administración de redes</i>	27
<i>Introducción al análisis económico empresarial</i>	28
<i>Negocios electrónicos</i>	30
<i>Verificación y validación de software</i>	31
CÓMPUTO EN LA CIENCIA A TRAVÉS DEL TIEMPO	35
¿CÓMO FUE LA SITUACIÓN EN MÉXICO?	36
HISTORIA DEL DISCO DURO	38
ALMACENAMIENTO LÓGICO DE LOS DATOS: ARCHIVOS, BASES DE DATOS, <i>DATA WAREHOUSE</i> Y <i>NOSQL</i>	40
ANÁLISIS DE DATOS	43
<i>GOPHER, ARCHIE, VERONICA Y JUGHEAD</i>	46
LA <i>WEB</i> Y LOS BUSCADORES	49
<i>BIG DATA</i>	52
<i>MERCADOTECHNIA DIGITAL</i>	56
ESTADÍSTICAS Y MÉTRICAS PARA SITIOS <i>WEB</i>	62
METODOLOGÍAS ÁGILES	66
DESARROLLO	73
SPRINT 1	76
SPRINT 2	78
RESULTADOS	81
ENTREGABLES DEL <i>SPRINT 1</i>	81
ENTREGABLES DEL <i>SPRINT 2</i>	84
CONCLUSIONES	92
FUENTES	98

ÍNDICE DE ILUSTRACIONES

Ilustración 1 Trejo, D. (2013). Modelo de 8 Vs de Big Data. Fuente: Apuntes sobre Big Data.....	54
Ilustración 2 Kotler, P. y Armstrong, G. (2013). Sistema de marketing. Fuente: Fundamentos de marketing.....	58
Ilustración 3 Satpathy, T. et. al. (2016). Flujo del Scrum. Fuente: SBOK.....	70
Ilustración 4 Burn up de complejidad estimada del sprint 1. Fuente: elaboración del autor.....	77
Ilustración 5 Burn up de complejidad estimada del sprint 2. Fuente: elaboración del autor.....	79
Ilustración 6 Arquitectura de la solución. Fuente: elaboración del autor.....	81
Ilustración 7 Administración de sitios. Fuente: elaboración del autor.....	82
Ilustración 8 Analíticos principales. Fuente: elaboración del autor.....	83
Ilustración 9 Gráfica de referencias al sitio. Fuente: elaboración del autor.....	83
Ilustración 10 Burn up de estimación vs ejecución del sprint 1. Fuente: elaboración del autor.....	84
Ilustración 11 Gráfica de flujos de usuario. Fuente: elaboración del autor.....	85
Ilustración 12 Administración de usuarios. Fuente: elaboración del autor.....	85
Ilustración 13 Formulario para alta de usuarios. Fuente: elaboración del autor.....	86
Ilustración 14 Agregar usuario a un grupo de trabajo. Fuente: elaboración del autor.....	87
Ilustración 15 Administración de grupos. Fuente: elaboración del autor.....	87
Ilustración 16 Formulario de grupo de trabajo. Fuente: elaboración del autor.....	88
Ilustración 17 Agregar sitios al grupo de trabajo. Fuente: elaboración del autor.....	89
Ilustración 18 Agregar usuarios al grupo de trabajo. Fuente: elaboración del autor..	90

Ilustración 19 Pantalla de ingreso al sistema. Fuente: elaboración del autor..... 90

Ilustración 20 Segmentación de visitantes al sitio. Fuente: elaboración del autor..... 91

Ilustración 21 Burn up de estimación vs ejecución del sprint 2. Fuente: elaboración del autor. 91

ÍNDICE DE TABLAS

Tabla 1 Lista priorizada de epics del proyecto. Fuente: elaboración del autor. 74

Tabla 2 Definición de personas. Fuente: elaboración del autor..... 75

Tabla 3 Criterios de terminado del proyecto. Fuente: elaboración del autor. 75

Tabla 4 Historias de usuario del sprint 1. Fuente: elaboración del autor..... 77

Tabla 5 Lecciones aprendidas del sprint 1. Fuente: elaboración del autor..... 78

Tabla 6 Historias de usuario del sprint 2. Fuente: elaboración del autor..... 79

Tabla 7 Lecciones aprendidas del sprint 2. Fuente: elaboración del autor..... 80

INTRODUCCIÓN

El análisis masivo de datos se ha convertido en un estándar adoptado por las empresas líderes en cualquier sector económico y particularmente en mercadotecnia es una ventaja competitiva que permite tomar decisiones oportunas en beneficio de las estrategias de venta.

La toma de decisiones efectivas es parte fundamental de las acciones que debe ejecutar una empresa para mantenerse con vida y gracias a las soluciones actuales de *Big Data* es posible adaptarse al cambio de manera eficiente y en algunos casos de forma anticipada.

Poco a poco el análisis mencionado ha generado que los servicios se personalicen y se tenga identificado claramente cuál es el segmento meta y cómo evoluciona el consumo de los productos ofertados; logrando obtener resultados que aporten valor a las empresas en tiempo real, además de generar, en algunos casos, un modelo de pronóstico que permite prever eventualidades.

Durante el desarrollo de este trabajo se pretende explorar un sector particular en el análisis de datos: los analíticos *web*, pues parte de la forma actual en que se realiza mercadotecnia comprende el darse a conocer a través de un sitio *web* en internet, la fuente de información con mayor alcance y con un mercado potencialmente inmenso.

Principalmente se busca poder apreciar el comportamiento que tienen los usuarios dentro de un sitio *web* e identificar si se está logrando el alcance previsto en el modelo de negocio, con base en **indicadores cuantificables** que expliquen de manera sencilla el éxito o fracaso de la exposición de un negocio en internet.

La solución propuesta fue desarrollada trabajando bajo el esquema de *Scrum*, que es una metodología ágil para el desarrollo de proyectos, garantizando así la mayor satisfacción del cliente.

Más adelante se explicará con mayor detalle las ventajas que ofrece utilizar una metodología ágil en comparación con un desarrollo tradicional en cascada para

analizar por qué existe actualmente una tendencia a utilizar este tipo de metodologías en el desarrollo de soluciones a la medida

PROBLEMA A RESOLVER

El análisis masivo de datos en escenarios de mercadotecnia es cada vez mayor y más utilizado, sin embargo, el poder generar modelos predictivos con base en datos multicanal lo hace bastante complicado, la tesis busca solucionar el análisis de datos y pronosticar posibles comportamientos al vuelo.

TIPO DE ESTUDIO (PROPUESTA METODOLÓGICA)

Esta investigación se plantea basada en la metodología deductiva que utiliza hechos generales comprobados para establecer una hipótesis sobre situaciones particulares.

Como menciona Mario Tamayo en su obra *El Proyecto de Investigación* (2004), "... El procedimiento que consiste en desarrollar una teoría empezando por formular sus puntos de partida o hipótesis básicas y deduciendo luego sus consecuencias con la ayuda de las subyacentes teorías formales se llama método hipotético-deductivo".

De esta forma, el proyecto de investigación incluye los Elementos del Proceso Investigativo que Tamayo propone en el capítulo 1 de esta misma obra como son los antecedentes del problema o caso de estudio, el problema de la investigación, los objetivos, un marco teórico y las hipótesis que plantearemos para nuestro proyecto.

En nuestro caso plantearemos una hipótesis basada en el éxito que en la actualidad tienen la mayoría de los sistemas de información que son implementados con la finalidad de optimizar procesos y reducir costos de operación en las instituciones.

OBJETIVO GENERAL

Aplicar conceptos de Ingeniería de *software* y matemáticas para el análisis de información sobre grandes volúmenes de datos bajo el esquema de desarrollo de una metodología ágil.

Objetivos particulares

1. Desarrollar una solución que aporte valor al negocio de mercadotecnia digital.
2. Aplicar un desarrollo ágil y satisfacer de manera oportuna las necesidades de negocio detectadas.
3. Mostrar el valor ofertado por la solución desde el inicio del proyecto.

JUSTIFICACIÓN

Existe un campo de oportunidades muy extenso para aplicar soluciones de *Big Data*, ya que existen innumerables canales de información que se pueden aprovechar. A través del presente se busca atacar alguna de estas áreas desarrollando una solución bajo un esquema de metodología ágil.

A partir de la investigación en aplicaciones de soluciones de *Big Data*, se identificó junto con el director de la presente que un caso de estudio poco desarrollado es sobre *mercadotecnia* digital.

Se justifica el desarrollo de una solución de *Big Data* para aportar información valiosa a estrategias de *marketing*, particularmente la publicación de un sitio *web* como canal de comunicación y el éxito o fracaso de la exposición de un negocio en internet.

Se aplicaron conocimientos de distintas materias cursadas a lo largo de la carrera en la Facultad de Ingeniería sobre todo materias relacionadas con Ingeniería de *Software*, negocios electrónicos, probabilidad y estadística, por mencionar algunas.

Esta investigación aporta conocimiento a la universidad en la adecuación de una metodología ágil al desarrollo de un proyecto que analice grandes volúmenes de datos, lo cual aporta difusión de áreas emergentes de conocimiento interdisciplinario (Mercadotecnia e Ingeniería de Software).

HIPÓTESIS

H1: Si se desarrolla un sistema de análisis de grandes volúmenes de datos bajo metodologías ágiles, se podrá tener una mejor respuesta al control de cambios durante el desarrollo del sistema.

MARCO DE REFERENCIA

De acuerdo con el Plan de Estudios del año 2010 de la Carrera de Ingeniería en Computación (Facultad de Ingeniería. Universidad Nacional Autónoma de México (UNAM)., 2010), se determina que las materias más relevantes para el desarrollo de esta tesis son las descritas a continuación.

Materias relevantes

Computación para ingenieros

Tiene por objetivo conocer la importancia de la computación e informática como herramienta para el desempeño académico y profesional de Ingeniería. Empleando el software básico que permita generar productos que resuelvan problemas matemáticos y de Ingeniería.

En esta materia se revisaron conceptos básicos relacionados con computación; qué es una computadora, cómo funcionan, qué aplicaciones de la vida real tiene el hacer uso del cómputo, cómo aprovechar los recursos para resolver problemas y las bases para desarrollar *software*.

En otras palabras, es el primer acercamiento que se tiene al uso de la tecnología como herramienta para resolver problemas.

Lo anterior está distribuido en dos secciones: una teórica y una práctica. En la teoría se exploran de manera abstracta los conceptos y en la práctica se aplicaron para resolver problemas de distintos tipos.

La teoría sienta un marco de referencia de conocimientos sobre el cómputo y cómo la Ingeniería ha aprovechado el desarrollo tecnológico para lograr un progreso social.

Considero que las prácticas resueltas en la materia contribuyeron a un mejor entendimiento de la forma en que debe diseñarse y desarrollarse un programa lo que me permitió enfrentar retos más complejos.

Con base en lo anterior pienso que esta materia es fundamental en el desarrollo profesional de un ingeniero en computación, ya que establece un punto de partida para aprender a aprovechar recursos de cómputo en la solución de diversos problemas.

El plan de la materia queda limitado a lo más básico de cada tema, sin embargo se complementa con materias más avanzadas, las cuales profundizan sobre los conceptos principales comentados en esta clase.

Programación avanzada y métodos numéricos

Tiene por objetivo emplear estructuras de almacenamiento de datos complejas para la resolución de problemas numéricos. Seleccionar y aplicar métodos numéricos para obtener soluciones aproximadas de modelos matemáticos que no se pueden resolver por métodos analíticos; y desarrollar programas tanto en lenguaje estructurado como orientado a objetos que implementen dichos métodos numéricos.

Los temas vistos están relacionados con los fundamentos de la programación. Principalmente se revisan dos paradigmas: programación estructurada y programación orientada a objetos.

Adicionalmente y como práctica de los conocimientos adquiridos en programación, se comentan temas sobre métodos numéricos, que sirven para aproximar una solución a problemas matemáticos cuya respuesta exacta no se puede calcular de manera tradicional.

Para ejemplificar lo anterior, tomemos el caso de las raíces cuadradas, algunas de ellas no tienen una solución exacta que pueda calcularse de manera directa como es el caso de la raíz cuadrada del número 2; no existe un número decimal no periódico que elevado al cuadrado dé como resultado el número 2.

En la materia se explica cómo manejar el margen de error obtenido al realizar aproximaciones, pues dependiendo del tipo de problema que se trate de resolver el error permisible puede ser mayor o menor.

Los métodos numéricos están basados en procesos repetitivos que fácilmente pueden ser programados para que una computadora los resuelva.

Aprovechando esto, en la materia se explican conceptos básicos de la programación, por ejemplo los tipos de datos, las estructuras, los ciclos iterativos y las operaciones básicas, entre otros.

Para el análisis de datos numéricos esta materia aporta el conocimiento suficiente que permite aplicar aproximaciones de métodos numéricos a modelos matemáticos que no tienen solución eficiente de manera analítica.

El plan de la materia está bien estructurado pues retoma conceptos de programación de la asignatura *Computación para ingenieros* y los aplica a problemas que comienzan a relacionarse directamente con la Ingeniería.

Considero que esta es otra de las materias fundamentales que ha aportado mucho valor en mi desarrollo profesional, pues constituye el siguiente escalón en la formación de un profesional capaz de aprovechar la tecnología para resolver problemas que involucren cálculos matemáticos.

Con respecto a los paradigmas mencionados, me parece correcto ir de lo estructurado a la programación orientada a objetos. Estos dos paradigmas conforman la base del enfoque más popular utilizado en la industria.

La mayoría de la oferta laboral actual (año 2017) demanda profesionales que dominen lenguajes de programación que están principalmente orientados a objetos, algunos ejemplos son C# y Java.

Particularmente en el curso que me tocó, las prácticas fueron desarrolladas con lenguaje Java, uno de los más comunes entre las empresas que se dedican al desarrollo de soluciones de *software*.

El programa de la materia fue bastante completo y gracias a los temas sobre aproximación numérica se logró un aprendizaje muy práctico sobre la forma en la que pueden procesarse los datos a través de un *software*.

Esta materia ha resultado muy útil en mi desarrollo profesional pues definió una línea base aprovechando los antecedentes de *Computación para ingenieros*. Gracias a ella fui capaz resolver problemas matemáticos tan básicos como una suma o tan complejos como la aproximación en la solución de un sistema de ecuaciones.

Algoritmos y estructura de datos

Tiene por objetivo resolver problemas de almacenamiento, recuperación y ordenamiento de datos, utilizando las estructuras para representarlos y las técnicas de operación más eficientes; así mismo analizar y diseñar algoritmos.

El diseño de algoritmos es otro aspecto fundamental en el ambiente profesional de un ingeniero en computación. Los procesos que ejecutan las soluciones desarrolladas se modelan a través de algoritmos y éstos a su vez emplean estructuras de datos que permiten ordenamientos.

Como parte del plan curricular, esta materia analiza distintos casos que se modelan a través de algoritmos para cumplir con ciertas reglas de negocio, con el fin de ejecutar un proceso con ayuda de una computadora.

Para el desarrollo de soluciones es muy importante que un ingeniero tenga la posibilidad de modelar un caso de negocio, pues a través del algoritmo diseñado un proceso puede ser ejecutado más rápido y con un menor margen de error.

En otras palabras, un algoritmo que esté bien diseñado permite implementar tecnologías que faciliten la ejecución de dicho proceso.

Por otro lado, la materia analiza la forma en que una computadora puede manejar los datos que se utilizan para calcular algún resultado. La forma en que se modela un proceso involucra el manejo de datos, los cuales pueden ser números, palabras, caracteres o *booleanos* entre otros.

Lo anterior se estudia a profundidad con el fin de poder diseñar estructuras más complejas que permitan cumplir con un proceso diseñado. Por ejemplo, para modelar un sistema de calificaciones escolares debe considerarse la forma en la que se almacenarán los datos de los alumnos, además de separar esos alumnos en grupos.

Otro de los objetivos de poder definir la estructura en la que el *software* ha de manejar la información es la capacidad de realizar ordenamientos sobre los datos.

Tomando como ejemplo el sistema de calificaciones, normalmente un grupo de alumnos está ordenado alfabéticamente con respecto al apellido paterno. Estos ordenamientos permiten a su vez realizar búsquedas eficientes.

La materia también sienta las bases para analizar la complejidad computacional de resolver un problema aplicando un cierto algoritmo.

Es importante que un profesional dedicado a desarrollar *software* considere la complejidad que implica aplicar algún algoritmo con el fin de decidir cuál puede resolverlo de manera más eficiente, ya que uno de los mayores beneficios de utilizar un *software* es que una computadora puede ejecutar el proceso más rápidamente que una persona.

En el campo laboral esta asignatura es particularmente una de las que más aporta a la formación profesional, pues sienta las bases para desarrollar soluciones de calidad que satisfacen de la mejor manera las necesidades de los clientes.

Probabilidad y estadística

Tiene por objetivo aplicar los conceptos y la metodología básicos de la teoría de la probabilidad y la estadística, para analizar algunos experimentos aleatorios que ocurren en la naturaleza y la sociedad, resaltando los correspondientes a la Ingeniería.

La estadística es la base para el análisis de datos. La estadística se encarga, entre otras cuestiones, de describir fenómenos con base en el análisis de los datos que éstos producen.

Uno de los módulos de especialización de la carrera es el de bases de datos. En él, los ingenieros optan por dedicarse al desarrollo de las estructuras que puedan manejar de la mejor manera los datos que permiten aplicar tecnologías de la información a procesos de negocio.

Con base en lo anterior, la estadística ofrece un área de oportunidad muy importante, pues puede aplicarse a los datos que se están analizando y así encontrar nuevas oportunidades.

Algunos de los temas tratados en el curso se dedican a explicar cómo se debe utilizar la estadística y de qué forma nos ayuda a entender el comportamiento de ciertos fenómenos. En esta primera parte se explica cuáles son los indicadores básicos que pueden obtenerse al analizar un conjunto de datos muestrales.

La materia trata también temas relacionados con la probabilidad, la cual básicamente se enfoca a encontrar la frecuencia con la que ocurre algún evento. Esto sirve principalmente para saber cuál de los posibles resultados tiene mayor tendencia a ocurrir en caso de realizar alguna acción.

Con base en lo anterior, se ha observado que esta asignatura es importante en el ámbito laboral ya que otorga un conocimiento base que permite la comunicación con analistas de datos, los cuales tienen una necesidad de soluciones de *software* que sean capaces de calcular datos lo más rápidamente posible y así obtener información de manera oportuna.

Por ejemplo, para medir la calidad del aire se pueden analizar distintas variables que influyen en la cantidad de partículas contaminantes que permanecen suspendidas en el ambiente.

Al analizar dichas variables se puede obtener un estudio sobre qué escenarios puede ocasionar el si se toman acciones que repercuten directamente sobre esas variables.

Para este ejemplo tomemos el caso del medio de transporte que utilizan las personas para desplazarse en la ciudad, esta decisión influye directamente en la cantidad de contaminantes que se generan por los vehículos que utilizan combustible y puede pronosticarse el escenario ideal para conservar un ambiente limpio.

Con base en la experiencia laboral, he observado que el plan de la carrera es suficiente para abordar los problemas básicos, sin embargo se propone considerar un plan que abarque problemas más complejos relacionados con pronóstico en el comportamiento de datos.

Los temas sobre pronóstico no son indispensables en todos los módulos de especialización pero es un aspecto que aportaría mucho valor al área de análisis de datos.

Los negocios actuales están muy interesados en el análisis de los datos para obtener información que permita tomar decisiones a futuro.

Dicho en otras palabras, los directores de empresas están buscando constantemente una solución que realice pronósticos confiables para tomar mejores decisiones con respecto a sus negocios.

Es importante que la Facultad esté actualizada en cuál es la demanda actual en la industria con el fin de educar líderes que cubran las necesidades del mercado actual.

Es por esto que esta materia se considera muy importante para un ingeniero en computación, pues muestra las oportunidades de un área de especialización que actualmente es de mucha utilidad para la sociedad.

Redacción y exposición de temas de Ingeniería

Tiene por objetivo mejorar la competencia en el uso de la lengua, a través del desarrollo de capacidades de comunicación en forma oral y escrita. Valorar también la importancia de la expresión oral y de la redacción en la vida escolar y en la práctica profesional. Al final del curso, habrá ejercitado habilidades de estructuración y desarrollo de exposiciones orales y redacción de textos sobre temas de Ingeniería.

El valor obtenido de esta materia radica principalmente en la habilidad de transmitir el conocimiento sobre proyectos desarrollados. Uno de los principales retos que se tiene y que además representa un aspecto fundamental en el desarrollo de proyectos de Ingeniería es la capacidad de expresar cuáles son las ventajas competitivas de un producto desarrollado de tal forma que el cliente vea el valor que obtiene al utilizarlo, además de despertar el interés en usuarios potenciales.

Lo anterior está basado en que normalmente los ingenieros desarrollan las presentaciones con base en características técnicas y por lo tanto mantienen el lenguaje de la conversación a un nivel que puede llegar a ser incomprensible y aburrido para el cliente.

Para solucionar esto, en esta materia se explora la forma en que una persona debe preparar y ejecutar una exposición, independientemente del tema a tratar.

Los temas relacionados con expresión oral que se estudian son la base que se tiene para poder realizar presentaciones y mantener el interés del público. Adicionalmente se explica la forma en que debe seleccionarse el tipo de lenguaje dependiendo del público que estará presente en dicha exposición.

Para un público que esté dedicado a dirigir un negocio, es esencial mantener un lenguaje con términos de negocio y describir las características de tal forma que se muestre el valor que genera en el negocio utilizar un producto.

Otro aspecto a tratar en la materia es la capacidad de generar textos técnicos en los cuales se exponga algún tema relacionado con Ingeniería.

Esta parte es de mucha ayuda para el ingeniero que ha de dedicarse a la investigación, pues implica explicar algún concepto o tecnología en términos que sean entendidos por algún otro profesional en el área.

En el campo laboral, los temas de redacción son muy útiles pues permiten la comunicación escrita entre elementos de una compañía.

Por ejemplo para redactar un informe al director del área, debe emplearse una correcta redacción que sea entendible por la persona que lo leerá. Otro caso sería el generar un manual de usuario, el cual puede facilitar el uso de un producto y por lo tanto el nivel de satisfacción de los consumidores.

Esta materia también implica sesiones dedicadas a la comprensión de lectura, las cuales son muy útiles cuando se quiere entender la idea principal de un escrito con el fin de determinar si será útil leerlo por completo.

La comprensión de lectura también facilita la comunicación entre colegas de una empresa, un ejemplo muy claro puede ser la documentación de una solución previamente desarrollada.

Con un buen nivel de comprensión de lectura puede entenderse más fácilmente la documentación de artefactos heredados de otros colaboradores que podrían o no continuar en la misma empresa.

Por último cabe recalcar que esta materia ha resultado muy útil para el desarrollo del presente trabajo, pues de ella se obtuvieron las herramientas básicas que han permitido una expresión clara de las ideas desarrolladas

Sistemas operativos

Tiene por objetivo obtener las bases para administrar un sistema operativo, así como diseñar y desarrollar *software* operativo.

En esta materia se estudiaron varios temas que explican el funcionamiento de los sistemas operativos y de qué manera aprovechan los recursos de una computadora.

Un sistema operativo es el *software* principal que permite interactuar con los componentes de una computadora para realizar distintas tareas, las cuales son ejecutadas a través de distintos programas.

En esta materia se analiza qué recursos están disponibles en una computadora y que permiten el correcto funcionamiento de los programas que se utilizan.

Se observa a detalle de qué manera se administran estos recursos con el fin de que los procesos puedan ejecutarse de la manera más eficiente posible.

Es importante conocer las capacidades de los procesos que se manejan en un sistema operativo de acuerdo con los recursos físicos de las computadoras porque sabiendo esto es posible como ingenieros desarrollar una solución que se adecúe al ambiente.

Por otro lado, también es muy útil conocerlo porque de esta manera se pueden plantear cuáles son los requerimientos mínimos que requerirá una computadora para ejecutar de manera adecuada un cierto *software* desarrollado.

Es muy necesario que se puedan identificar cuáles son los requerimientos mínimos de ejecución de un producto para que los consumidores estén conscientes de las características que deben considerar para poder utilizar dicho producto.

Otro aspecto que resulta muy útil es el estudio de la forma en que un sistema operativo administra el espacio de memoria de la computadora.

Con un buen conocimiento básico sobre esto es posible diseñar *software* que no cause conflictos con otros procesos ejecutados concurrentemente en la misma computadora.

Esta materia ha resultado muy útil en el campo laboral, pues gracias a ella se conoce la forma en que una computadora funciona a un nivel muy bajo y por lo tanto se conocen diferentes ambientes para desarrollar *software*.

De los temas tratados se observa que la asignatura tiene un enfoque muy técnico sobre el conocimiento de los sistemas operativos, se sugiere que se incluyan algunos temas en los que se analicen de manera comparativa los sistemas operativos más utilizados.

Otra razón por la cual esta materia es muy relevante, es que abre una oportunidad muy grande de negocio: el poder crear un sistema operativo propio.

Parte del objetivo de la materia establece que el alumno podrá diseñar y desarrollar un sistema operativo que funcione correctamente.

Lo anterior resulta demasiado ambicioso en el tiempo que está planeado para completar el curso, sin embargo la materia brinda los conocimientos suficientes para que un ingeniero interesado en el tema logre ejecutarlo.

Ingeniería de *software*

Tiene por objetivo aplicar en un proyecto procesos y herramientas mediante las cuales se analice diseño e implemente un proyecto de *software*.

Esta materia fue muy útil como introducción al campo laboral del módulo de *software* para la carrera. Se explicaron cuáles son los aspectos básicos que componen el desarrollo de un proyecto de *software*.

Estos temas son actividades que realiza un ingeniero en computación en el día a día y que permiten el desarrollo de soluciones asegurando un alto nivel de calidad.

El proceso que se sigue para el desarrollo de soluciones, sin importar la metodología que se esté siguiendo, se compone de cinco aspectos fundamentales: requerimientos, diseño, implementación, pruebas y corrección de defectos.

Durante el curso de la materia se revisaron con detalle cada uno de estos aspectos y se adaptó el proceso a una metodología de desarrollo en cascada.

También se revisó brevemente cuál es la tendencia de metodologías de desarrollo que se utilizan en la industria, sin embargo se considera oportuno planear más tiempo para estudiar estos temas pues se sigue enseñando un enfoque poco actualizado.

Dicho en otras palabras, actualmente se enseña cómo adaptar un desarrollo de *software* al proceso en cascada, sin embargo la mayoría de las empresas no utiliza dicha metodología pues se ha conseguido una mejor satisfacción de clientes aplicando metodologías de tipo ágil.

El plan de la materia no contempla el estudio del enfoque ágil para el desarrollo de soluciones, se sugiere dedicar más horas a analizar cuáles son las metodologías más utilizadas y comparar en qué casos resulta más conveniente optar por una o por otra.

Un tema muy importante que se discute, es el manejo de los cambios en las especificaciones iniciales del proyecto ya que se ha observado en el campo laboral que todos los proyectos sufren cambios.

Con base en lo anterior se refuerza la idea de incluir un tiempo dedicado al estudio de metodologías ágiles, pues la administración de cambios es más eficiente en este tipo de metodologías.

Por otro lado, también se considera que la materia se desarrolla de manera muy teórica.

En general se abarcan temas que son difíciles de adaptar a un ejemplo práctico porque lograr esto significaría que los alumnos siguieran un proceso ya establecido.

Sin embargo lo importante no es aprender un proceso en específico sino las bases y los aspectos fundamentales que lo componen para que una vez cursada la materia, el alumno sea capaz de adaptarse.

Las empresas en el campo laboral tienen un proceso definido, pero no existen dos empresas con exactamente el mismo proceso y es por esto que no resulta práctico

seguir un único proceso, el mejor enfoque es entender bien lo básico para lograr adaptarse a cualquier empresa que ya tenga algo definido.

Incluso podría llegar a diseñarse un proceso que permita un buen desarrollo si se consideran los aspectos fundamentales que deben componerlo.

A pesar de todo, se sugiere también incluir un taller más práctico que esté enfocado en la aplicación de los conocimientos teóricos.

Este taller puede reforzar en gran medida los conocimientos que se estudian en la materia.

Durante el curso de la materia se realizó el desarrollo de un proyecto a lo largo del semestre y cada fase del proceso se debía ejecutar a la par de la explicación teórica de los temas.

El problema fue que en el tiempo dedicado al proyecto no se tenía contacto con el profesor de la asignatura y por lo tanto las dudas no podían resolverse de manera oportuna.

De cualquier manera la teoría estudiada y el desarrollo del proyecto contribuyeron de manera efectiva al aprendizaje del proceso de desarrollo de soluciones a un nivel de ejecución.

Al final de la materia se tenían las capacidades suficientes para planear y ejecutar un desarrollo que asegurara la satisfacción de un cliente.

Lenguajes formales y autómatas

Tiene por objetivo describir la teoría y la técnica para el diseño de lenguajes de computadora, así como los aspectos formales de la teoría de los lenguajes.

En esta materia se analizan varios temas que tiene que ver con la teoría de lenguajes desde el punto de vista lógico.

Se explora cuáles son las características de los lenguajes desde el punto de vista formal con el fin de poder utilizarlos y aplicarlos en la resolución de problemas.

Existen muchos casos de uso en los que se puede aplicar algún lenguaje formal o un autómata como parte del algoritmo que los resuelven.

Un ejemplo muy claro es la capacidad de determinar si un dato que ingresa el usuario a un sistema es correcto, como su nombre que no es posible que contenga números o símbolos que no sean letras.

Esta materia ha resultado muy útil en el desarrollo profesional, pues en ella se estudian conceptos que ayudan a resolver problemas muy específicos.

Algunos de los ejemplos resueltos en la materia tienen que ver con lograr identificar si las palabras del lenguaje están siendo utilizadas de una manera gramaticalmente correcta.

Esto es la base para poder identificar si el código escrito en algún lenguaje de programación es comprensible o no para una computadora, tanto en su sintaxis como en su semántica.

Para un ingeniero en computación es importante que pueda diseñar algoritmos que sean capaces de formalizar las sentencias de un lenguaje ya que ésta es la forma en la que se logra traducir instrucciones en un formato que las computadoras puedan procesar.

Con base en lo anterior esta materia sirve también para sentar un precedente de conocimientos base en el desarrollo de lenguajes de programación y de compiladores.

Por un lado un lenguaje de programación que esté bien formado nos permitirá escribir código que sea fácil de traducir a lenguaje máquina.

En otras palabras, gracias esta asignatura es posible diseñar un lenguaje de alto nivel que pueda ser traducido a instrucciones de bajo nivel.

Por otra parte un compilador es una solución de software que realiza la tarea de traducir un lenguaje de alto nivel en código de bajo nivel.

Con una base sólida de lenguajes formales ya autómatas se puede desarrollar este tipo de soluciones e incluso mejorar las ya existentes aplicando algoritmos que sean más eficientes.

Así mismo un compilador no sirve únicamente para traducir un código escrito en lenguaje de alto nivel, también resulta útil para detectar errores sintácticos que podrían causar una falla grave en el sistema.

Realizar esta detección oportunamente ayuda a reducir el tiempo y costo de desarrollo pues por un lado no se invierte tanto tiempo en corregir defectos y por otro no es necesario gastar en componentes que pudieran haberse dañado a causa de un defecto en un sistema de *software*.

En resumen, esta materia aporta mucho valor para el pensamiento lógico que permite resolver problemas de tipo semántico, así mismo otorga antecedentes que ayudan al aprendizaje de conocimientos en otras materias.

Lenguajes de programación

Tiene por objetivo explicar la importancia de estudiar las características y paradigmas de los lenguajes, además de discernir, de entre los diferentes lenguajes existentes, los óptimos para desarrollar sistemas de software de alta calidad; y diseñar nuevos lenguajes para computadora.

El estudio de los lenguajes de programación es importante ya que ésta es la manera en que es posible indicarle a una computadora qué pasos debe seguir para realizar un proceso.

En el curso de esta asignatura se revisó de manera breve cuáles fueron los orígenes de la programación para entender cómo se ha ido modificando a través del tiempo hasta llegar a tener sus características actuales.

Comprender de manera adecuada estas características es importante para un ingeniero ya que entre mejor sea su entendimiento podrá traducir de manera más sencilla un algoritmo a un código entendible para una máquina.

En consecuencia se entiende que esta asignatura es muy importante para la formación de un profesional de tecnologías de la información.

Una de las grandes ventajas competitivas en el campo laboral es la capacidad de construir un producto que atienda a una necesidad de negocio de manera rápida.

Con base en lo anterior se puede observar la importancia que tiene estudiar esta materia, pues un factor que particularmente ayuda a reducir el tiempo de desarrollo es el paradigma de programación con el que se planea resolver el caso de estudio.

Dado que hay una gran variedad en la naturaleza de los problemas a resolver, es posible que para un mismo problema un paradigma resulte más adecuado que otro.

Por ejemplo, en el caso de un sistema que lleve el control de calificaciones de un curso se necesita un enfoque que permita manejar fácilmente los datos y la relación entre ellos, por lo tanto lo más adecuado es utilizar un lenguaje que esté orientado a objetos.

En cambio, un programa que requiera tomar decisiones de manera lógica a partir de evidencia, como la inteligencia artificial de un videojuego, es mejor desarrollarlo en un lenguaje que implemente un paradigma lógico.

En resumen, esta asignatura trata temas relacionados con los diversos paradigmas de programación y en ella se analiza qué tipo de problemas se resuelven más fácilmente con qué tipo de lenguajes.

Adicionalmente se explica a profundidad cuáles son las características básicas de los lenguajes de programación en general.

Esto se hace con el fin de que el alumno tenga las herramientas suficientes para diseñar un lenguaje propio.

A pesar de que esta intención genera una buena oportunidad de desarrollo académico, se observa que el tiempo previsto para concluir con el plan es muy reducido y por lo tanto resulta poco práctico determinar que uno de los objetivos es desarrollar un lenguaje nuevo.

Sin embargo el tener este conocimiento básico no limita a sólo crear algo nuevo, también puede darse el caso en que un ingeniero logre perfeccionar la estructura base de un lenguaje ya existente a partir de lo aprendido en esta asignatura.

Para concluir, en este curso se presenta una amplia gama de posibilidades para utilizar como lenguaje de desarrollo, la elección del lenguaje siempre depende del problema a resolver y se considera que el plan es muy completo y otorga las herramientas suficientes para que el alumno pueda tomar esta decisión.

Además proporciona conocimientos adecuados para que, si se llega a dar el caso, el ingeniero tenga la capacidad de diseñar un nuevo lenguaje o incluso enriquecer alguno de los que ya existen.

Ética profesional

Tiene por objetivo fortalecer la vocación humana y profesional en la conciencia de la dignidad, el deber y el ejercicio de la libertad a través del conocimiento de la ética y el entendimiento de las responsabilidades.

Las materias que son de tipo filosófico, como la ética, suelen ser un contraste muy radical en la forma de trabajar de los estudiantes de Ingeniería.

Normalmente estas asignaturas no son consideradas muy importantes, sin embargo se puede observar que son indispensables para la formación de un ingeniero.

Esto se debe a que uno de los principales objetivos de la Ingeniería es mejorar la calidad de vida de la sociedad.

Gracias a la Filosofía es que una persona logra entender cuáles son las necesidades generales de la humanidad además de incluirlas como parte de sus objetivos personales.

Particularmente el estudio de la Ética ayuda a entender cómo es el deber del actuar de una persona en el contexto cultural en que vive.

Esta materia se dedica a estudiar cuáles son los aspectos que una persona debe entender de una sociedad para poder ejercer una vida profesional que aporte en la mejora continua de la humanidad.

Asimismo, la asignatura pretende infundir un sentido de conciencia y valores que se deben aplicar en el ejercicio de la profesión.

Definitivamente es una asignatura que directamente se aplica en el día a día, no de manera técnica sino con el fin de buscar un beneficio para la sociedad.

En conclusión, esta disciplina forma parte fundamental de cualquier profesionista, definitivamente una materia que aporta mucho en el aspecto humano del ingeniero y que debe tenerse presente para que con el cumplimiento del código de Ética, el egresado contribuya con la mejora de la situación social.

Bases de datos

Tiene por objetivo explicar los conceptos y principios en los que se fundamenta la teoría de Bases de datos, los cuales permiten diseñar, usar e implementar sistemas de Bases de datos.

Las bases de datos son herramientas indispensables en la forma en que se desarrollan soluciones de *software* actualmente y es por esto que esta materia es de especial importancia en el desempeño profesional.

De manera muy general la materia aporta conocimientos suficientes para entender qué son y cómo se pueden utilizar las bases de datos para construir un producto de *software* que pueda satisfacer ciertas necesidades de negocio.

El aspecto atendido por estas estructuras es el manejo de la información de negocio que está involucrada en un proceso.

La materia revisa varios de los temas que apoyan el razonamiento abstracto de los alumnos con el fin de que puedan modelar correctamente una estructura que englobe todos los posibles ámbitos de la información que se maneja en un negocio.

Con base en esto puede notarse que la materia tiene especial importancia en el área de desarrollo de *software* ya que aporta conocimientos que son fundamentales para construir productos de alta calidad.

Por otro lado, se aprecia que el plan considera impartir los temas desde un punto de vista muy teórico al no considerar un tiempo que se dedique a reafirmar los conceptos de manera práctica en un laboratorio.

En el semestre cursado, la materia se fue desarrollando con algunas clases prácticas y se considera que estas pocas horas de práctica resultaron muy útiles para reafirmar ciertos temas.

Al final de día es importante que el conocimiento no se quede en la fase de teoría, pues en el campo laboral es un recurso muy utilizado y es una gran ventaja competitiva el tener experiencia práctica en la implementación de bases de datos.

En contraste con lo anterior, la materia provee de las bases suficientes, explica por completo cuáles los fundamentos teóricos que permiten realizar implementaciones prácticas.

A pesar de no tener muchas horas de práctica se considera que el nivel competitivo que se alcanza es bastante bueno.

Adicionalmente, esta asignatura es un antecedente del módulo de salida de bases de datos y por lo tanto es una materia clave en la formación profesional de un ingeniero en tecnologías de la información.

Es por esta razón que se resalta la importancia de no descuidar las horas dedicadas al estudio de la teoría, aunque sí se sugiere aumentar algunas horas al plan para incluir un laboratorio práctico.

En resumen, esta área de conocimiento representa una oportunidad de desarrollo muy grande para los ingenieros en computación, se observa que el plan ha resultado muy útil y que podría mejorarse agregando un laboratorio para reforzar de manera práctica los conocimientos teóricos.

Por último, la asignatura menciona algunas tendencias actuales en el manejo de los datos, sin embargo, se considera que sería bueno dedicar más tiempo a conocer cuáles son las tecnologías utilizadas en la industria.

Por ejemplo, actualmente están surgiendo muchas soluciones que implementan esquemas *NoSQL* para el manejo de los datos. Éste es un enfoque que podría en el

largo plazo reemplazar totalmente el uso de las bases de datos y aun así no se menciona demasiado en la materia.

Administración de proyectos de *software*

Tiene por objetivo aplicar en un proyecto el proceso general mediante el cual se administran los proyectos de *software*.

La administración de un proyecto es la manera en que una empresa formaliza el proceso de desarrollo.

Esta asignatura revisa varios temas que ayudan a entender cuál ha sido el enfoque más utilizado en el campo laboral para conseguir este objetivo.

Lo anterior se considera muy importante porque es una manera en que el alumno pueda comenzar a familiarizarse con los pasos que sigue una empresa que se dedica al desarrollo de proyectos de *software*.

Sin embargo la materia enseña una ideología de tipo *Project Management Office (PMO)*, la cual da más importancia a la parte administrativa de la empresa y del proceso que a la satisfacción del cliente.

A pesar de que es una metodología muy común existen otras que están generando tendencia gracias a que gestionan los recursos de tal manera que la prioridad radica en resolver una necesidad de negocio en lugar de documentar exhaustivamente el proceso.

Con base en lo anterior se considera que deberían añadirse temas que estén orientados a analizar cuál es el comportamiento de las tendencias para elegir una metodología que gestione el desarrollo.

Se hace mención especial en las metodologías ágiles que han logrado obtener niveles de satisfacción más altos por parte de los clientes además de mejorar la eficiencia en el desempeño de los miembros del equipo de trabajo.

Por otro lado, esta materia ha resultado muy útil en el campo laboral ya que los temas explicados permiten entender cuáles son los aspectos básicos que deben

considerarse para la gestión de un proyecto y con base en ellos es posible que el ingeniero se adapte a procesos nuevos.

Cada tema aborda un elemento distinto de la gestión de los proyectos y a pesar de que las tendencias en la industria tengan un nivel distinto de prioridades, las metodologías nuevas aplican los mismos elementos para poder lograr sus objetivos.

Adicionalmente, esta asignatura muestra a los alumnos un área importante en la cual se pueden desarrollar profesionalmente.

La administración de proyectos es un área que siempre está presente en las empresas dedicadas a desarrollar *software* ya que permite tener un mayor control sobre la planeación del tiempo de los miembros de la empresa.

Asimismo, en el curso se forman habilidades directivas en los alumnos lo que representa una ventaja competitiva en el campo laboral.

En resumen, esta materia aporta un conocimiento muy valioso en el desarrollo de proyectos, sin embargo, se considera relevante que se incluya también un enfoque más actualizado de los temas además de dedicar tiempo a comparar el desempeño entre las nuevas tendencias y lo tradicional.

Redes de datos

Tiene por objetivo comprender y aplicar los conocimientos de protocolos, métodos y estándares sobre redes de datos dentro de las siete capas del modelo ISO/OSI.

Las soluciones y tecnologías actuales involucran la comunicación de datos entre varios dispositivos. Esta asignatura revisa los estándares y protocolos que se utilizan actualmente para resolver distintos problemas de comunicación tecnológica.

Por esta razón, ésta es una de las materias que más importancia tiene en la formación de un ingeniero en computación, es indispensable entender y poder utilizar adecuadamente el modelo de transmisión de datos entre computadoras.

La materia estudia diversos temas que permiten diseñar esquemas para comunicar dispositivos además de implementar un cierto nivel de seguridad para que sólo personas autorizadas puedan obtener los recursos disponibles.

Esto es un requerimiento esencial al desarrollar soluciones que aporten valor a un negocio, ya que normalmente la infraestructura de los sistemas únicamente debe de permitir las acciones de personal autorizado.

Adicionalmente las redes han permitido el envío de la información de manera rápida y confiable para que los datos necesarios sean adquiridos casi inmediatamente.

Un ejemplo muy claro de esto es el internet, el cual es una red que comunica prácticamente todos los dispositivos que existen.

Esto representa un área de oportunidad de negocio bastante grande para los expertos en tecnologías de la información ya que en el mercado existe cada vez una mayor demanda por adquirir información rápidamente y de manera segura.

Los temas que están descritos en el plan de estudios de la asignatura abarcan lo fundamental para que se puedan realizar las tareas básicas que están relacionadas con las redes de datos.

Esta materia ha resultado muy útil en el campo laboral pues gracias a los conocimientos adquiridos es posible atender los aspectos que involucran la comunicación de dispositivos en soluciones de *software*.

Además, la calidad en los productos desarrollados es superior ya que se contempla un aspecto muy importante en soluciones de negocio: la seguridad.

A pesar de que esta materia no desarrolla muy a fondo lo relacionado con seguridad, algunos de los temas están dedicados a explicar nociones básicas sobre ello.

Por lo tanto otra de las aportaciones es el área de seguridad, a manera de introducción se aprecia cuál es esa área de estudio y por qué se aplica en la industria.

Otro de los conceptos que tiene mucha relevancia es el de los protocolos más comunes en la capa de aplicación del modelo OSI.

Éste se explica en los últimos temas del plan y se estudia con el fin de que se conozcan cuáles son los protocolos que existen actualmente, para qué se usan y que de esta manera las aplicaciones que se desarrollen sigan los estándares.

Cada protocolo tiene una utilidad distinta y por lo tanto es importante conocerlos, pues se pueden aplicar en situaciones diferentes durante el desarrollo profesional.

Finalmente, se puede decir que ésta es una de las materias más relevantes en la carrera ya que aporta una formación en temas que se aplican directamente en la industria de las tecnologías de la información, además de que sirve como introducción a una de las áreas de especialización de la carrera.

Administración de redes

Tiene por objetivo conocer, identificar y aplicar diferentes enfoques, metodologías y técnicas que permitan planear, organizar, integrar, dirigir y controlar redes de datos dentro del esquema de la Administración.

Esta es una asignatura que complementa los temas tratados en “Redes de datos” profundizando en el aspecto administrativo de las redes.

Por esta razón se consideran más temas que tienen que ver con seguridad informática ya que estos conocimientos permiten asegurar que la información esté siempre disponible y que sólo pueda ser consultada por personal autorizado.

En su mayoría se estudian los protocolos que deben mantenerse para asegurar que la infraestructura de red de una organización sea segura.

Los conceptos que son explicados no están únicamente relacionados con el control de la red, también se revisa cuáles son cada una de las fases que deben seguirse en el proceso administrativo.

Primeramente debe existir una planeación de la implementación de la red, para poder realizar una propuesta correcta deben investigarse qué requerimientos se tienen.

Aplicado al campo laboral, esto resulta muy útil ya que existen múltiples ejemplos en los que el desarrollo de soluciones implica una red de comunicación.

Por ejemplo, todos los servicios que están disponibles en internet requieren de una infraestructura de red que les permita operar, ésta infraestructura puede planearse desde un inicio considerando ciertas características de la aplicación.

Algunos de las consideraciones pueden incluir el número de equipos de cómputo que son necesarios para ejecutar la aplicación, accesos internos para administradores o servidores de respaldo entre otros.

Posteriormente se diseña la red y se realiza una implementación. Esta implementación incluir otros dispositivos adicionales a las computadoras de la organización, por ejemplo teléfonos IP.

Varios de los temas que se estudian en la materia se pueden ver reflejados directamente en el desarrollo profesional, ya que un ingeniero puede dedicarse a esta área de especialización y estar constantemente atendiendo las necesidades de comunicación entre los dispositivos de una organización.

El curso se desarrolla de manera que el alumno pueda comprender los conceptos a un nivel básico y posteriormente pueda escalarlos a proyecto más grandes.

Además, algo que se considera muy útil para la asignatura fue que incluye un laboratorio dedicado a observar cómo aplicar los protocolos estudiados de manera práctica.

De esta manera la preparación es más integral y los conocimientos se refuerzan para poder enfrentarse a este tipo de tareas en proyectos del mundo real.

En resumen, esta asignatura es de aplicación directa en el mundo real pues los aspectos de seguridad informática están presentes en cualquier proyecto de tecnología en la actualidad, además de presentar de forma práctica y teórica en qué consiste el área de especialidad de redes y seguridad de la carrera.

Introducción al análisis económico empresarial

Tiene por objetivo conocer el funcionamiento de importantes conceptos económicos que contribuirán al exitoso desempeño profesional del ingeniero como empresario.

Ser empresario es una de las contribuciones sociales más influyentes, pues ayudan a estabilizar la economía y al crecimiento del país.

Con base en esta idea, la materia desarrolla temas que están relacionados principalmente con las características que conlleva crear una empresa además de mantenerla activa para que siga creciendo.

Lo aprendido en este curso tiene dos utilidades: la primera es que creando una empresa puede obtenerse un ingreso mayor al que si sólo se desempeña la vida profesional como trabajador.

Por otro lado una empresa tiene una contribución social muy importante al ayudar al crecimiento económico del país en general.

Esto se logra creando empleos para que más gente genere ingresos y aumente su calidad de vida, además que el flujo de dinero ayuda al bienestar de distintas áreas del mercado.

El curso está orientado a que el alumno adquiera conocimientos que sirvan para la fundación de empresas y que éstas puedan tener un correcto posicionamiento en su segmento de mercado.

Adicionalmente se estudia cómo debe interactuar la iniciativa privada en colaboración con los organismos de gobierno y de esta forma se aprenda cuál es el marco legal en el que debe estar operando la empresa.

Lo anterior ayuda a mantener el cumplimiento del código de ética para que además de mantener los valores morales de la universidad, se ejerzan las actividades empresariales cumpliendo con la normatividad del país.

Las aportaciones de esta materia para el desarrollo profesional de un ingeniero son numerosas pero en general aporta una visión de negocios, con la cual pueden posicionarse proyectos de Ingeniería de manera que cause impacto social y sea económicamente redituable.

Adicionalmente, en lo relativo a las tecnologías de la información, esta asignatura ayuda a adquirir una capacidad de detectar necesidades de negocio que representan oportunidades laborales para un ingeniero en computación.

Finalmente se estudian también algunos temas administrativos, los cuales son importantes para mantener saludable a la empresa y planear un crecimiento a futuro.

En conclusión, éste es uno de los cursos que proveen de ventajas competitivas a los egresados pues ayuda en el desarrollo de habilidades directivas y visión empresarial, lo cual es muy importante para complementar la parte técnica de la carrera.

Negocios electrónicos

Tiene por objetivo utilizar las metodologías de los negocios electrónicos en un proyecto a lo largo del semestre.

En esta asignatura se estudiaron varios temas relacionados con negocios de tecnología, y se analizaron algunos casos particulares con el fin identificar qué aspectos representan oportunidades de negocio para el desarrollo tecnológico.

El curso comienza por explicar conceptos teóricos que son básicos para entender cómo funciona el mundo de los negocios y cómo la tecnología puede proveer de ciertas ventajas en el mercado.

Principalmente las soluciones de *software* se implementan en un área conocida como inteligencia de negocios, la cual obtiene información estratégica sobre el comportamiento del mercado para que los negocios aprovechen de mejor manera las oportunidades.

La inteligencia de negocio es un área de conocimiento de las tecnologías de la información que representa una oportunidad profesional en el campo laboral muy demandada actualmente.

Es por esta razón que se considera esta materia muy relevante en la carrera, pues presenta una visión sobre las necesidades de soluciones tecnológicas que tiene este segmento de mercado.

Adicionalmente, la materia se desarrolla manteniendo el lenguaje a nivel de negocio, con lo cual se aprende a mostrar de manera efectiva cuáles son los beneficios que una solución de *software* puede aportar a una empresa.

Es aquí donde se puede apreciar la utilidad de esta asignatura en el desarrollo profesional, pues en la industria es necesario hacer ver el valor de un producto a nivel de negocio.

Este es uno de los retos más grandes para los ingenieros porque normalmente se está acostumbrado a describir las características técnicas, sin embargo no todos los clientes conocen el lenguaje y por lo tanto tienen dificultades para apreciar el valor que aporta a su negocio.

Por otro se analiza otro segmento de mercado que actualmente es muy popular: el comercio electrónico.

Existen muchos servicios que pueden contratarse de manera fácil y rápida a través del internet y esto representa una oportunidad muy grande, pues el alcance de clientes potenciales es muy grande.

No obstante, es importante señalar que el posicionamiento de un negocio electrónico no se da de manera inmediata y esta asignatura se dedica a estudiar algunos casos de éxito que ayudan a comprender cuáles son las acciones que favorecen el éxito.

La tecnología ayuda en varios aspectos a mejorar el proceso en que se hacen negocios ya que provee de información que puede utilizarse para realizar una planeación estratégica además de agilizar la producción y en la distribución.

En resumen, la asignatura es relevante en el desarrollo empresarial del ingeniero aportando una visión de cómo la tecnología influye en la forma actual de hacer negocios logrando esto a partir del estudio de conceptos básicos y de casos de éxito de empresas actuales.

Verificación y validación de *software*

Tiene por objetivo aplicar, en un proyecto, la metodología y herramientas de la validación y verificación del *software*.

Esta materia estudia técnicas para realizar pruebas sobre una solución de *software*, las cuales están orientadas a encontrar defectos en el sistema que puedan ser corregidos y que con esto aumente la calidad de los productos desarrollados.

Dicha calidad es un aspecto muy influyente en la satisfacción de los consumidores de las soluciones y asegurando un alto nivel de calidad es posible mantener y aumentar la cantidad de clientes.

Principalmente, se analizan cuáles son las distintas formas que existen para garantizar que un *software* cumpla con las necesidades de los clientes.

Es por esto que actualmente en la industria muchas empresas dedican esfuerzos para lograr este objetivo, pues se ha demostrado que la calidad de los productos y servicios es un aspecto sumamente importante para los negocios.

Los temas vistos son de aplicación directa en el desarrollo profesional, ya que el aseguramiento de la calidad es una ventaja competitiva en el desarrollo de soluciones a la medida.

Ésta materia ha sido muy útil pues ha contribuido a que se eviten fallas graves en los sistemas que se desarrollan.

Por otro lado, las pruebas ayudan a asegurar que se cumplan los requerimientos de negocio de los clientes.

La asignatura dedica un tiempo a explicar que además de las características técnicas como velocidad de procesamiento o costo también es necesario considerar cuáles son los pasos que sigue el proceso de un negocio, esto con el fin de asegurar que la solución desarrollada contribuya a realizar ese proceso de manera más eficiente.

Adicionalmente, se estudian algunas de las causas más comunes por las cuales los proyectos fracasan.

Este conocimiento resulta muy útil pues permite prever ciertos riesgos que ya fueron identificados por otros colaboradores de la industria.

El curso en el que se tomó la materia se desarrolló de manera muy teórica y con pocas horas de práctica.

A pesar de que el objetivo es aplicar un proceso de aseguramiento de la calidad en un proyecto de *software*, la materia no considera un tiempo de prácticas que permitan reforzar y demostrar la teoría.

Es cierto que el proyecto del curso es probar un proyecto realizado en asignaturas anteriores, pero resultaría más eficiente dedicar tiempo a realizar prácticas más específicas a cada tipo de pruebas.

Debido a que la calidad es fundamental al desarrollar un producto, esta asignatura resulta fundamental para el desarrollo profesional del ingeniero en tecnologías y se considera oportuno dar un enfoque más práctico a la materia.

Cada una de las materias aquí mencionadas aportó algo útil en mi formación como ingeniero.

Por un lado se adquirieron recursos técnicos con los cuales me es posible desarrollar prácticamente cualquier tipo de proyecto que esté relacionado con tecnología.

Por otro lado aprendí habilidades empresariales y de liderazgo que me han permitido tener una visión de negocios con la que me es posible identificar diversas oportunidades.

Adicionalmente, al conjuntar ambos aspectos he logrado presentar propuestas de soluciones que atienden las necesidades de los clientes de una forma en la que los involucrados en el proyecto puedan ver cuál es el valor que aporta la propuesta al negocio.

También he observado que la carrera imparte una cultura de trabajo en equipo, sin embargo me parece que se puede mejorar el plan de la carrera en este aspecto pues las acciones que deben tomarse para impartir estas habilidades van más allá de pedir que los trabajos sean entregados en grupos de n personas.

Esta parte de la colaboración en equipo es difícil de aprender ya que normalmente los ingenieros consideramos que el proyecto puede realizarse por un solo elemento.

Sin embargo resulta mucho más eficiente realizar la producción de soluciones en un equipo de trabajo porque esto ayuda a entregar el proyecto en el tiempo determinado por el cliente.

Por otro lado, un enfoque individualista es muy impráctico cuando se enfrentan problemas o proyectos muy complejos pues no es posible ser experto en todas las áreas de conocimiento que existen.

Para solucionar esto, se sugiere pensar en proyectos que tengan como objetivo colaborar con áreas ajenas a la Ingeniería, por ejemplo con licenciados en derecho, arquitectos o incluso filósofos por mencionar algunos.

En otras palabras, se propone que el plan considere un taller más práctico sobre desarrollo de proyectos interdisciplinarios para adquirir una experiencia básica de cómo debe llevarse proyectos de Ingeniería relacionados con otras áreas de conocimiento.

También se observó que el plan de la carrera tiene varias asignaturas que están relacionadas y resultaría oportuno considerar un plan de desarrollo de proyectos más integral desde el punto de vista de carrera.

Es decir, que los proyectos que se piden desarrollar en cada una de las asignaturas se planeen de tal forma que se lleve el mismo proyecto y al cual cada materia aporte algo.

Por ejemplo, podría desarrollarse el mismo proyecto juntando un grupo de administración de proyectos, uno de Ingeniería de *software* y uno de validación y verificación.

Para lograrlo sería necesario coordinar un tiempo para integrar los equipos de trabajo y que los alumnos se organicen para que cada quien realice las acciones que corresponden a la asignatura que están cursando.

Finalmente, es apropiado mencionar que durante la carrera he adquirido todos los conocimientos suficientes para desarrollarme profesionalmente de forma que los proyectos en los que he colaborado tienen algún impacto social positivo.

El área de especialización que elegí fue el de Ingeniería de *software* y debido a esto he decidido dedicarme al desarrollo de proyectos de tecnología.

La presente tesis abarca el desarrollo de una solución de *software* que analiza un gran volumen de datos y para entrar en contexto se explica a continuación el origen de esta área de conocimiento de la Ingeniería de *software*, además de los conceptos básicos que explican el área donde se identificó la oportunidad de una solución de *Big Data*.

Cómputo en la ciencia a través del tiempo

Los desarrollos tecnológicos en computación y los descubrimientos científicos han ido evolucionando a la par conforme pasa el tiempo. La computación provee de herramientas para que los investigadores puedan procesar datos y entender los fenómenos que estudian y por otro lado los científicos aumentan el tamaño de las colecciones de datos que utilizan lo que pone a prueba y empuja el alcance y las capacidades de la tecnología. El cómputo en la ciencia es muy importante porque conforme va evolucionando la tecnología, los científicos van adquiriendo una mayor capacidad en el análisis de sus datos.

Desde tiempos muy antiguos el hombre ha utilizado números para describir fenómenos que lo rodean, por ejemplo, el crecimiento de una población o su tasa de mortalidad. Este tipo de descripciones son el campo de estudio de la estadística descriptiva. Formalmente el autor Lacourly (2010) define a la Estadística como *“una rama del método científico que se ocupa del manejo de datos empíricos, es decir, de datos obtenidos al contar o medir fenómenos naturales, con instrumentos y métodos que generalmente entregan resultados inciertos. Además, ofrece métodos para la recolección, la agregación y el análisis de esos datos”*.

La estadística es el principal apoyo de las investigaciones en la ciencia, ayuda a describir fenómenos a través de números que son procesados y computados para que los resultados permitan entender el mundo que nos rodea. Para fenómenos muy complejos se requiere de mejores tecnologías que sean capaces de entregar resultados en un tiempo aceptable.

Uno de los primeros casos de estudio de la computación fue el procesamiento de números y de datos estadísticos con el fin de obtener resultados en menor tiempo, lo que permitía analizar cantidades más grandes de datos y así obtener mejores resultados.

El cómputo en la estadística comenzó desde los años 1920s, mucho antes de la invención de la primera computadora personal, cuando las universidades realizaron investigaciones en el uso de los tabuladores de tarjetas perforadas de IBM (Grier, s. f.). La mayoría de las primeras investigaciones en estadística fueron enfocadas hacia

estudios económicos y rápidamente se empezaron a aplicar las herramientas y técnicas en otro tipo de problemas (Grier, s. f.).

Los estudios en agricultura ayudaron a impulsar el avance tecnológico para el procesamiento de datos, ya que se buscaba aprovechar de la mejor manera posible las condiciones climáticas y de suelo para que la producción agrícola lograra cubrir las necesidades de la población. Principalmente se describía la productividad de los campos incluyendo distintas variables y esto mostraba qué semillas era más conveniente sembrar durante qué temporadas y eventualmente los estudios se enfocaron a la genética de las semillas para conseguir que los cultivos fueran mucho más productivos (Grier, 2000).

A través del tiempo se fueron dando más aplicaciones tecnológicas en investigaciones científicas y sociales con temas muy diversos y donde los proyectos están relacionados con distintas disciplinas.

J. A. Navarro y R. C. Barrientos (2013) establecen que *“las ciencias de la computación son la parte frontal y el centro de las ciencias en general, y de la biología y las ciencias ambientales en particular. La computación provee de medios para resolver problemas en áreas tan disímiles como genética, medicina, modelación fisiológica, mapeo geográfico y genético de cultivos, modelación del clima; provee también de herramientas para el manejo de bases de datos descriptivas de la biodiversidad así como predictivas del ambiente (e.g., salinidad, precipitación pluvial, etc.)”*.

El objetivo del cómputo es poder procesar suficientes datos para obtener resultados confiables y a su vez conseguir estos resultados en el menor tiempo posible para aprovechar el análisis y lograr que el resultado sea de valor para la investigación.

¿Cómo fue la situación en México?

En México la tecnología de computación tuvo sus primeras apariciones con la adquisición de máquinas de IBM especializadas en tareas administrativas en empresas alrededor de los años 1920s (Rodríguez y Cervantes, 2015).

En la Facultad de Ciencias de la Universidad Nacional Autónoma de México (UNAM), surgió un proyecto de investigación por parte de Sergio Beltrán López para aprender cómo se realizaban cálculos en la Universidad de California. Ahí descubrió que se utilizaba la IBM 650 y propuso al rector de la UNAM para conseguir una y realizar estudios científicos. El Centro de Cálculo Electrónico, después Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, utilizada la IBM 650 en para cálculos relacionados con física, astronomía, aplicaciones matemáticas y para investigación de lenguaje máquina y lenguaje ensamblador para programar (Rodríguez y Cervantes, 2015).

En 1960 se compró una computadora Bendix G-15 que se utilizó para dar varios cursos de computación en distintas universidades del país, pero el proyecto terminó cuando la computadora se incendió en un viaje de Monterrey a la Ciudad de México. Estos viajes despertaron interés en varias ciudades. El Centro de Cálculo Electrónico compró otra G-15 que fue posteriormente reemplazada por una Bull Gamma 30 con cintas magnéticas y que se programaba utilizando Fortran (Rodríguez y Cervantes, 2015).

A partir de esto otras Universidad es como el Instituto Politécnico Nacional (IPN) o el Instituto Tecnológico y de Estudios Superiores de Monterrey (ITESM) compraron computadoras e iniciaron investigaciones sobre temas de computación (Rodríguez y Cervantes, 2015).

El primer artículo mexicano sobre computación fue publicado en 1958, escrito por Raúl Pavón y tenía el título de "*The Mexican Light and Power Company Introduces a Direct Way for Fast Computation of Industrial Services with Power Factor Adjustment*". El artículo describe un método numérico utilizado para calcular la raíz cuadrada del factor de potencia en los transformadores (Rodríguez y Cervantes, 2015).

Posteriormente otras organizaciones adquirieron sus propias computadoras y comenzaron a utilizarlas para tareas administrativas o para cálculos operativos de su sector. Algunas de éstas fueron: Ingenieros Civiles Asociados (ICA), Instituto

Mexicano del Seguro Social (IMSS), Petróleos Mexicanos (PEMEX), Cámara de Diputados, entre otras (Rodríguez y Cervantes, 2015).

Un evento importante que se puede mencionar fue que en 1970 el censo general de población fue procesado de manera electrónica por primera vez. Algunos años más tarde los semáforos comenzaron a ser controlados a través de una computadora. Fue durante la década de los 70s que varias universidades crearon sus carreras de Ingeniería en sistemas, maestrías y algunos otros cursos académicos relacionados con las ciencias de la computación (Rodríguez y Cervantes, 2015).

En 1972 el Banco Nacional de México (BANAMEX) lanzó el primer cajero automático y las redes bancarias facilitaron pago crediticio a lo largo del país.

A finales de 1980 surgieron las computadoras personales de bajo costo y esto ocasionó que en los 90s la mayoría de las instituciones educativas considerasen la computación como una herramienta útil en cualquier área de conocimiento (Rodríguez y Cervantes, 2015).

Historia del Disco Duro

Se ha visto que el cómputo es relevante para la ciencia porque ayuda al procesamiento de los datos en las investigaciones. Sin embargo, existe también otra necesidad que es importante considerar: el almacenamiento de estos datos.

Antes de la tecnología de la computación la información era guardada en archivos, hojas de papel almacenadas en estantes.

Cuando surgen los tabuladores de tarjetas perforadas, los datos se podían representar en las tarjetas. Unas tarjetas eran para leer datos de entrada, otras para procesar los datos y otras para representar datos de salida.

El siguiente paso fue almacenar la información en cintas magnéticas. El primer computador en utilizar cintas magnéticas para almacenar y procesar archivos grandes fue la *Universal Automatic Computer (UNIVAC)*. La lectura y escritura de la información se hacía de manera secuencial. Esto permitía realizar respaldos y

recuperación de los datos sin pérdidas en caso de que se presentara una falla en el funcionamiento del sistema (Daniel, 1999).

Posterior a la cinta magnética surgió la idea de guardar los datos dentro de discos magnéticos. El funcionamiento de los discos era más rápido que el de las cintas magnéticas, porque el diseño de los discos permitía un acceso aleatorio a los datos, a diferencia de las cintas donde la información se procesaba de manera secuencial (Daniel, 1999).

Eric D. Daniel señala que durante 1957 IBM desarrolló investigaciones para desarrollar un disco magnético capaz de almacenar información que fue conocido como IBM 350, éste formaba parte de un producto conocido como *Random Access Method of Accounting and Control (RAMAC)*. Se menciona que el siguiente paso fue incluir discos intercambiables en los sistemas de almacenamiento y desarrollar una interfaz para intercambiar información entre las tarjetas perforadas y los discos magnéticos (Daniel, 1999).

Para mejorar la capacidad de almacenamiento de los discos se fue mejorando la tecnología de las cabezas magnéticas con las que se escribía y se leía la información, así como aumentando la densidad de área de los discos. También se buscó mejorar los tiempos de acceso reduciendo el tamaño del diámetro de los discos. El objetivo era lograr discos que tuvieran la posibilidad de contener más datos ocupando un menor tamaño y con el menor costo posible (Daniel, 1999).

Físicamente los componentes tecnológicos se han ido adaptando para poder almacenar los datos recolectados y los resultados obtenidos del procesamiento de esos datos. El avance tecnológico de componentes ha favorecido al cómputo en varios aspectos.

Por un lado, los componentes tienen mayor confiabilidad y por lo tanto son menos propensos a perder información al tener menos fallas causadas por un defecto de fabricación; también tienen un mayor tiempo de vida útil.

Se tiene también un mejor desempeño de los dispositivos. Como se ha mencionado, los dispositivos tienen una mayor capacidad de almacenamiento con cada nueva generación y también van reduciendo sus tiempos de operación; guardan, recuperan

y eliminan registros de manera más eficiente. Gracias a esto, se reduce el tiempo de computación para los datos utilizados y también se logra procesar más datos en el mismo tiempo.

El principal beneficio resulta entonces un ahorro en el costo y tiempo de las investigaciones, el avance tecnológico de dispositivos ha permitido explorar áreas que antes parecían imprácticas para la investigación científica.

Almacenamiento lógico de los datos: archivos, bases de datos, *data warehouse* y NoSQL

Durante el desarrollo de los sistemas informáticos, los desarrolladores se enfrentaron también a otro desafío: estructurar los datos almacenados para lograr cierta compatibilidad entre distintos programas.

En los inicios de la computación los datos de entrada utilizados en los programas debían programarse cada vez que se pretendiera ejecutar el sistema. No existía una necesidad de tener almacenados estos datos, pues las entradas y las salidas de estos sistemas se escribían en papel y se programaban cada vez que fuera necesario (Jiménez, s. f.).

Como se mencionó anteriormente, el almacenamiento de la información comenzó con las cintas magnéticas, cuyo uso principal era automatizar los datos a procesar y almacenar respaldos de los datos de salida, el problema era que las computadoras no tenían forma de saber por sí mismas qué contenía la cinta y dónde exactamente estaban guardados los datos que necesitaba para ejecutar cierto programa. Para resolver esto se diseñó un sistema de archivos que almacena la información en bloques de información identificados dentro de la memoria con la información suficiente para poder encontrar fácilmente los datos necesarios. Las computadoras crean un archivo estructurado de cierta manera y guardan un registro del nombre, el tamaño del archivo y en qué parte de la unidad de memoria (cintas magnéticas, discos duros, entre otros) se puede encontrar.

Este sistema de archivos representó un gran avance, sin embargo, al empezar a trabajar con cantidades grandes de información (que para la época pudieran ser unos 50 o 100 archivos) se identificaron algunas problemáticas:

- La redundancia en los datos. Dentro de archivos distintos se repetía la información (Jiménez, s. f.).
- Inconsistencia en la información. Si se tiene alta redundancia en los datos que se almacenan, es normal que se generen inconsistencias en los distintos archivos donde se estén almacenados los datos (Jiménez, s. f.). Por ejemplo, en un registro reciente se tiene a una organización registrada como “Autos S. A.” y si lo comparamos con un registro anterior de la misma organización es posible que lo encontremos registrado con un nombre diferente, tal vez “Automóviles S. A.”
- Costos elevados. El problema de inconsistencia, generó la necesidad dedicar tiempo y recursos a tratar de evitar estos problemas, esto implica que el costo de los sistemas sea mayor. La tecnología de almacenamiento físico en esta época tenía poca capacidad, por esta razón se requerían varios dispositivos para guardar una gran cantidad de archivos y entre más datos se manejaban, más costoso resultaba mantener la consistencia de los datos.
- Baja seguridad. No se tenía contemplado controlar el acceso a los archivos de datos así que cualquier programa tenía la capacidad de leerlos y modificarlos.

A partir de estas problemáticas y en la década de 1960 surge el primer acercamiento a las bases de datos definiéndose dos modelos para almacenar datos: el modelo jerárquico y el modelo de red. Una base de datos es una colección de datos almacenados de forma permanente que se comparten entre múltiples usuarios y programas (Jiménez, s. f.).

El principal objetivo de las bases de datos era recolectar los datos en un mismo repositorio para poder tener acceso a ellos desde distintos programas, además de resolver las problemáticas expuestas en el sistema de archivos.

Primero surgió el modelo jerárquico que se basa en relaciones de tipo padre – hijo. Cada registro es un conjunto de datos que pertenecen a una entidad padre o una

entidad hija y la información se representa a manera de árboles de jerarquías, cada árbol tiene un padre y n hijos (Elmasri y Navathe, 2011).

Posteriormente se diseñó el modelo de base de datos en red. Éste se basa mucho en la teoría matemática de conjuntos para representar los datos y las relaciones entre los registros (Elmasri y Navathe, 2011).

Estos modelos permitieron comenzar a estructurar los datos para lograr evitar redundancia y reducir tiempos de búsqueda de la información, aunque tenían ciertas desventajas en comparación con los sistemas actuales, principalmente tenían un alto grado de complejidad para su diseño e implementación.

En la década de 1970 surge el modelo más usado hasta la actualidad (hablando de bases de datos estructuradas): el modelo relacional. Éste modelo fue definido por Edgar Frank Codd quien redactó 12 reglas que describen los requerimientos de un sistema manejador de base de datos relacional. En este modelo la información es representada a manera de tablas donde cada registro es una fila de la tabla y se definen relaciones entre las tablas utilizando un identificador único como referencia (Elmasri y Navathe, 2011).

Después del desarrollo de Codd, surgieron otros modelos que se adaptan al modelo relacional. Uno de ellos fue el modelo entidad relación, propuesto por Peter Chen alrededor de 1976. Este modelo se basa en la idea de que el mundo real está compuesto por entidades (abstracciones de algún elemento) y que estas entidades tienen ciertos atributos y relaciones, donde una relación se define como una asociación que existe entre las entidades que componen a la base de datos (Ying y Bhalla, 2012). Algunos ejemplos de sistemas manejadores de bases de datos (*DBMSs* por sus siglas en inglés) que utilizan este modelo son: *Oracle*, *IBM DB2* y *MariaDB*.

Los *DBMSs* actuales utilizan un lenguaje estándar llamado *Structured Query Language (SQL)* que sirve para definir la estructura utilizada para modelar los datos y para insertar, actualizar, borrar y consultar los datos.

Las bases de datos son la forma en que actualmente muchos datos son almacenados para manejar distintos procesos de negocio y de investigación, han

resultado muy útiles para estructurar datos crudos y lograr un análisis más simple para obtener información de valor. Almacenar los datos con diferentes modelos y estructuras ha permitido una mayor organización y simplicidad para tomar datos, procesarlos y obtener un resultado, cada modelo aporta algo nuevo al análisis desde realizar una operación entre dos números hasta relacionar ese número con algo más cercano a la realidad como una calificación o un total a pagar en un ticket de venta.

Los datos por sí solos permiten obtener los resultados que se definen directamente para un modelo, por ejemplo, en un punto de venta donde los datos almacenados representan los tickets de compra de los clientes, se guarda la cantidad de artículos adquiridos, el precio unitario de cada producto y el monto total a pagar de cada compra; sin embargo es posible obtener información que en un principio no se tenía contemplada si se realiza un análisis diferente; tomando el ejemplo anterior, analizando todos los tickets de un periodo podríamos obtener los productos más vendidos o los meses en que más ventas se tienen.

Análisis de datos

Los sistemas de bases de datos son muy eficientes para manejar el guardado de datos, pero cuando se requiere realizar múltiples consultas que involucren muchos datos o quizás funciones agregadas a los registros (como la cuenta o el promedio de un cierto campo) presentan ciertas dificultades.

Estas dificultades fueron detectadas porque dentro del mundo de los negocios las empresas comenzaron a notar el beneficio de analizar los datos de sus sistemas agrupándolos por fechas, por tipo de producto, entre otros. Esto dio lugar a un nuevo desafío para el procesamiento de datos: obtener información útil sobre un negocio a partir de los datos generados en el día a día como notas de venta o registros de clientes con el fin de realizar toma de decisiones, planeación y coordinación de los procesos de una empresa (Bojicic et al., 2016).

El problema principal es la necesidad de aplicar varias operaciones a los datos almacenados en una base de datos para obtener cierta información, se necesitan contar cuántos registros existen que cumplan cierta condición, sumar el total de

cierto dato para los registros que cumplan algún otro criterio. Tal como se dijo antes, las bases de datos están optimizadas para el almacenamiento de los datos y les resulta más costoso (en tiempo de procesamiento) obtener los datos almacenados y si se reflexiona sobre esta premisa se puede observar el por qué resulta lento obtener el análisis de grandes cantidades de información utilizando los datos de una base de datos, lo que dio paso al planteamiento nuevos modelos y nuevas estructuras para almacenar los datos.

Alrededor de 1993 surgen dos soluciones para el análisis de datos: el *data warehouse* (*DW*), que en español puede traducirse como almacén de datos y el *data mart* (*DM*) o mercado de datos.

Un *DW* es “*un modelo de un sistema de negocio en concreto que representa un conjunto de todos los estados de ese sistema durante un intervalo de tiempo dado [...] Uno de los principales problemas relacionados con el desarrollo y mantenimiento de un DW es la inconsistencia entre el sistema y el DW, que aumenta con el paso del tiempo*” (Bojicic et al., 2016).

Básicamente un *DW* se construye guardando cada cierto tiempo los registros almacenados en una base de datos utilizando procesos de tipo *Extract, Transform and Load* (*ETL*) que extraen los datos de los repositorios de información, los transforman (generalmente para limpiarlos y homologarlos) y finalmente los cargan al *DW*. Para obtener datos analizados simplemente se hace una consulta al *DW* definiendo los parámetros necesarios de tiempo, ubicación o cualquiera que sea el objeto de análisis. En comparación resulta costoso construir el *DW* pero obtener estadísticas es mucho más rápido que si se realizan operaciones directamente en una base de datos.

El inconveniente de los *DW* es que sacrifica la nula redundancia que se logra cuando se utiliza una base de datos pero a cambio otorga un esquema mucho más simple de analizar.

Las primeras definiciones sobre los *DW* y los *DM* fueron acuñadas por los autores Inmon (2002) y Kimball (2002). Ambos desarrollaron escritos independientes sobre qué son y cómo deben implementarse los *DW* y los *DM*. Para Kimball (2002) un *DW*

es el conjunto de datos de una organización representando los diferentes estados y áreas a través del tiempo y cuya estructura está diseñada específicamente para realizar consultas de análisis con alto desempeño y facilidad de uso.

Un *DM* es un subconjunto lógico del total de los datos almacenados en un *DW*. Según los escritos de Inmon (2002) un *DM* representa una estructura extraída de un *data warehouse* y que se alimenta de los datos almacenados en él donde los datos se encuentran desnormalizados para satisfacer las necesidades de información requeridas para un departamento o área específico de una organización.

Es evidente la forma en la que estas dos estructuras lógicas de almacenamiento de los datos funcionaron como un primer acercamiento para resolver las necesidades que hoy en día son cubiertas por las herramientas modernas de *Big Data*. Los *DW* y *DM* extraen grandes volúmenes de datos constantemente, los datos son variados y se unifican según sea la necesidad a resolver, se obtiene un análisis rápido al permitir consultas simples y todo eso permite que se tomen decisiones respecto al curso que debe tomar el negocio en periodos futuros.

Un par de años más tarde apareció otro término muy relacionado con el análisis de datos: las bases de datos *NoSQL*. *NoSQL* significa *Not Only SQL*, lo que significa que las bases de datos *NoSQL* almacenan y manejan datos no estructurados, conocidos también como no relacionales (Bakshi, 2010).

Las bases de datos *NoSQL* son más sofisticadas que las tradicionales porque permiten realizar operaciones con datos estructurados y con datos no estructurados; esto resulta en una gran flexibilidad para modificar la estructura de los datos, pues su diseño no incluye la definición de un esquema.

En 2011, el autor Jing habló sobre el teorema de CAP, propuesto por el profesor Eric Brewer en el año 2000 y que define tres principales necesidades sobre los datos de análisis: consistencia, disponibilidad y tolerancia a la partición de red. De acuerdo con este teorema sólo es posible atender dos de las 3 necesidades.

La clasificación propuesta por Jing et al. (2011), nos habla sobre tres tipos de bases de datos *NoSQL*: llave – valor, bases columnares y bases de datos de documentos; teniendo como ejemplo Redis, Cassandra y MongoDB respectivamente.

A pesar de sus mejoras con respecto a bases de datos tradicionales, este paradigma tiene algunos puntos donde se ve más limitado que el enfoque tradicional, por ejemplo: en el manejo de transacciones. Los sistemas *SQL* tienen mejor desempeño en el manejo de operaciones transaccionales sobre los diseños *NoSQL*.

Los esquemas *NoSQL* y los *DWs* dieron paso a un concepto muy importante en análisis de información y que es tema de estudio para esta tesis: el *Big Data*.

La intención del *Big Data* es integrar la información estructurada (bases de datos) y la no estructurada (documentos, *blogs*, redes sociales, entre otros) para realizar un análisis que muestre información sobre el comportamiento de un fenómeno y así ayudar a la toma de decisiones y la planeación.

Big Data es un concepto que surge de la percepción de la gran cantidad de datos crudos que se generan al día como resultado de la navegación de los usuarios en internet (redes sociales, noticias, *blogs*, entre otros) y cuyo volumen de información generado alcanza los Peta Bytes (PB) (Sruthika y Tajunisha, 2015).

Como principal antecedente se puede hablar del desarrollo de los buscadores web. Cuando comenzó a utilizarse popularmente el internet y muchas páginas se crearon, surgió como problemática a resolver el poder encontrar información relevante entre tantas referencias. Varios investigadores de universidades desarrollaron herramientas de búsqueda que categorizaban las páginas de internet y permitían encontrar información relacionada con ciertas palabras clave.

Gopher, Archie, VERONICA y JUGHEAD

Un primer acercamiento fue el desarrollo de los sistemas *Gophers* (nombrados así por la mascota oficial de la Universidad de Minnesota, donde se desarrolló el sistema), que como lo define Lindner (1992) son sistemas de “*búsqueda y adquisición de información distribuida en ordenadores remotos conectados a través de Internet*”.

Gopher comenzó como un Sistema de Información del campus de la Universidad de Minnesota y fue creado por el equipo de apoyo de microordenadores de la

universidad. El prototipo original era un explorador de árbol multicolumna en el cual cada columna de desplazamiento tenía una lista de títulos (Khare, 1999).

Tal como se menciona en varias publicaciones, el funcionamiento de un sistema *Gopher* está basado en una arquitectura cliente / servidor donde el servidor almacena el título de documentos distribuidos en diferentes nodos de la red y ciertas características sobre su contenido; y el cliente permite realizar peticiones de búsqueda al servidor. El servidor define una cierta jerarquía para buscar los documentos y esta jerarquía se construía a partir de vínculos entre carpetas/directorios y servidores (Adell, 1992). El cliente y el servidor se ejecutaban en computadoras diferentes y por lo tanto podían ejecutarse en plataformas diferentes.

El *Gopher* es principalmente un protocolo mediante el cual se comunican dos aplicaciones que residen en ordenadores diferentes y que permite compartir información referente a los recursos disponibles en una red (Adell, 1992).

Gopher combinaba características de servicios de boletines electrónicos y bases de datos para lograr la búsqueda de documentos a partir de la jerarquía definida y de palabras clave (Lindner, 1993). La idea del protocolo *Gopher* era mostrar un menú con los recursos disponibles para realizar una búsqueda más ordenada de los archivos.

Gopher soporta una amplia gama de datos que se puede buscar utilizando la aplicación cliente. Principalmente realiza búsquedas de directorios, archivos de texto, guías telefónicas o elementos multimedia entre otros (Lindner, 1993).

Los *Gophers* nos ayudan a explorar los recursos de la red tomando en cuenta palabras clave como criterios de búsqueda y tienen la capacidad de seleccionar archivos de cualquier nodo en la red incluyendo servidores FTP. En sí *Gopher* fue un protocolo adaptado en varios buscadores, por ejemplo: *Archie*, *JUGHEAD* y *VERONICA* (Gaffin, 1994). Estos buscadores recibieron sus nombres de los personajes de los famosos comics de *Archie* y cada uno atacaba problemas particulares.

El *Gopher Archie* era una base de datos que contenía los nombres de los archivos alojados en servidores *FTP* y fue desarrollado originalmente por Alan Emtage, Bill Heelan y J. Peter Deutsch durante 1990 (Seymour, Frantsvog, Kumar, 2011). Fue el primero de los tres en surgir y su objetivo era mostrar un catálogo de los archivos que contenía un servidor y realizar búsquedas de palabras clave en el nombre de cada archivo (Seymour, Frantsvog, Kumar, 2011). Sin embargo, *Archie* no indexaba el contenido de los archivos porque la cantidad de recursos contenidos en un servidor era muy reducida y fácilmente se podía realizar una búsqueda manualmente (Seymour, Frantsvog, Kumar, 2011).

VERONICA (*Very Easy Rodent-Oriented Netwide Index to Computerized Archives*) fue creado por Fred Barrie y Steve Foster a finales de 1992 en la universidad de Nevada como un servidor *Gopher*. Se encargaba de navegar los distintos catálogos de *Gopher* para buscar archivos y crear una lista indexada de servidores de acceso público, para así lograr una búsqueda más rápida en un mayor número de servidores *Gopher* (Frana, 2004)

La ventaja principal de este sistema es la indexación de archivos. Los índices reducen el tiempo y aumentan la eficiencia al buscar en un conjunto grande de archivos así que las búsquedas realizadas con *VERONICA* eran mucho más rápidas que las consultas hechas con algún otro cliente *Gopher*.

La primera aplicación de *VERONICA* fue un motor de búsqueda llamado *Digital NeXT* que fue utilizado para realizar consultas en una biblioteca. El motor buscaba e indexaba la información de directorios y archivos para después realizar consultas booleanas y así localizar la información (Frana, 2004).

En paralelo al desarrollo de *VERONICA* surgió también el buscador *JUGHEAD* (*Jonzy's Universal Gopher Hierarchy Excavation And Display*), cuya única diferencia con *VERONICA* fue que limitaba sus búsquedas a los sistemas de *Gopher* específicos en los que residía.

Según el autor Frana (2004) *JUGHEAD* fue elaborado por Rhett "Jonzy" Jones en el centro de cómputo de la Universidad de Utha en 1993 y consistía en un buscador de texto sin formato. *JUGHEAD* fue usado principalmente en las bibliotecas donde

...limitaba las jerarquías de archivos logrando resultados relevantes más rápido gracias a su indexación.

La *web* y los buscadores

Uno de los grandes avances en *Big Data* fue la implementación de estos servicios de búsqueda, pues con el surgimiento de la *web*, la humanidad fue generando una cantidad enorme de información que actualmente sigue en constante expansión. Las tecnologías se adaptaron y el uso de recursos en internet se facilitó gracias a la indexación implementada por los buscadores.

La *web* es un conjunto de recursos y enlaces interconectados en el cual los usuarios pueden captar el significado de un documento así como determinar su relevancia, cosa que una maquina aun no puede interpretar tan fácilmente (Sadeh, Walker., 2003). Como se ha mencionado, la cantidad de recursos disponibles en la *web* está en constante incremento y el enfoque principal para consultar la información ha sido filtrar los registros por palabras o frases clave que se relacionen con lo que se espera encontrar.

A principios de 1993 no existían motores de búsqueda para la *web*, así que la forma de consultar la información disponible era a través de catálogos que eran actualizados a mano constantemente. Uno de los primeros motores de búsqueda fue el *W3Catalog* que tomó como base una serie de rutinas escritas en lenguaje Perl por Oscar Nierstrasz de la Universidad de Ginebra, que se encargaban de mantener actualizado un catálogo de contenido *web* en un formato estándar (Seymour, Frantsvog, Kumar, 2011).

En ese mismo año, Matthew Gray del MIT, escribió un programa que se dedicaba a navegar en las páginas con el fin de dimensionar el tamaño de la *World Wide Web* (Seymour, Frantsvog, Kumar, 2011). Esto sentó un precedente a los *web crawlers*, que pretendían generar un índice de las páginas web disponibles en internet al visitar la mayor cantidad de sitios posibles. Los índices mencionados fueron utilizados para construir los motores de búsqueda que se utilizan actualmente como método para navegar y recuperar recursos de internet.

A partir de todos estos catálogos se fueron desarrollando diferentes motores de búsqueda que permitían recuperar recursos e información de internet utilizando palabras clave para encontrarlos. El funcionamiento básico de los motores de búsqueda consiste en un *bot* (también conocidos como *web crawlers*) que se dedica a navegar en las páginas *web* y que extrae palabras clave para construir un índice, el cual es utilizado posteriormente para mostrar recursos que estén relacionados con la búsqueda realizada por el usuario.

Existen varios ejemplos de motores de búsqueda como *Lycos* o *AltaVista* (motor que fue la base para el buscador de *Yahoo!*), pero el más conocido y utilizado actualmente es *Google* que fue creado en 1997 por Larry Page y Sergey Brin, compañeros de la Universidad Stanford (Hormby 2015). A pesar de que ahora *Google* domina el mercado de motores de búsqueda no siempre fue así, pues no fue el primer motor que se desarrolló.

Los primeros motores realizaban una comparación entre las palabras buscadas por el usuario y palabras clave indexadas a partir de los *web crawlers*, así se calificaba una relevancia del contenido de las páginas para ordenar el resultado de la búsqueda y mostrar primero las páginas más relacionadas con las palabras buscadas (Hormby 2015). Sin embargo este método resultó poco eficiente, pues no siempre se lograba encontrar el contenido deseado; consideremos el ejemplo explicado por Hormby (2015): “[...] al realizar una búsqueda para “Microsoft”, las páginas de vendedores de productos Microsoft podrían estar categorizadas como más relevantes que la página corporativa de Microsoft porque en una sola página pueden estar listadas docenas de productos Microsoft.”

La ventaja principal por la que *Google* llegó a ser líder en motores de búsqueda fue el algoritmo empleado para calificar la relevancia de una página para las palabras clave utilizadas en la búsqueda (Hormby 2015). *Google* califica no sólo el contenido de la página *web* sino que también aplica un cierto peso utilizando las referencias de otras páginas hacia la página a analizar y las referencias de otras páginas relacionadas que están incluidas dentro de la página a analizar (Hormby 2015). Es hasta cierto punto como buscar un artículo en una publicación académica, los más

relevantes al tema seguramente son citados repetidamente por otros autores e incluyen varias fuentes de información.

El éxito de *Google* fue tan grande que *Yahoo!*, su principal competidor, tuvo la necesidad de integrar el motor de *Google* al suyo para no salir completamente del mercado (Hornby 2015).

Otro competidor en cuanto a motores de búsqueda (pero quizás no tan popular como *Yahoo!* o *Google*) es el buscador *Bing*. *Bing* surgió a partir del esfuerzo que puso *Microsoft* en este campo, comenzando por un motor llamado *MSN Search* (lanzado en 1998) y que integraba los resultados de los motores *Inktomi* y *Looksmart*, aunque fueron reemplazados por los resultados de *AltaVista* durante un breve periodo en 1999 (Seymour, Frantsvog, Kumar, 2011). En 2004, *Microsoft* comenzó a utilizar tecnología basada en su propio *web crawler*, conocido como *msnbot*, y finalmente en 2009 se realizó una alianza entre *Microsoft* y *Yahoo!* para definir que *Yahoo! Search* estaría impulsado por la tecnología *Microsoft Bing* (Seymour, Frantsvog, Kumar, 2011).

El caso de *Google* es un buen ejemplo del reto que representa el desarrollo de herramientas para *Big Data*. El problema en ese momento era categorizar la relevancia en las búsquedas a páginas y al mismo tiempo incluir la mayor cantidad de fuentes para asegurar un buen resultado; aquí el volumen de información analizada se considera muy grande pues se está hablando de la posibilidad de indexar el contenido de todas las páginas publicadas en internet para realizar búsquedas que ordenen los resultados por su relevancia. La solución puede resultar obvia una vez que se conoce pero se debe notar que fue un gran avance para la época, por eso existieron tantos motores distintos (*Lycos*, *WAIS*, *AltaVista*, *Yahoo!*, *HotBot*, entre otros) que dejaron de ser conocidos una vez que *Google* comenzó a popularizarse entre los usuarios de internet.

Big Data

Las soluciones de *Big Data* consisten en la recolección y análisis de grandes cantidades de información. Sin embargo deben considerarse algunos aspectos cuando se quiere implementar una solución de *Big Data* para una empresa.

Lo primero es definir bien un objetivo alcanzable que resulte del análisis de los datos, por ejemplo en el caso de los buscadores *web* el objetivo fue categorizar las páginas y encontrar la información más relevante para cada tema de tal forma que los usuarios pudieran encontrar los recursos a partir de palabras clave.

Es necesario también encontrar la mejor forma para presentar los datos. Retomando el caso de los buscadores, la mayoría consiste en un diseño minimalista con un sólo campo de texto donde se introducen las palabras clave y los resultados se muestran ordenados por relevancia (principalmente, porque es posible aplicar distintos tipos de ordenamiento, por fecha por ejemplo).

Para dar una definición más formal se tomará otro caso en que el *Big Data* tuvo sus orígenes. El autor Shmarzo (2013b) presenta una breve descripción de cómo surge la necesidad de las herramientas de *Big Data Analytics*. En su obra *Big Data: El poder de los datos*, incluye un capítulo dedicado a redactar cómo las empresas fabricantes de bienes de consumo comienzan a analizar datos de venta provistos por el mercado minorista.

El análisis parte de las auditorías que entregaba Nielsen, una empresa líder en información de mercado, a Procter & Gamble, posteriormente utilizando un sistema de escáneres de terminal de punto de venta (TPV) y códigos de barra que permitía llevar un mejor control de inventarios y que otorgó a las empresas de un panorama del comportamiento de consumo de los usuarios.

Para este caso el origen de los datos son las fuentes de información de los tickets de compra generados a partir de puntos de venta.

Tal como se menciona en el libro de Shmarzo (2013), el volumen de datos generado por los escáneres de TPV aumentó considerablemente y se buscaban herramientas que dieran a las empresas un análisis introspectivo con el fin de generar una base

sólida para tomar decisiones que permitieran optimizar la cadena de suministro, predecir la demanda de productos, categorizar las compras por temporada o tipo de cliente e incluso desarrollar programas fidelización al cliente.

El análisis continuo de estos datos mostró a las empresas hábitos de compra de sus usuarios, segmentación de mercado y otros indicadores que permitieron tomar decisiones efectivas para su crecimiento.

Entender este análisis puede resultar muy simple o muy complicado dependiendo mucho de cómo es que se muestran los resultados al interesado. Por ejemplo, observar en una gráfica de línea el monto total de ventas por mes en un año permite apreciar fácilmente si durante el año analizado las ventas fueron en aumento, disminución o si se mantuvieron estables. Imaginemos ahora que estos mismos datos se muestran en una tabla plana, el impacto es diferente ya que toma más tiempo entender el comportamiento de las ventas si sólo se observan números. Por lo tanto un análisis es bueno tanto por su objetivo (en este caso el comportamiento de las ventas) como por la presentación de los resultados (a manera de gráfica).

Cuando se habla de *Big Data*, se hace referencia a un gran volumen de información que debe ser analizada, sin embargo los autores Sruthika y Tajunisha (2015) explican que el concepto varía de empresa a empresa, donde para algunos manejar Tera Bytes (TB) se consideran *Big Data* y para otros *Big Data* sería hablar de Peta Bytes (PB). Estos autores mencionan que *“Big Data analytics se refiere al proceso de recolectar, organizar y analizar grandes cantidades de datos que son importantes para un negocio. Para simplificar nuestro entendimiento de Big Data, nos podemos apoyar en tres grandes características”*. Las características que mencionan son el volumen y variedad de los datos y la velocidad con que se generan y procesan. Otros autores como Chawda y Thakur (2016) mencionan otras 2 características extra: Valor y la Veracidad.

Es importante considerar los 5 aspectos dependiendo del análisis que se requiera. Cada una de estas características va a ser particular para cada empresa que aplique una solución de *Big Data*. El volumen se refiere a la cantidad de datos generados, normalmente se habla de conjuntos superiores a 1 TB. La velocidad es un indicador

del volumen de información generada en un tiempo determinado y que tan rápido se procesa esa información. La variedad significa que los datos pueden encontrarse en diferentes fuentes y formatos, por ejemplo: emails, imágenes, videos, bases de datos, documentos, entre otros. La veracidad de los datos se refiere a que los datos sean correctos y confiables, normalmente los datos utilizados no están limpios, pero las herramientas de *Big Data* son capaces de trabajar con datos “sucios”. El valor que pueden ofrecer los datos va a depender mucho del análisis que se quiere realizar, si los datos permiten obtener un resultado para el análisis propuesto entonces se consideran valiosos.

El Doctor Daniel Trejo (2013) agrega 3 propiedades más para construir un esquema más completo de qué es *Big Data* en los negocios. El esquema (Ilustración 1) muestra las características a manera de pilares y refleja claramente que si la solución carece de alguna de ellas entonces no es óptima.

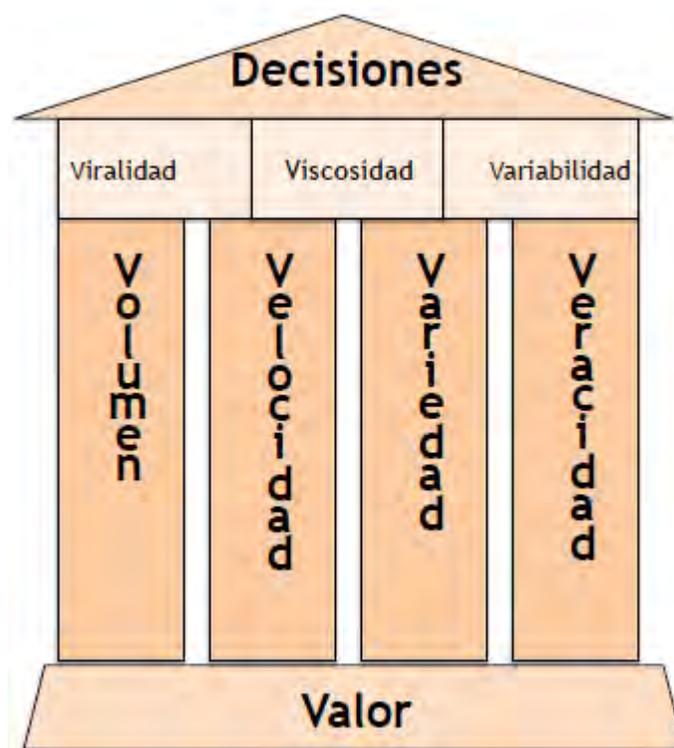


Ilustración 1 Trejo, D. (2013). Modelo de 8 Vs de Big Data. Fuente: Apuntes sobre Big Data.

Big Data se puede entender entonces como una enorme colección de datos procedentes de fuentes diversas y recolectados con cierta frecuencia. El análisis y

procesamiento de estos datos resulta útil para identificar patrones y obtener información relevante respecto de un cierto tema. Para algunas empresas el tema de interés son los hábitos de consumo de sus compradores, para otras podría ser el sentimiento en redes sociales sobre acontecimientos; independientemente del objeto de estudio todo se puede resumir en el análisis oportuno de grandes cantidades de datos que aporta valor en la toma de decisiones.

Existen actualmente muchas empresas que ofrecen soluciones de *Big Data* para que las compañías realicen la introspección necesaria que les brinde información oportuna con el fin de seguir siendo competitivos en el mercado actual. Algunos de los principales competidores en el mercado son *Attivio*, *Watson* (herramienta de *IBM*) y *TIBCO* (principalmente con un generador de analíticos y de tableros llamado *Spotfire*) entre otros.

En general las herramientas ofertadas para soluciones de *Big Data* aprovechan la información de los clientes siguiendo los siguientes pasos:

1. La recolección y clasificación de los datos. Como primer acercamiento se consulta qué clase de información maneja el cliente y qué indicadores se pueden obtener al analizar la información. El objetivo de este paso es conocer y entender los datos que se manejan y qué información pueden ofrecer.
2. El análisis de los datos. Después de definir qué valor pueden ofrecernos los datos que se manejan, se analizan los datos y se genera la información que permite tomar decisiones estratégicas, en otras palabras, se generan los resultados que se van a presentar.
3. La unificación y presentación de la información. Al final se identifica cómo está conectada la información de toda la empresa y se presenta una visualización fácil de digerir y de entender para los clientes (gráficas, tableros, colores, entre otros).

Mercadotecnia digital

Existen muchas aplicaciones para las soluciones de *Big Data*, particularmente se identificó durante las investigaciones que existe una gran oportunidad en el área de *Mercadotecnia digital*.

Los autores Philip Kotler y Gray Armstrong redactaron una visión completa de qué significa hacer *marketing*, estableciendo que está orientado a ocuparse de los clientes más que de alguna otra función de negocio. Como definición se tiene que es el “*proceso mediante el cual las empresas crean valor para sus clientes y generan fuertes relaciones con ellos para, en reciprocidad, captar valor de los clientes*” (Kotler y Armstrong, 2013).

La ventaja principal de realizar un proceso eficiente de *mercadotecnia* está en que los productos desarrollados aportan un valor adecuado a un precio razonable ya que se tienen bien definidas las necesidades y deseos de los clientes, logrando así que el proceso de venta se vuelva más eficiente.

Lo anterior significa que la *mercadotecnia* se encarga de desarrollar relaciones entre un negocio y sus clientes, haciendo que el negocio escuche y entienda las necesidades que tienen sus clientes para ofrecer un producto que las satisfaga adecuadamente.

Con lo descrito en el párrafo anterior un negocio no solamente logra conservar a sus clientes por largos periodos de tiempo, también capta la atención de un mayor número de consumidores pues su producto es recomendado verbalmente de persona a persona consiguiendo así que sean los consumidores los que procuren al negocio.

Kotler y Armstrong ligan el mercado y los clientes a cinco conceptos fundamentales:

1. Necesidades, deseos y demandas. Las necesidades y los deseos están muy relacionados, sin embargo no se deben confundir; una necesidad se refiere a un estado de carencia percibido, por ejemplo una persona “necesita” comer, sentir afecto o adquirir nuevos conocimientos (Kotler y Armstrong, 2013). Por otro lado un deseo es la forma en la que esa persona satisface su necesidad y

dependen de su cultura y personal individual, por ejemplo una persona que necesite comer podría desear una pizza o unos tacos (Kotler y Armstrong, 2013). Cuando un deseo está respaldado por el poder adquisitivo se convierte en “demanda” (Kotler y Armstrong, 2013).

2. Ofertas de mercado. Los deseos y necesidades de los consumidores se satisfacen mediante las ofertas de mercado, que pueden ser productos servicios o experiencias (Kotler y Armstrong, 2013). De ejemplos pueden mencionarse una hamburguesa como producto, un banco como servicio y un viaje turístico como experiencia. Lo óptimo es que la oferta sea integral e involucre, en la medida de lo posible, estos tres aspectos.
3. Valor y satisfacción. Existe una gran variedad de productos y servicios que pueden satisfacer una cierta necesidad, ante esta disyuntiva, los consumidores forman expectativas sobre el valor y satisfacción que debe entregar una oferta para considerarla como la mejor; constantemente eligen productos nuevos, comparan y se quedan con el que aporta más valor a un menor costo (Kotler y Armstrong, 2013). El valor del cliente y su satisfacción son los bloques fundamentales para desarrollar y gestionar relaciones entre el negocio y los consumidores (Kotler y Armstrong, 2013).
4. Intercambios y relaciones. La mercadotecnia se basa en las relaciones de intercambio que satisfacen las necesidades y deseos de las personas (Kotler y Armstrong, 2013). Los mercadólogos procuran provocar una respuesta a una oferta de mercado, que puede ir más lejos que una simple venta, por ejemplo un político que desea obtener votos (Kotler y Armstrong, 2013). Lo que se busca es crear, mantener y hacer crecer las relaciones de intercambio con un público meta (Kotler y Armstrong, 2013).
5. Mercados. Tal como se define en el libro de los autores Kotler y Armstrong (2013), *“desde el punto de vista del marketing, un mercado es el conjunto de todos los compradores reales y potenciales de un producto o servicio ofertado”*.

Estos cinco conceptos complementan la definición dada anteriormente. Para llevar a cabo un proceso eficiente de *marketing*, una empresa debe poner especial atención

en cuáles son las necesidades y deseos de los consumidores, qué características les dan más valor y satisfacción, qué producto pueden ofertar que incluya dichas características, cómo lograr crear y mantener relaciones con sus clientes y finalmente cuál es el segmento de los consumidores a los que va dirigido el producto.

Kotler y Armstrong (2013), exponen que generalmente se piensa en la mercadotecnia como una comunicación unidireccional que va de los vendedores hacia los consumidores, sin embargo los consumidores también realizan *mercadotecnia* en el momento en el que interactúan con las empresas para realizar compras.

Lo que los autores quieren expresar con esta declaración es que los consumidores realizan un proceso de *mercadotecnia* al comparar los distintos productos ofertados y analizar cuál de ellos cubre mejor sus necesidades y deseos. En este caso la relación es creada del consumidor hacia el negocio.

El proceso es una constante interacción entre clientes y negocios en la cual los clientes generan expectativas y los negocios intentan cumplirlas, los negocios mejoran sus productos en la medida que evolucionan las expectativas de los clientes y de esta forma crean nuevas relaciones y conservan las actuales. El siguiente esquema muestra los elementos que componen un sistema de *marketing*.

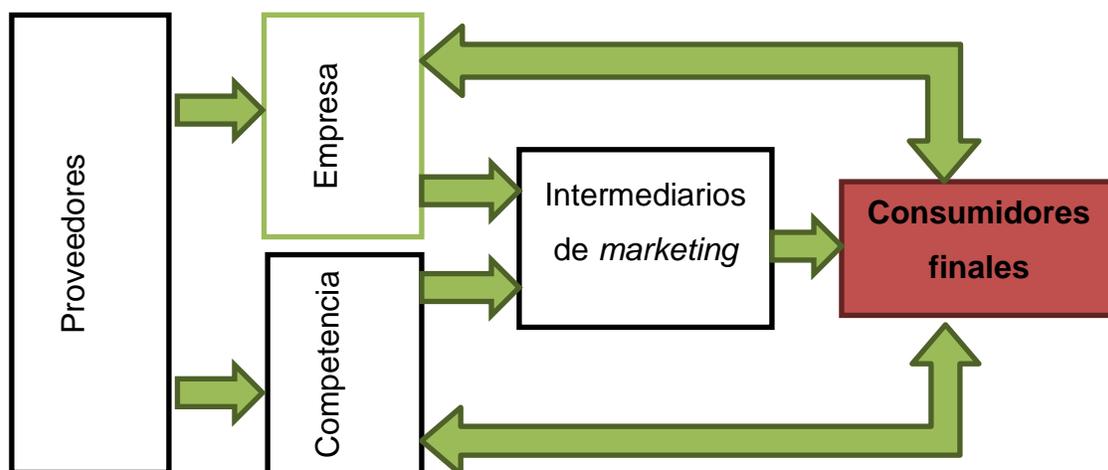


Ilustración 2 Kotler, P. y Armstrong, G. (2013). Sistema de marketing. Fuente: Fundamentos de marketing.

Las flechas representan las relaciones que deben desarrollarse y gestionarse para generar valor del cliente y relaciones rentables con los clientes. Cada componente es importante para agregar valor para el siguiente nivel, por lo que el éxito no depende únicamente de la empresa, por ejemplo en el caso de un supermercado, la tienda no podría ofrecer precios bajos si sus proveedores no le surten mercancía a bajo costo (Kotler y Armstrong, 2013).

De manera concreta, para realizar *mercadotecnia* se necesita conocer cuáles son las necesidades y deseos del mercado, cómo son sus hábitos de consumo y cuál es su cultura, dicho en otras palabras, se requiere analizar una gran cantidad de información que está segmentada y que puede observarse de manera muy granular.

Es aquí donde la tecnología comienza a involucrarse con el *marketing*, pues a mayor cantidad de consumidores existe un mayor volumen de datos que se deben analizar para obtener información estratégica.

El desarrollo de la tecnología ha apoyado el conocimiento que tienen los negocios de sus clientes pues además de facilitar el análisis de los datos, también ha creado nuevas fuentes de las cuales se puede extraer más información.

La era digital ha abierto nuevos canales de comunicación entre las empresas y los consumidores, logrando que los mercadólogos tengan nuevas formas de aprender sobre los consumidores y así crear productos y servicios cada vez más personalizados a las necesidades y deseos individuales (Kotler y Armstrong, 2013).

El internet es el principal canal digital de comunicación que modificó la forma en la que se realiza *mercadotecnia* (Kotler y Armstrong, 2013). Prácticamente cualquier persona está conectada con internet, la mayoría tiene un perfil en alguna de las redes sociales y adquiere información navegando en la *web*.

Como ejemplo de lo anterior se puede considerar como fuente de información una encuesta de satisfacción aplicada después de cada compra, gracias a la cual un negocio llega a conocer el grado de satisfacción de sus clientes.

La encuesta es voluntaria y por lo tanto existe la siguiente interrogante ¿qué opinan todos aquellos que no decidieron tomarla? La tecnología permite a los negocios

ampliar la visión en la percepción de sus clientes ya que muchos de ellos prefieren expresar su opinión a través de redes sociales.

Analizar los comentarios que generan los consumidores en internet permite al negocio tener una retroalimentación que complementa a las encuestas que pueda aplicar.

Esto aumenta considerablemente el volumen de datos que debe procesar para obtener información y por lo tanto debe apoyarse de soluciones tecnológicas que permitan conocer las conclusiones del análisis de manera eficiente y oportuna.

Por otro lado, los procesos de venta y la forma de realizar negocio se ha digitalizado y esto ha permitido que surjan nuevas oportunidades para analizar el comportamiento de los consumidores.

Para este análisis se considera la información que genera una persona en internet de forma indirecta como por ejemplo el tipo de páginas *web* que visita constantemente o los productos que compra a través de tiendas en línea.

Con esta información, las personas encargadas de tomar decisiones con respecto a un negocio son capaces de aplicar estrategias efectivas para mantener y aumentar su cartera de clientes.

Lo mencionado anteriormente ha generado una oportunidad estratégica que involucra el análisis del comportamiento de consumo: personalizar la experiencia de los consumidores.

El mundo de los anuncios digitales y las tiendas en línea ha logrado personalizar las ofertas de los usuarios al mostrar el contenido más relevante a su perfil de consumo.

La empresa Amazon fue pionera en este tipo de experiencia. La tienda analiza productos adquiridos por sus consumidores e identifica qué productos tienen alta probabilidad de ser comprados por la misma persona aún sin estar directamente relacionados.

Por ejemplo una relación directa serían libros del mismo autor o del mismo género, mientras que una relación indirecta sea quizás que una persona de cierta edad que

compró un cierto libro de suspenso tiende a tener afinidad por los grupos musicales de *rock and roll*.

Existen varias soluciones de *software* que apoyan el proceso de *mercadotecnia* de los negocios, particularmente en este trabajo se explora la posibilidad de aplicar un análisis de datos de internet como fuente de información para *mercadotecnia* digital.

La *Mercadotecnia* digital es una parte de la *Mercadotecnia* que se encarga de atender principalmente los canales de comunicación relacionados con tecnologías de la información (por ejemplo internet) que tiene una empresa con su segmento meta (Kotler y Armstrong, 2013).

Basado en lo descrito anteriormente, publicar un sitio *web* en internet abre un canal de comunicación importante entre un negocio y sus clientes. Particularmente esta investigación está orientada a descubrir patrones de uso en el sitio *web* de una empresa con el fin de obtener información que apoye al proceso de *mercadotecnia* digital.

Para aplicar *mercadotecnia* digital de manera efectiva es importante considerar cuál es el objetivo de negocio que se quiere lograr. Este objetivo es la respuesta a la pregunta ¿por qué existe mi sitio *web* en internet? El ejemplo más claro para esto podría ser un *e-commerce*, el cual tendría por objetivo principal de negocio la venta de productos en línea.

Avanish Kaushik (2009) explica que estos objetivos deben características *Doable. Understandable. Managable. Beneficial (DUMB)*, lo que significa que deben ser factibles, entendibles, manejables y benéficos. También sugiere que se contemple una visión amplia y global del negocio ya que en muchos casos si no están bien definidos los objetivos, el análisis puede mostrar un falso fracaso del negocio.

En el ejemplo del *e-commerce*, se pueden considerar más objetivos que amplían el criterio de éxito, como la suscripción de usuarios o la publicidad consultada.

Una vez que se han definido claramente los objetivos se analiza qué métricas nos sirven de indicadores sobre el fracaso o éxito de esos objetivos.

Estadísticas y métricas para sitios web

Los autores Jason Burby y Angie Brown en colaboración con el resto del comité de estándares de la asociación de analíticos web (*Web Analytics Association, WAA*) publicaron en 2007 un documento que conjunta las definiciones más importantes sobre analíticos importantes que deben considerarse para un sitio web.

En la primera sección, los autores definen los siguientes términos sobre analíticos en general:

“Existen 3 tipos de métricas de analíticos web – cuentas, razones y KPIs:

- **Cuenta.** *Es la unidad más básica de medida. Un número generalmente entero (Visitas = 12,398), pero no necesariamente (Total de Ventas = \$52,126.37).*
- **Razón.** *Generalmente una cuenta dividida por otra cuenta, aunque una razón puede utilizar ambos una cuenta o razón en el numerador o denominador. [...] Al ser una razón la palabra ‘por’ se encuentra usualmente en el nombre como en ‘Visitas a Páginas por Sesión’ [...]*
- **KPI (Key Performance Indicator o Indicador Clave de Desempeño).** *Un KPI puede ser una cuenta o una razón, frecuentemente es una razón. [...] un KPI considera una estrategia de negocio – de ahí el término ‘Clave’ – y por lo tanto el conjunto de KPIs apropiado normalmente varía entre sitios y tipos de procesos.*

Un cuarto tipo de definición se incluye por términos que describen conceptos en lugar de números.

- **Dimensión.** *Una fuente de datos que puede ser usada para definir varios tipos de segmentos o cuentas y representa una dimensión fundamental del comportamiento del visitante o dinámica del sitio. Algunos ejemplos son eventos y referencias. Se pueden interpretar igual que las cuentas mencionadas pero normalmente deben ser más segmentadas para ser de interés real. Por lo tanto, éstas definen una clase de métricas más general y representan una dimensión de datos que puede ser asociada con cada visitante de manera individual.*

Las métricas son medidas en dimensiones.

Una métrica puede aplicar a tres diferentes universos:

- **Agregado.** *Tráfico total del sitio para un periodo de tiempo definido.*
- **Segmentado.** *Un subconjunto del tráfico del sitio para un periodo de tiempo definido, filtrado de alguna manera para ganar mayor introspección analítica: por ejemplo, por campaña (e-mail, banner, PPC), por visitante (nuevo vs. no nuevo, compradores repetidos), por referencia*
- **Individual.** *Actividad de un único visitante Web en un periodo de tiempo definido.”*

A continuación se mencionan los términos más relevantes que fueron definidos por el comité de estándares de la WAA que se distribuyen en cuatro categorías:

1. Términos base:

- **Sesiones.** Una sesión es una interacción entre un individuo y el sitio web que consiste en varias peticiones de contenido (páginas). Si el individuo no realiza ninguna otra acción relevante en el sitio por un periodo de tiempo especificado (normalmente 30 min), la sesión se da por terminada. Como métrica se toma la cuenta del número de sesiones generadas en el periodo analizado.
- **Visitantes.** Un visitante es un individuo “humano” (se excluyen *bots* y *crawlers*) que interactúa con el sitio navegando a través de sus páginas o realizando otro tipo eventos. Se pueden categorizar considerando si ya han realizado visitas anteriormente. Se cuenta el número de visitantes que llegan al sitio en el periodo analizado.
- **Visitas a páginas.** Un sitio está compuesto por varias páginas y una página se define como una unidad de contenido. Se cuentan las veces que cada recurso o página fue consultada por los usuarios.

2. Caracterización de la visita:

- **Página de entrada.** Es la página en la que un visitante inicia una sesión. Es una dimensión que ayuda a analizar las visitas a páginas.

- **Landing page.** Es la página que originalmente está diseñada como el inicio de la experiencia de un visitante en el sitio, es la página de bienvenida al sitio. Es una dimensión que aporta al análisis de las visitas a páginas.
- **Página de salida.** Es el último recurso consultado por el visitante antes de terminar su sesión. Es una dimensión que ayuda a analizar las visitas a páginas
- **Duración de sesión.** El tiempo que un usuario permanece en interacción con el sitio. Es un valor numérico normalmente contabilizado en minutos.
- **Referencia.** Es la URL de la página que ocasionó que el visitante realizara una petición al recurso actual que está consultando. Explicado de otra forma, es la visita anterior a la página que se está visitando actualmente. Existen casos en los que no se tiene una referencia definida, por ejemplo si el usuario conoce de memoria la liga de acceso al sitio la petición se realiza de manera directa y no existe una visita anterior. Se catalogan en referencias internas, externas y de búsqueda.
- **Número de clics.** Se contabilizan los clics que los usuarios realizan sobre los enlaces de una página y sirve como métrica para analizar la interacción del visitante. Se puede considerar como cuenta o como una razón.
- **Páginas visitadas por sesión.** Es la razón del número de visitas a páginas entre el número de sesiones totales.

3. Caracterización de contenido.

- **Razón de página de salida.** Es la razón del número de veces que una página es la última visitada en una sesión entre el número de visitas que tiene esa misma página.
- **Sesiones de una vista de página (Rebotes).** Es la cantidad de sesiones que consisten en una única página visitada. El visitante llega al sitio y lo abandona sin realizar alguna otra interacción con él.

- **Porcentaje de rebote.** Es la razón de la cantidad de rebotes dividida entre la cantidad de sesiones.

4. Métricas de conversión.

- **Evento.** Se define como una acción que sea registrada con una fecha y hora. Como métrica se analiza la cuenta del número de veces que ocurre. Un ejemplo puede ser la visita a una página o la descarga de cierto archivo.
- **Conversión.** Una conversión ocurre cuando un visitante realiza un evento específico que se tiene como objetivo de la página. Se puede medir como cuenta o razón dependiendo del análisis que se realice o incluso puede usarse como una dimensión (por ejemplo, sesiones en las que se tienen conversiones). Un ejemplo sería un *e-commerce* donde el objetivo es que los visitantes realicen una compra y la conversión ocurre cuando un visitante compra algún producto.

Cada una de estas métricas es importante para detectar los aspectos buenos y malos de un sitio *web*. No todas las métricas van a representar algo indispensable para todos los negocios pero la mayoría aportan información valiosa con la cual pueden adoptarse estrategias para mejorar, el análisis va depender de los objetivos de negocio que se definan.

Adicionalmente, resulta muy valioso detectar cómo está segmentado el conjunto de usuarios que tienen interacción con el sitio. Esta segmentación se utiliza principalmente para aplicar una estrategia de venta efectiva que logre captar nuevos consumidores además de conservar a los clientes frecuentes.

Algunas de las características que se consideraron para el análisis que realiza la solución desarrollada, son las características tecnológicas con las que el usuario interactúa con el sitio. Concretamente se adquiere información de qué navegador y qué tipo de dispositivo (celular, *tablet* o pc) utiliza para navegar, esto resulta muy importante porque se puede detectar si el diseño del sitio es adecuado para el tipo de dispositivo que utilizan sus usuarios.

Metodologías ágiles

Como ya fue mencionado, el desarrollo de los sistemas informáticos fue evolucionando conforme aumentaba la complejidad de los problemas a resolver, en un principio no existía un protocolo o guía que indicara los pasos a seguir para garantizar un desarrollo óptimo y de calidad. Esto no fue necesario en un principio porque la simpleza de las soluciones permitía simplemente realizar programas fáciles de usar y que no requerían de conocimientos tan avanzados.

Poco a poco comenzaron a surgir algunos problemas que causaban insatisfacción en las personas que utilizaban un *software* desarrollado por un tercero. Principalmente se desconocía el uso y la tecnología implicada, además de que no siempre realizaba los procesos requeridos de forma correcta.

Fue así como surgió una época conocida como la crisis del *software* que fue un periodo de tiempo en el que surgieron varios términos como el de Ingeniería de *software* que fueron dando más formalidad a la producción de soluciones tecnológicas. La crisis del *software* fue un momento en la historia cuando no existían estándares para el desarrollo de *software*, los productores ponían poco esfuerzo en garantizar la satisfacción de sus consumidores y no se tenía un marco de referencia para exigir una calidad mínima en las soluciones adquiridas (Sommerville, 2011).

Esto fue evolucionando y comenzaron a surgir estándares y metodologías de desarrollo que garantizaban productos funcionales y que cubrían en un mayor grado los requerimientos de los clientes.

El término Ingeniería de *software* fue formalmente publicado en 1990 en el estándar 610.12 de la *IEEE* definiéndose como “*la aplicación de un enfoque sistemático, disciplinado y cuantificable para el desarrollo, operación y mantenimiento del software*” (Sommerville, 2011).

El objetivo principal fue cubrir de manera más eficiente las necesidades de la sociedad que constantemente eran identificadas como áreas de oportunidad para la aplicación de la tecnología.

Actualmente e independientemente de la metodología de desarrollo aplicada; realizar Ingeniería de *software* implica 5 fases principalmente: toma de requerimientos, diseño, implementación, pruebas y corrección de defectos. Cada parte del proceso involucra una documentación que garantiza un mejor control de la administración del proyecto y en el enfoque tradicional, las soluciones se desarrollan en cascada lo que significa que las fases se ejecutan en orden y no se inicia una sin haber terminado y validado la anterior.

Este enfoque representa un conflicto con la naturaleza de los proyectos de *software* pues es normal que los requerimientos cambien constantemente a lo largo del ciclo de vida del proyecto y además es prácticamente imposible extraer todas las características y detalles implicados en la solución desde el inicio del proyecto, por lo general los clientes sólo tienen una idea vaga del producto que podría atender a sus necesidades de negocio. Modificar los requerimientos iniciales en este modelo implica repetir el ciclo de vida desde la primera fase, se debe actualizar la documentación para incluir las nuevas peticiones y generar una evidencia del impacto que éstas ocasionan sobre el desarrollo del proyecto. El desarrollo en cascada busca siempre tener los requerimientos completos desde el inicio para evitar modificaciones, pues los cambios implican un re trabajo exhaustivo.

El desarrollo en cascada causaba mucho descontento por parte de los clientes en proyectos muy grandes (más de 6 meses) ya que el valor obtenido no se apreciaba a corto plazo. Para ofrecer mayor calidad, los desarrolladores adoptaron una metodología incremental, en la cual un proyecto se dividía en varios sub proyectos y por lo tanto se definía más de un entregable, obteniendo valor en etapas tempranas de desarrollo.

El proceso continuó adaptándose y en 2001 surgió una ideología de desarrollo llamada “ágil”. Los desarrollos ágiles estaban basados en un manifiesto redactado y firmado por varios críticos de modelos de mejora del desarrollo de *software*, quienes identificaron 4 postulados principales (Fowler y Highsmith, 2001):

Individuos e interacciones sobre procesos y herramientas.

Software funcionando sobre documentación extensiva.

Colaboración con el cliente sobre negociación contractual.

Respuesta ante el cambio sobre seguir un plan.

De manera breve, el objetivo de las metodologías ágiles es que el desarrollo aporte valor al negocio desde el inicio del ciclo de vida del proyecto y enfocarse principalmente a las necesidades del cliente en lugar de darle prioridad al proceso del proveedor.

Existen varias metodologías ágiles, las dos más conocidas son *eXtreme Programming (XP)* y *Scrum*. Se optó por desarrollar la solución propuesta aplicando la metodología *Scrum*.

“Scrum es una metodología adaptable, iterativa, rápida, flexible, y efectiva diseñada para entregar valor significativo de forma rápida y constante a lo largo de un proyecto. Scrum asegura la transparencia en la comunicación y crea un ambiente de responsabilidad colectiva y progreso continuo (Satpathy et. al., 2016)”.

Scrum divide el trabajo a realizar en esfuerzos cortos e iterativos llamados “*sprints*”, busca conformar un equipo empoderado y auto-organizado que realiza las actividades dándoles prioridad con base en qué aspectos aportan mayor valor a las necesidades del cliente (Satpathy et. al., 2016).

La metodología implica que el control de procesos se lleve a cabo de manera empírica, de tal forma que se adapten con base en la observación en lugar de ejecutar un plan riguroso (Satpathy et. al., 2016).

Con respecto al tiempo, *Scrum* lo considera como una restricción limitada e intenta definir el plazo para la realización de las actividades de tal forma que se marque un ritmo de trabajo constante para todos los involucrados en el proyecto (Satpathy et. al., 2016).

Scrum define 4 roles principales que influyen en el desarrollo de los proyectos. Por un lado, se tiene un *Product Owner* que es la persona que aporta una necesidad de negocio definiendo la prioridad de las características a crear, representando así la voz del cliente durante el proyecto. Quienes ejecutan el desarrollo son los miembros del equipo (*Scrum Team*), que son coordinados por un *Scrum Master*, cuya función

es facilitar la comunicación entre el *Scrum Team* y el *Product Owner*, además de atender las necesidades que tenga el equipo para realizar las actividades del proyecto (Satpathy et. al., 2016). Alrededor del proyecto existen más personas involucradas como los usuarios finales o personas que pueden obtener un beneficio del producto, por ejemplo, suponiendo un sistema de cobro para una tienda de autoservicio un usuario final sería el cajero que realiza el cobro y otro involucrado, que no necesariamente es usuario final, sería el cliente que efectúa un pago. El término que utiliza *Scrum* para todos los demás involucrados en el proyecto es “*stakeholders*” (Satpathy et. al., 2016).

Scrum utiliza épicas e historias de usuario para redactar los requerimientos del proyecto. Cada épica define de manera genera una parte del producto y las historias de usuario se van a un nivel más detallado, especificando ¿qué? quiere ¿quién? y ¿para qué lo necesita? (Satpathy et. al., 2016).

La Ilustración 2 muestra un diagrama del proceso que se debe seguir cuando se aplica la metodología. Primero se realizan reuniones con el cliente para entender y definir cuál es la necesidad detectada, esto se llama “*Project Vision Statement*” o en español “Enunciado de visión del proyecto”. A partir de este enunciado se realiza una lista de épicas, que posteriormente se descomponen en historias de usuario, los cuales conforman el “*Product Backlog*” que debe ser priorizado entre el *product owner* y el *scrum master* para distribuir el desarrollo en *sprints* y calendarizar los entregables del proyecto. Una vez que se definió el calendario se ejecutan iterativamente el número de *sprints* definidos (Satpathy et. al., 2016).

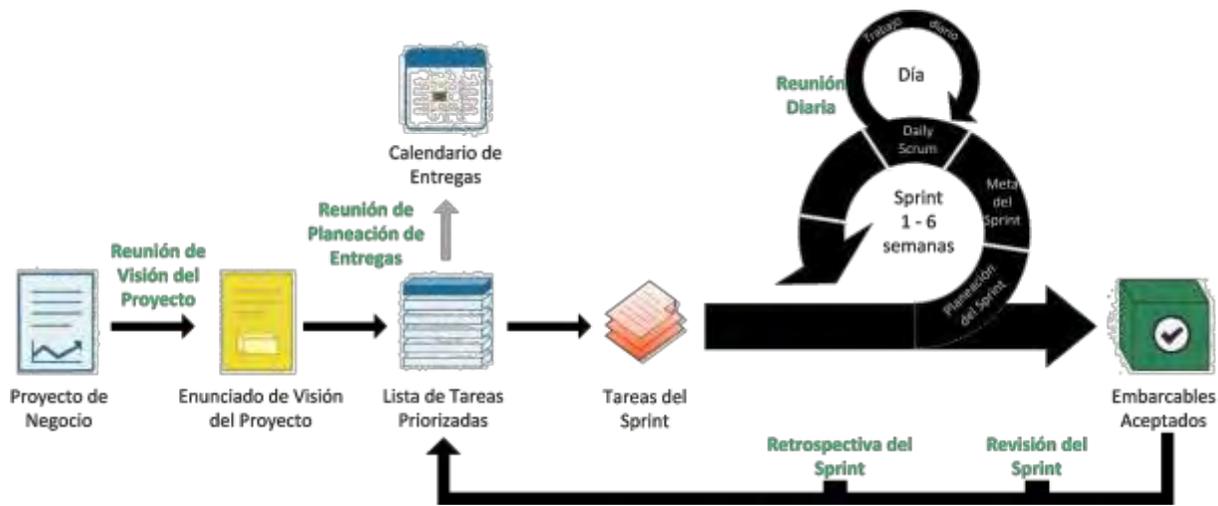


Ilustración 3 Satpathy, T. et. al. (2016). Flujo del Scrum. Fuente: SBOK.

El *sprint* inicia tomando las épicas estimadas para el mismo, se definen actividades a realizar para completar las historias de usuario de acuerdo a las expectativas del *product owner* y se ejecutan hasta haber cumplido con la definición de entregable terminado provista por el *product owner* y el *scrum master* (Satpathy et. al., 2016).

Diariamente se realizan reuniones breves (15 min máximo) para revisar el estado del proyecto y atender los riesgos que se presenten, es responsabilidad del *scrum master* asegurarse que el equipo esté progresando de acuerdo con la estimación inicial y resolver los impedimentos que el equipo vaya encontrando. Durante estas reuniones se revisa qué actividades se terminaron desde la reunión anterior, cuáles se tiene planeado terminar para la siguiente y finalmente los riesgos detectados que provoquen atrasos en el desarrollo (Satpathy et. al., 2016).

Cada *sprint* inicia con una reunión que tiene por objetivo informar al equipo de trabajo sobre las necesidades del cliente, qué historias de usuario abarca el *sprint* y con cuanto tiempo se dispone para ejecutarlo. Al finalizar la ejecución del *sprint* se realiza una reunión de cierre y retrospectiva del *sprint* donde se revisa la ejecución para detectar qué se hizo bien y qué puede mejorarse en las siguientes iteraciones.

Para poder dar por terminado un *sprint* el *product owner* revisa los entregables para validar que cumplen con los criterios establecidos al inicio, una vez validado esto se

revisa también el *product backlog* para actualizar la prioridad de las épicas e historias de usuario que están pendientes (Satpathy et. al., 2016).

DESARROLLO

Con base en las investigaciones realizadas sobre mercadotecnia digital y estadísticas *web*, se pensó en la oportunidad de desarrollar un proyecto que resuelva la necesidad de indicadores de éxito en un sitio *web*.

Como resultado se obtuvo la definición formal de qué aspectos se deben considerar como mínimo para que la solución aporte valor en ésta área, siguiendo los pasos de la metodología *Scrum*.

Primero se identificaron algunos roles principales.

Como **product owner** se consideró al encargado de llevar el sitio *web* de una empresa que actualmente no utiliza ninguna solución para obtener estadísticas *web*.

Posteriormente se estudió el caso de negocio para identificar cuáles son las necesidades principales que el proyecto debe cubrir y así redactar el siguiente enunciado de visión de proyecto:

Se busca analizar el tráfico generado en un sitio web con el fin de generar información estadística que ayude a estudiar el comportamiento de los usuarios en el sitio así como a generar estrategias de marketing.

A partir de esto se definió el **Scrum Master** del proyecto, que según la metodología debe ser un miembro del equipo de trabajo que sea capaz de facilitar la comunicación y los recursos entre el equipo y el *product owner*, es decir que es la persona encargada de atender los impedimentos que surjan para el equipo y asegurar que el proceso de *Scrum* se siga correctamente (Satpathy et. al., 2016).

Además se identificaron también a los **Stakeholders**. Éstos son personas que se ven afectadas positiva o negativamente por el desarrollo del proyecto.

En este caso se observó que el proyecto impacta principalmente a: la empresa que desea obtener métricas sobre sus sitio *web*, los usuarios del sitio *web*, el área de mercadotecnia de la empresa, el área de tecnologías de la información (TI) de la empresa y potencialmente cualquier otro negocio que maneje un sitio *web*.

El siguiente paso fue seleccionar a los miembros del **Scrum Team**. Como el proyecto fue desarrollado en conjunto con la empresa, el equipo de trabajo incluía algunos consultores de la misma.

Continuando con la definición del proyecto, se realizó la creación de un **product backlog**, el cual está compuesto entre otras cosas por una lista de **epics** que fue priorizada tomando en cuenta principalmente las necesidades de negocio y las limitaciones técnicas.

Cada *epic* es un enunciado que define una funcionalidad general que debe realizar el sistema y que posteriormente se debe descomponer en historias de usuario (Satpathy et. al., 2016).

Para construir el *product backlog*, se realizó una reunión entre el equipo de trabajo y el *product owner* en la que se identificaron las siguientes *epics*.

No.	Epic
E1	Como analista de mercadotecnia digital deseo una herramienta que genere indicadores del uso de mi sitio <i>web</i> para identificar qué estrategias ayudarán a crecer el negocio.
E2	Como director de mercadotecnia quiero un módulo para delegar el análisis de varios sitios <i>web</i> a distintos analistas, además de restringir el acceso sólo a los sitios que estén autorizados.
E3	Como analista de mercadotecnia deseo una herramienta para identificar cuál es el perfil de usuarios que entran a mi sitio <i>web</i> .

Tabla 1 Lista priorizada de *epics* del proyecto. Fuente: elaboración del autor.

Adicionalmente, *Scrum* propone describir **personas** que habrán de interactuar con el producto final, esto con el fin de definir claramente cuál es el perfil general de los usuarios.

La especificación redactada en el SBOK explica que se debe crear un personaje ficticio con las características del segmento meta del producto (Satpathy et. al., 2016). Esto se sugiere para que el equipo de trabajo tenga presente las consideraciones especiales para los usuarios.

La siguiente tabla muestra a las personas definidas para el proyecto.

No.	Persona
P1	Francisco es un hombre de 32 años que se dedica a escribir un blog en internet. Actualmente no tiene forma de conocer cuántos lectores activos tiene y no sabe si este número ha aumentado desde que inició.
P2	Victoria es una mujer de 28 es analista de mercadotecnia digital de una empresa que se dedica a edición de contenido de varias revistas. A ella le gustaría conocer cuáles de las revistas que tiene a su cargo son las más populares, además de saber cuál es el perfil que tienen sus lectores.
P3	Carlos es un hombre de 38 años que maneja un sitio <i>web</i> . Él desea monetizar el uso de su sitio vendiendo espacios de publicidad, sin embargo desconoce cuánto podría cobrar por dichos espacios.

Tabla 2 Definición de personas. Fuente: elaboración del autor.

Por otro lado, es muy importante que las partes involucradas (cliente y proveedor) en el proyecto tengan conocimiento de cuáles son los criterios que al cumplirse los elementos del proyecto se consideran como terminados.

Esto ayuda a que el equipo de desarrollo asegure que el producto final tenga una calidad alta y que con esto el cliente quede satisfecho.

También es útil para que el cliente pueda apreciar desde el inicio de la planeación cuál será el valor que puede esperar al concluir el proyecto y que de esta manera pueda aprobar o rechazar los entregables.

Scrum nombra a estos criterios como **done criteria** o criterios de terminado. La definición de estos criterios se muestra en la Tabla 3.

No.	Criterio
CT1	La solución estará instalada en un servidor con acceso a internet y podrá ser consultada a través del dominio https://blackrabbit.dsasoluciones.com
CT2	Se podrán dar de alta usuarios y sitios y la solución debe restringir el acceso de los usuarios a la información de los sitios que tiene asociados.
CT3	El cálculo de las estadísticas será validado por el equipo de pruebas.

Tabla 3 Criterios de terminado del proyecto. Fuente: elaboración del autor.

Después de haber definido los criterios de terminado, se realizaron algunas reuniones para planear la calendarización de las entregas.

Esto incluye la definición del tiempo que habrá de durar el proyecto la cual se realiza dividiendo el desarrollo en **sprints**. Según *Scrum*, un *sprint* es un esfuerzo limitado a un tiempo específico que puede ir de una a seis semanas y que se realiza para cumplir con una parte funcional del proyecto (Satpathy et. al., 2016).

Con base en esta definición, se estimó el proyecto para una duración de 8 semanas, por lo que se tomó la decisión de repartir el desarrollo del proyecto en dos sprints de cuatro semanas cada uno.

En este punto ya se tiene una definición general de cuáles son los requerimientos del proyecto, una visión de cuál es el valor esperado por el cliente y de cuánto tiempo se dispone para completar el desarrollo.

El siguiente paso fue ejecutar cada uno de los *sprints*, cada uno conlleva la realización de una reunión de planeación para crear el *backlog* del *sprint* además de la definición y estimación de las tareas, reuniones diarias para llevar un control del avance realizado, elaboración de las tareas estimadas, validación de los entregables con los criterios de terminado y por último una reunión retrospectiva que se realiza para identificar los aspectos positivos y negativos de cada *sprint* con el fin de tener un mejor desempeño en iteraciones siguientes (Satpathy et. al., 2016).

El *backlog* de cada *sprint* contiene una definición más detallada y a nivel más técnico de los requerimientos que deben ser cubiertos al concluir el *sprint*.

Dichos requerimientos se redactan a manera de historias de usuario con la estructura ¿quién? necesita o quiere ¿qué cosa? para ¿qué?; y son desglosadas a partir de las *epic(s)* consideradas para el *sprint* en curso (Satpathy et. al., 2016).

A continuación se describe la ejecución de los dos *sprints*.

Sprint 1

Para el primer *sprint* se acordó entregar una solución que atendiera la *epic* E1, la cual está descrita en la Tabla 1.

Las historias de usuario redactadas son las siguientes.

No.	Historia de usuario
H1	Como analista de mercadotecnia necesito una pantalla que me permita dar de alta sitios para calcular sus estadísticas <i>web</i> .
H2	Como analista de mercadotecnia quiero un componente que dé seguimiento al uso de mi sitio <i>web</i> para obtener estadísticas <i>web</i> .
H3	Como analista de mercadotecnia necesito visualizaciones del análisis del uso de sitios <i>web</i> , con el fin de que la información generada sea fácil de analizar
H4	Como analista de mercadotecnia quiero un repositorio de información que me permita observar las estadísticas <i>web</i> en distintos periodos de tiempo.
H5	Como encargado de un sitio <i>web</i> quiero un componente para facilitar la instalación de la solución en mi sitio <i>web</i> .

Tabla 4 Historias de usuario del sprint 1. Fuente: elaboración del autor.

A partir de ellas se crearon y estimaron tareas, dando como resultado la siguiente gráfica de tipo *burn up* que muestra una calendarización estimada del avance del proyecto.

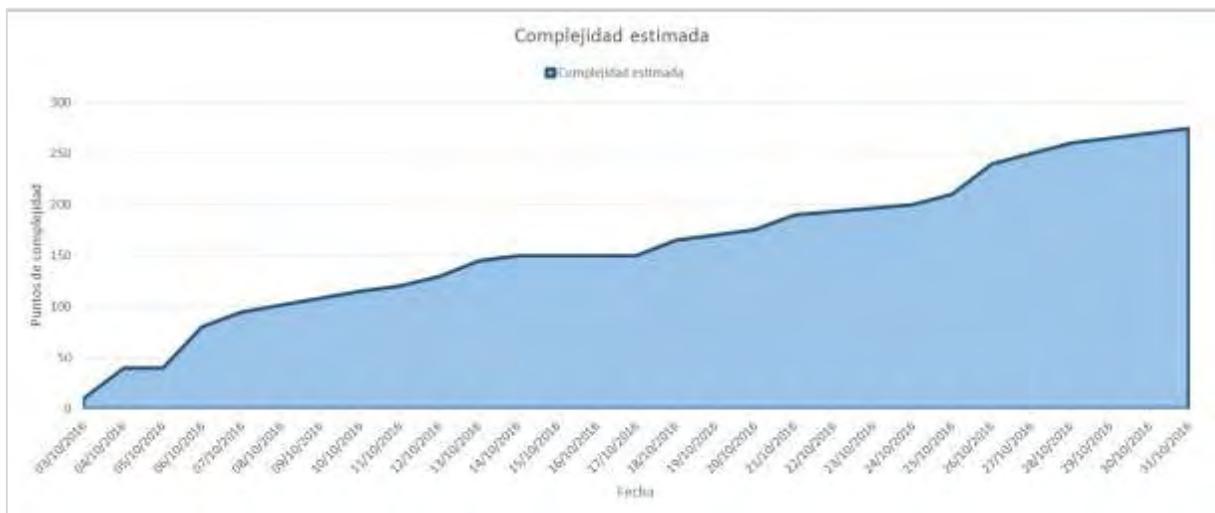


Ilustración 4 Burn up de complejidad estimada del sprint 1. Fuente: elaboración del autor.

La ejecución de las tareas dio como resultado lo expuesto en la sección de resultados.

En la reunión retrospectiva del *sprint* se comentaron las siguientes lecciones aprendidas.

No.	Lecciones aprendidas
L1	El desarrollo de <i>mockups</i> ayudó a estandarizar las vistas, pero creo que sería mejor si se realizaran también las pantallas en HTML, pues detalles como el tamaño de un botón, una tabla, estilos particulares y colores específicos todavía podrían variar. También la estandarización de mensajes fue una buena idea, sin embargo, creo que se pudo haber acelerado si se buscaban desde el comienzo los mensajes de error propios de Java (en inglés y en español) y se modificaban a nuestro criterio (cosa que no se hizo pues no se conocía a tal profundidad la herramienta)
L2	Quedaron varias ambigüedades en la definición de los <i>mockups</i> , considero que debería estandarizarse mejor desde el principio detalles como los colores o la posición de los filtros. Se podrían estandarizar los mensajes de error y de éxito de las vistas. Las ambigüedades deben consultarse con el <i>scrum master</i> para ahorrar tiempo de desarrollo, pues es él quien tiene el contacto directo con el <i>product owner</i> . Definir una nomenclatura ahorra tiempo de creación de código. Me pareció muy bueno que se decidiera en equipo las reglas de programación para que en caso de que alguien más tome algún módulo sea más fácil entender el código.
L3	Me parece que cuando se trabaja con temas visuales nuevos toma más tiempo que si se utiliza algo ya conocido.
L4	Mediante la creación de <i>mockups</i> se logró notar una mejoría en el proceso de creación de vistas y se pudo estandarizar los diseños de las pantallas. Por lo cual se precisa ser una buena práctica a implementar en futuros proyectos. A pesar de las afectaciones provocadas por las actividades fuera del proyecto, el equipo sacó adelante las tareas con muy poco retraso respecto al planeado.

Tabla 5 Lecciones aprendidas del sprint 1. Fuente: elaboración del autor.

Sprint 2

En el segundo sprint se construyó la funcionalidad que cumple con lo descrito en las *epics* E2 y E3.

Las historias de usuario redactadas son las siguientes.

No.	Historia de usuario
H4	Como director de mercadotecnia deseo una pantalla para dar de alta analistas de mi equipo.
H5	Como director de mercadotecnia deseo una pantalla para asociar a mis analistas con los sitios <i>web</i> a los que están

	autorizados a dar seguimiento.
H6	Como usuario quiero una pantalla para autenticarme y que la solución me de acceso a las funciones que estoy autorizado a realizar.
H7	Como analista de mercadotecnia quiero una pantlla que muestre de manera visual cuál es la segmentación de personas que utilizan el sitio <i>web</i> .
H8	Como encargado del sitio <i>web</i> quiero una pantalla que muestre cuál es el común de la navegación de los usuarios para saber si las páginas más relevantes son tan visitadas como se espera

Tabla 6 Historias de usuario del sprint 2. Fuente: elaboración del autor.

Con base en éstas se crearon y estimaron tareas, cuyo esfuerzo calendarizado se describe en la siguiente gráfica de tipo *burn up*.

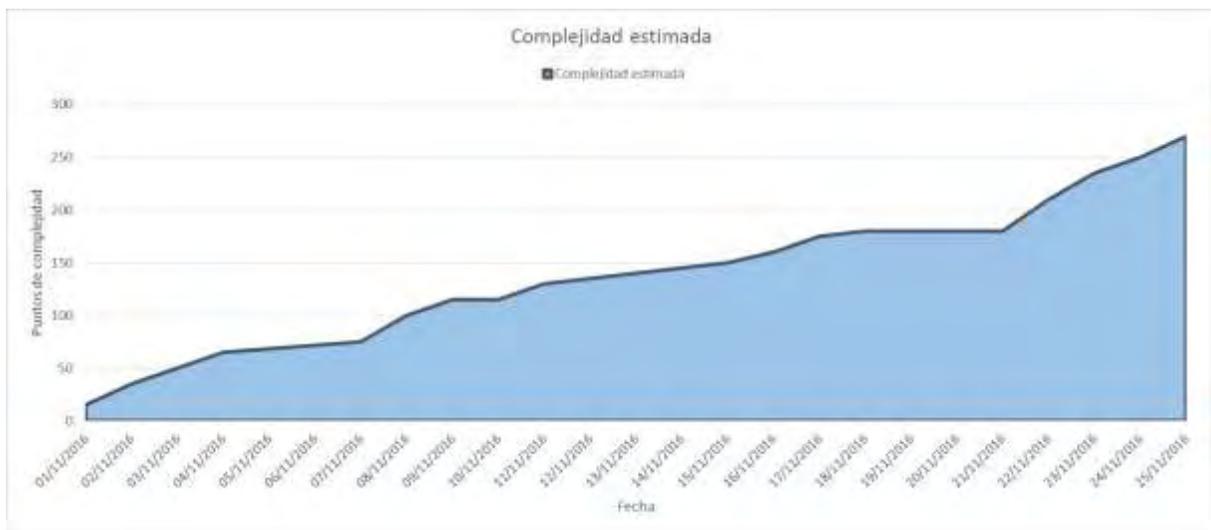


Ilustración 5 Burn up de complejidad estimada del sprint 2. Fuente: elaboración del autor.

Los entregables desarrollados a partir de la ejecución de las tareas están detallados en la sección de resultados.

Durante la reunión retrospectiva del *sprint* se comentaron las siguientes lecciones aprendidas.

No.	Lecciones aprendidas
L1	Utilizar los componentes que incluye el framework ahorra tiempo de desarrollo en las vistas, sin embargo no todos los componentes tienen una visualización que concuerde con la plantilla de la solución. Es necesario considerar esto en la planeación de tareas para futuros proyectos.

L2	Las pruebas unitarias en el desarrollo evitan defectos y reducen el tiempo de tareas de pruebas y de correcciones por lo que es importante no liberar un módulo sin haberlo probado.
L3	Las ambigüedades en los requerimientos deben resolverse directamente con el <i>product owner</i> , al hacer esto surgirán cambios pero si dichos cambios serán autorizados y además no se realizará trabajo extra por corregir malas suposiciones.

Tabla 7 Lecciones aprendidas del sprint 2. Fuente: elaboración del autor.

RESULTADOS

Entregables del *sprint* 1

Las primeras actividades del *sprint* estuvieron orientadas al diseño del sistema.

Como primer paso, se analizaron cuáles eran las características más importantes para una solución que analice el uso de sitios *web* dando como resultado el siguiente diagrama de la arquitectura del *software*.

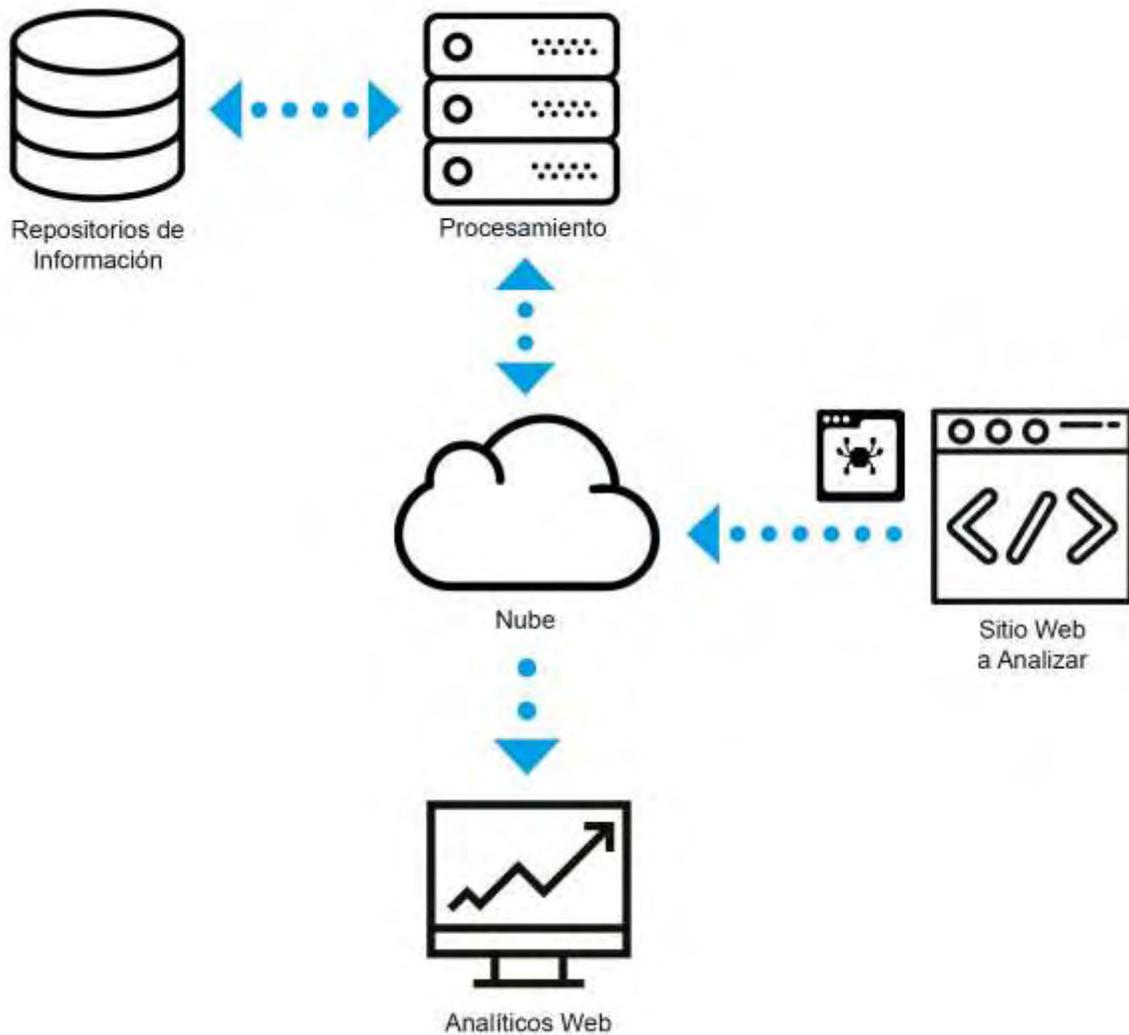


Ilustración 6 Arquitectura de la solución. Fuente: elaboración del autor.

Posteriormente se definieron las herramientas a utilizar como repositorios de información.

Por un lado se optó por una base de datos OLTP para el manejo de los módulos administrativos del sistema.

Además se utilizó una base de datos *NoSQL* orientada a documentos para guardar el registro de visitas al sitio *web* y extraer de manera eficiente las métricas del sitio.

Adicional a los repositorios de información se desarrolló un componente *web* que se comunica vía un servicio *web* con la solución para realizar el registro de las visitas de usuarios al sitio *web*.

Finalmente, se construyó también un sitio que permitiera realizar las configuraciones a la herramienta y que adicionalmente mostrara los analíticos calculados.

Para esto, se diseñaron las siguientes pantallas que cumplen con las historias de usuario del *sprint*.

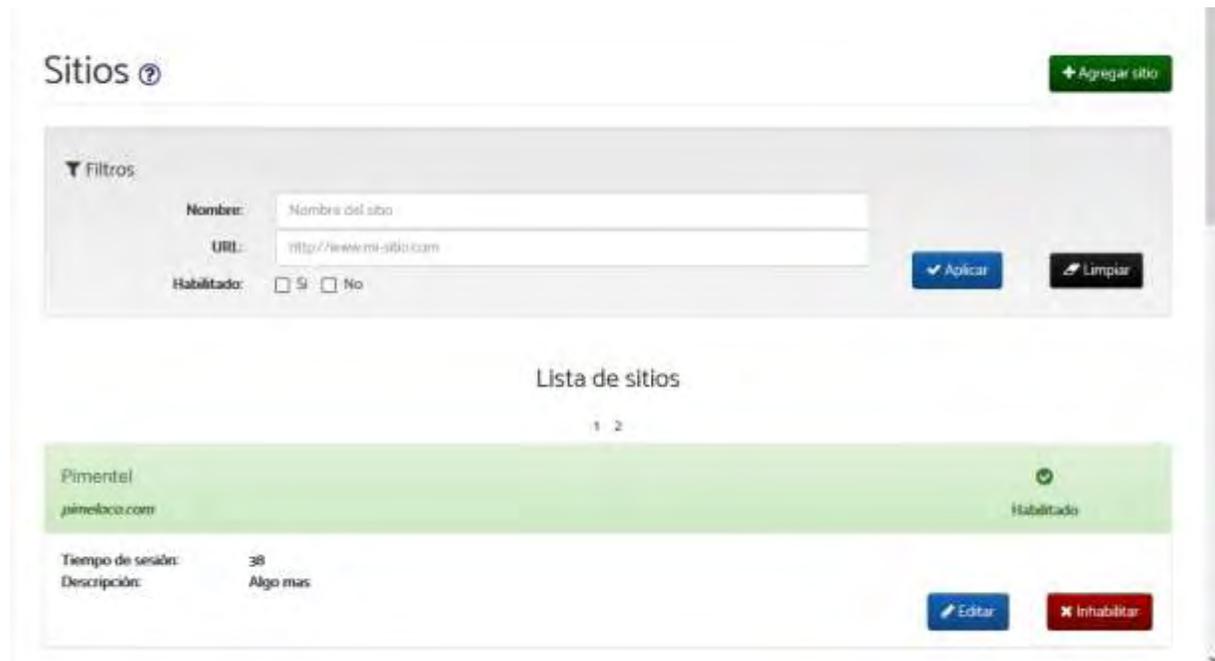


Ilustración 7 Administración de sitios. Fuente: elaboración del autor.

Desde esta pantalla es posible registrar un sitio *web* para generar el código que debe incluirse dentro del mismo.

Además, el usuario puede observar cuáles son los sitios *web* que ya ha registrado anteriormente así como editar su información o inhabilitar el seguimiento.



Ilustración 8 Analíticos principales. Fuente: elaboración del autor.

En esta vista el usuario puede observar cómo fue el uso de su sitio en un cierto periodo de tiempo. Aquí mismo es posible filtrar la información de tal forma que se obtenga la información de una página en lugar del sitio en general



Ilustración 9 Gráfica de referencias al sitio. Fuente: elaboración del autor.

Esta gráfica y tabla muestran al analista cuáles son los tipos de canales por los que los usuarios llegan a su sitio *web*. Es posible que el usuario deira click en un enlace

desde alguna red social o desde un buscador *web*. Esta información es útil para saber cuáles son los canales de adquisición más relevantes para el negocio.

Finalmente para anteder las historias de usuario H2 y H5, se generó un componente que registra cada visita realizando una consulta a un *web service* y con esto se registra la fecha y hora de entrada a la página que se solicita.

Para poder instalar dicho componente sólo es necesario agregar un tag al código html de cada página que compone al sitio.

La ejecución del sprint con respecto a la planeación puede analizarse en la gráfica de *burn up* agregando una serie que representa los puntos de complejidad de los artefactos construidos a lo largo de la etapa de desarrollo.

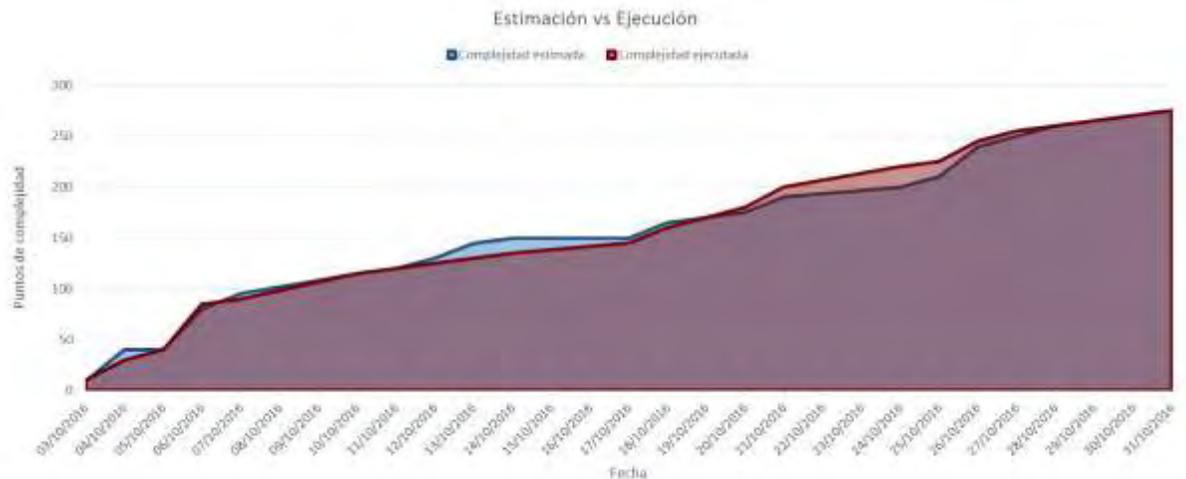


Ilustración 10 Burn up de estimación vs ejecución del sprint 1. Fuente: elaboración del autor.

Entregables del *sprint 2*

En este segundo sprint se realizaron algunas modificaciones al procesamiento de los datos de las visitas de los usuarios, de manera que se guardaran algunos datos adicionales para poder analizar cuál es el flujo de la navegación los usuarios en el sitio.

Para mostrar esto, se creó una gráfica de flujo dentro de las vistas de la solución.

Flujo de usuarios recopilado

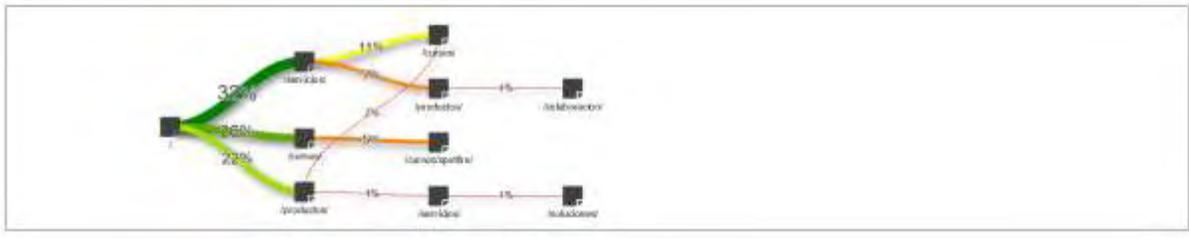


Ilustración 11 Gráfica de flujos de usuario. Fuente: elaboración del autor.

Cada cuadrado de la gráfica representa una página del sitio y las líneas que los unen representan la cantidad de usuarios que han seguido cada ruta, entre más gruesa sea la línea significa que una mayor cantidad de usuarios ha seguido esa ruta.

Posteriormente se agregaron otros módulos administrativos para poder dar de alta usuarios, organizar usuarios en grupos de trabajo, asociar los grupos de trabajo con sitios a los que están autorizados a analizar y finalmente la funcionalidad para iniciar y cerrar sesión en el sistema.

Las pantallas que se muestran a continuación fueron las diseñadas para cumplir con estos requerimientos.



Ilustración 12 Administración de usuarios. Fuente: elaboración del autor.

Desde esta vista es posible que el director de mercadotecnia registre analistas para que puedan consultar los analíticos de los sitios.

The image shows a web form titled "Agregar usuario" (Add user) with a close button in the top right corner. The form has two tabs: "Datos del usuario" (User data) and "Grupos del usuario" (User groups). The "Datos del usuario" tab is active. The form contains the following fields:

- Nombre***: Text input field with placeholder "Nombre".
- Apellido paterno***: Text input field with placeholder "Apellido paterno".
- Apellido materno**: Text input field with placeholder "Apellido materno".
- Correo electrónico***: Text input field with placeholder "Correo electrónico".
- Nombre de usuario***: Text input field with placeholder "Nombre de usuario".
- Contraseña***: Password input field with placeholder "lg!+_F2DjNw".
- Habilitado**: A checkbox that is checked.

At the bottom right of the form, there are two buttons: "Cerrar" (Close) and "Guardar" (Save).

Ilustración 13 Formulario para alta de usuarios. Fuente: elaboración del autor.

El formulario solicita información necesaria para la alta de usuarios. Y dando click en la pestaña “Grupos de usuario” el director de mercadotecnia puede definir a qué equipo de trabajo pertenecerá el analista creado, tal como se muestra en la siguiente ilustración.

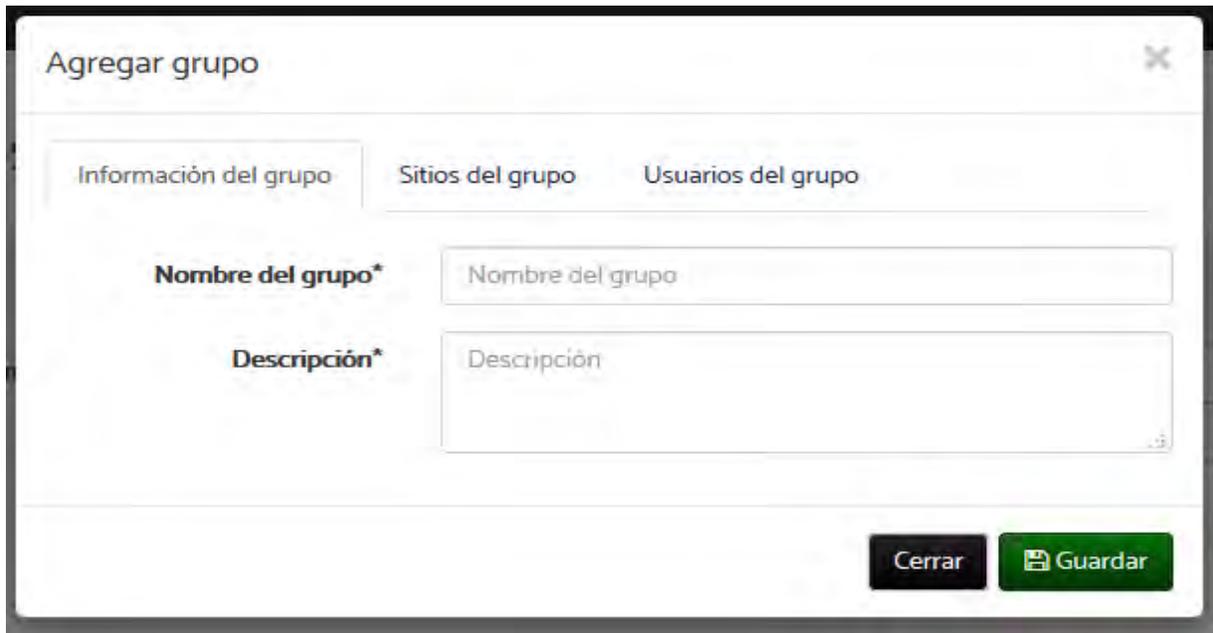


Ilustración 14 Agregar usuario a un grupo de trabajo. Fuente: elaboración del autor.



Ilustración 15 Administración de grupos. Fuente: elaboración del autor.

Este es el módulo que utiliza un director de mercadotecnia para definir cuáles son los grupos de trabajo, aquí se pueden observar y editar los grupos además de agregar grupos nuevos. Para agregar un grupo debe llenarse la información mostrada en el formulario siguiente:



Agregar grupo

Información del grupo Sitios del grupo Usuarios del grupo

Nombre del grupo*

Descripción*

Cerrar Guardar

Ilustración 16 Formulario de grupo de trabajo. Fuente: elaboración del autor.

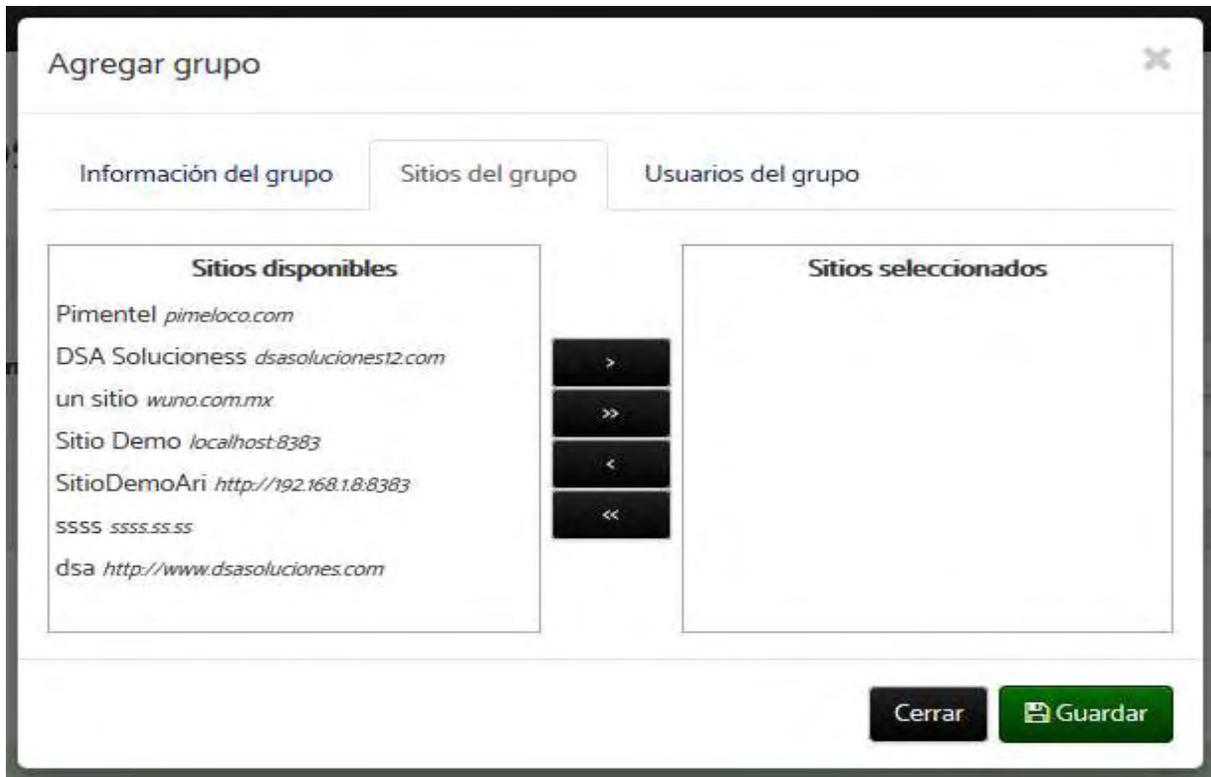


Ilustración 17 Agregar sitios al grupo de trabajo. Fuente: elaboración del autor.

Los sitios que se asignen al grupo de trabajo son los sitios a los que tendrán acceso los usuarios que pertenezcan al grupo. La solución permite que un usuario forme parte de varios grupos de trabajo.

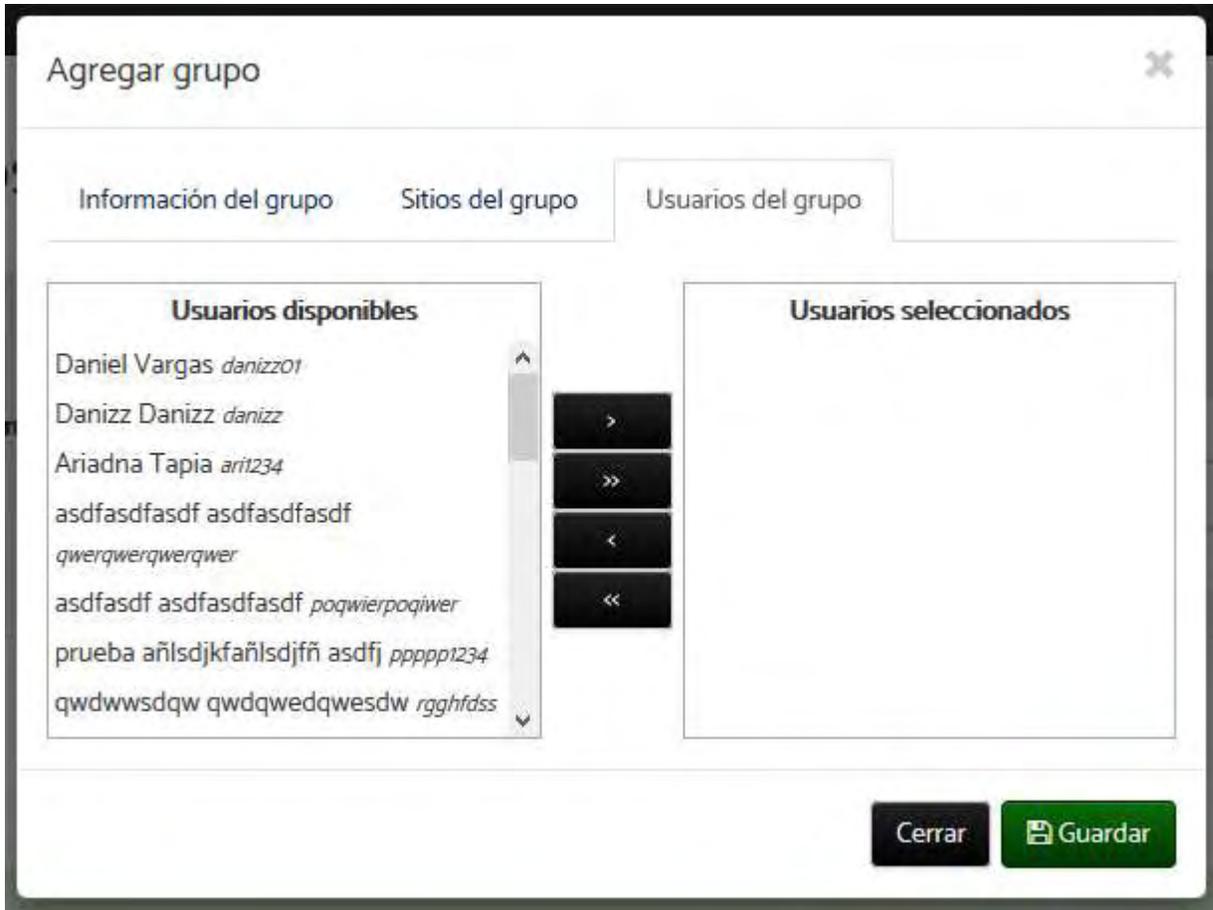


Ilustración 18 Agregar usuarios al grupo de trabajo. Fuente: elaboración del autor.

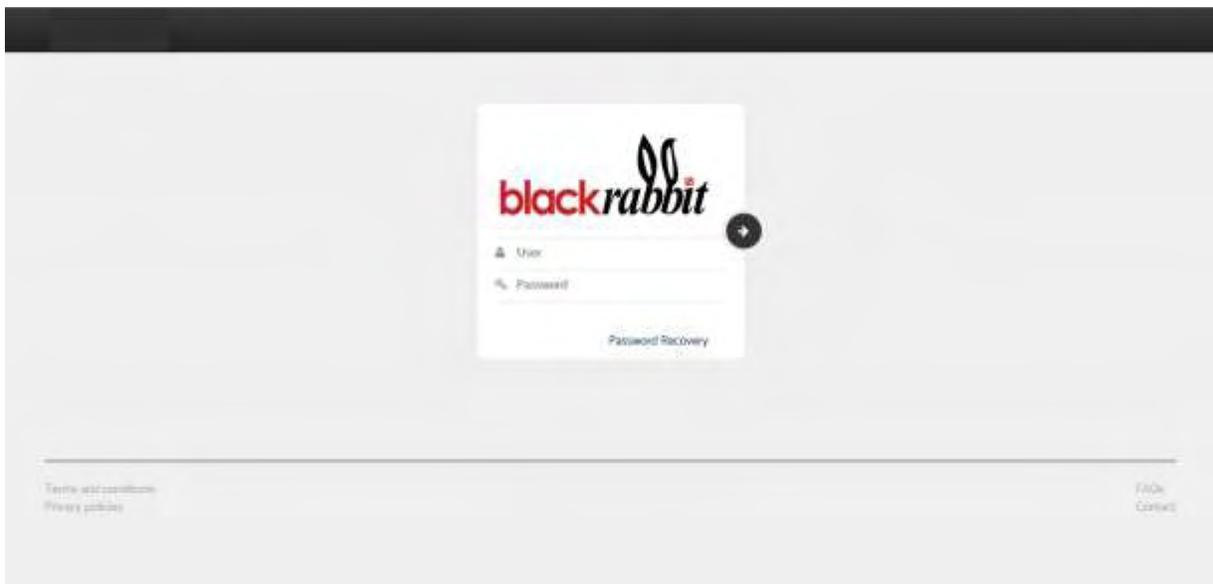


Ilustración 19 Pantalla de ingreso al sistema. Fuente: elaboración del autor.

Lo siguiente fue diseñar un algoritmo con el cuál se conociera cuál es el perfil de los usuarios que visitan el sitio *web* analizado.

Esta información se puede observar a manera de gráficas tal como se muestra en la siguiente imagen

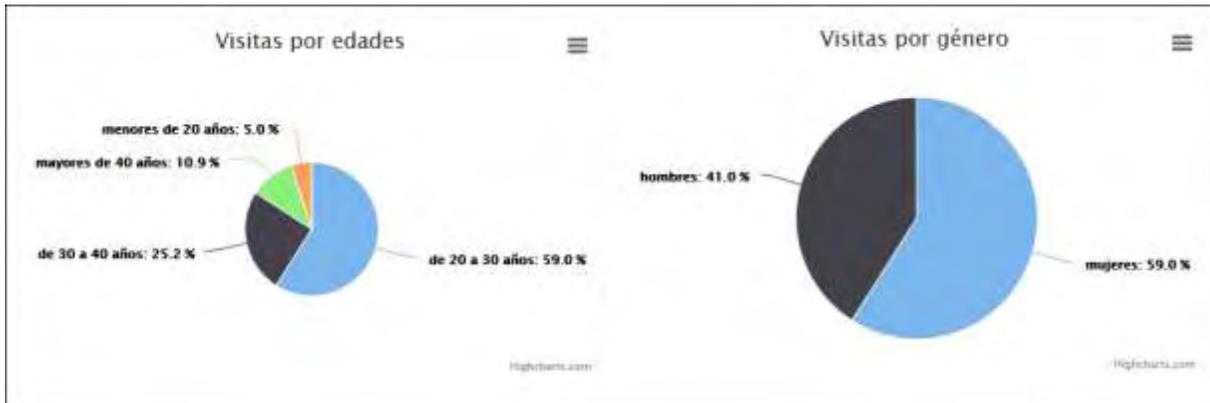


Ilustración 20 Segmentación de visitantes al sitio. Fuente: elaboración del autor.

Estas gráficas nos muestra el porcentaje de hombres y mujeres que utilizaron el sitio, además de dividirlos en grupos de edades.

Por último, el desfase en la ejecución puede observarse en la gráfica de tipo *burn up* al comparar la calendarización planeada con el avance obtenido en las fechas que abarca el *sprint*.

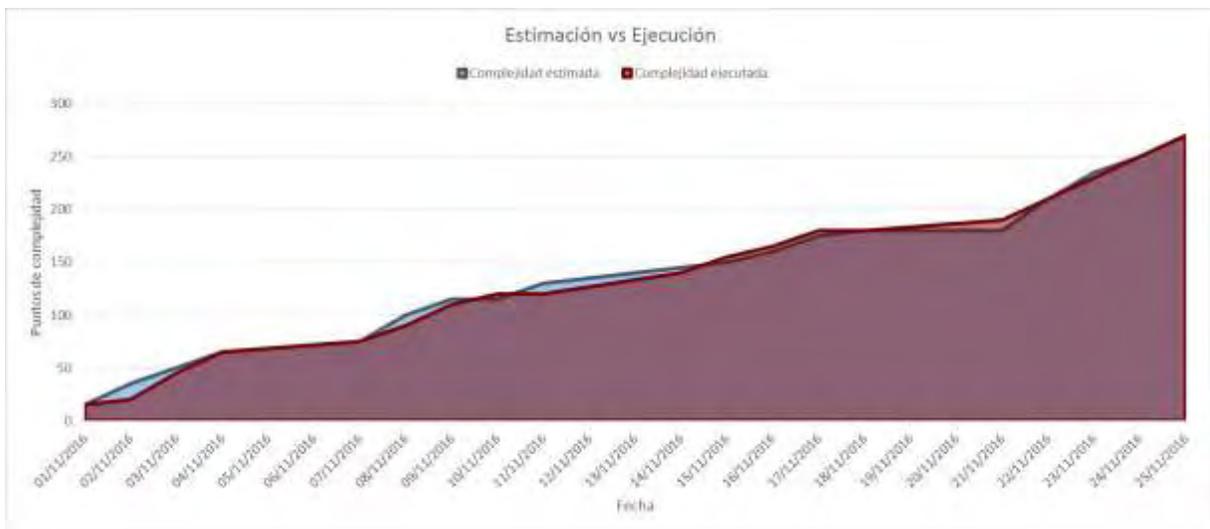


Ilustración 21 Burn up de estimación vs ejecución del sprint 2. Fuente: elaboración del autor.

CONCLUSIONES

Con base en lo estudiado a lo largo de la carrera de Ingeniería en Computación siguiendo el plan de estudios 2010 de la Facultad de Ingeniería de la UNAM, se propuso realizar el presente trabajo para resolver un caso de estudio identificado en el área de mercadotecnia digital, aplicando el concepto de grandes volúmenes de datos y metodologías ágiles.

Cada una de las materias cursadas fue de utilidad en distintos aspectos del desarrollo de esta tesis.

Durante la ejecución se aplicaron varios temas relacionados con ingeniería de *software* ya que el proyecto propuesto fue el desarrollo de una solución que pudiera analizar grandes cantidades de datos.

Por un lado se aplicaron los conocimientos administrativos adquiridos en materias como Ingeniería de *software*, Administración de proyectos y Verificación y validación del *software*.

En contraste con lo enseñado en la Facultad se optó por realizar el desarrollo bajo el esquema de la metodología ágil *Scrum*, la cual está orientada principalmente a que el valor de los proyectos se pueda apreciar desde los inicios de la planeación.

Lo aprendido en la Facultad aportó las bases de lo administrativo de los proyectos de *software*, sin embargo se puede observar con base en las gráficas de *burn up* que comparan la estimación con la ejecución, que aplicar una metodología ágil ayuda a que el proyecto se ejecute de manera correcta.

En ambos *sprints* hubo ligeros desfasamientos, pero en ambos casos se logró terminar las actividades planeadas en el tiempo establecido en la fase de planeación y calendarización del *sprint*.

Se considera que en el plan de estudios podría dedicarse más tiempo a instruir este tipo de metodologías que desde hace seis años son tendencia en el campo laboral y que han demostrado ser una buena elección para llevar los proyectos de alto valor, en contraste con metodologías consideradas ya obsoletas como los de *PMP –Project Manager Professional-*, sin embargo quizás la poca experiencia de los profesores en

el mundo laboral real no les ha impulsado el conocerlas para después impulsarlas con los alumnos.

Por otro lado de las materias computación para ingenieros, programación avanzada y métodos numéricos, algoritmos y estructuras de datos, lenguajes formales y autómatas, bases de datos, sistemas operativos y lenguajes de programación se aprendieron los conocimientos que permitieron el diseño e implementación de los algoritmos aplicados en la solución desarrollada, sin embargo hubo que explorar a mayor detalle los conocimientos dado que la profundidad requerida para ejecutar fue mayor en ciertos casos.

Si bien es cierto que una correcta administración de proyectos ayuda a cumplir con los entregables en tiempo y forma, es importante notar que los conocimientos técnicos son importantes pues de ellos puede realizarse el desarrollo de los elementos que componen una solución de *software* de manera rápida y asegurando su calidad.

Los temas enseñados en estas materias son muy importantes y han contribuido mucho a mi desarrollo profesional en áreas relacionadas con la ejecución de los proyectos.

Otro aspecto importante a considerar es la parte socio humanística del proyecto, la cual está sustentada en los conocimientos adquiridos de las asignaturas de ética profesional, redacción y exposición de temas de Ingeniería.

Ambas contribuyeron a 2 cosas principalmente.

Primero se logró mantener la solución dentro de un marco ético que siguiera los valores de la Universidad y del código moral profesional de un ingeniero en computación. Adicionalmente, la comunicación de características del proyecto (por ejemplo las historias de usuario) fue eficiente gracias a la correcta redacción de las mismas.

Las necesidades de negocio y el objetivo del proyecto se escribieron de tal forma que todos los involucrados (*product owner*, *scrum master* y *scrum team*) tuvieran presente cuáles eran y qué beneficio aportan al área de mercadotecnia digital.

Con respecto a las materias relacionadas con la infraestructura de red, se puede observar que han aportado un conocimiento que ayudó en la realización de este proyecto (y que resultan útiles en la mayoría de los proyectos actuales), aunque fue básico, es necesario se entregue mayor cantidad y calidad en esas materias, que lo permite el tiempo y la calidad de algunos profesores.

La solución resultado de esta tesis, como se observó se dedica a dar seguimiento a las consultas y peticiones de sitios *web*, los cuales están compuestos por recursos que se envían a través de internet (una red mundial de computadoras) por lo tanto es importante conocer cuáles son los protocolos de red y los elementos necesarios para que se pueda realizar el análisis mencionado.

Considerando las materias de competencias profesionales de negocios electrónicos e Introducción al análisis económico empresarial, se puede mencionar que los temas estudiados en ellas sirvieron para identificar las necesidades a observar o desarrollar por la solución desplegada.

Propongo que es importante conocer cuáles son los casos de estudio que existen y tener una visión amplia de dónde puede aplicarse una solución de *software*, esto con el fin de aportar valor a los negocios actuales y que en consecuencia dicha solución pueda comercializarse de manera exitosa.

Además, se adquirieron conocimientos básicos para poder plantear una estrategia que permita vender soluciones de *software*, por esta razón y por el valor que la solución desarrollada aporta al área de mercadotecnia digital, se concluye que es posible ofertarla para obtener ganancias a partir de su venta.

Durante la ejecución del proyecto se analizaron principalmente estas características sobre el plan de estudios de la carrera y adicionalmente se aprendieron conocimientos nuevos.

La práctica en el campo laboral me enseñó un aspecto muy importante y que podría mejorarse el enfoque actual que tiene la Facultad hacia él: el trabajo en equipo.

A lo largo de la carrera los profesores se preocupan por mencionar que aprendamos a trabajar en equipo y normalmente los proyectos que se proponen están pensados para ser realizados en equipos de trabajo.

Sin embargo, lo más común termina siendo que una únicamente una parte del equipo realiza todo el desarrollo y esto impide que se adquieran estas habilidades, las cuales son sumamente importantes en el campo laboral pues lograr trabajar en equipo permite enfrentar retos que superan las capacidades individuales de una persona.

Puede dedicarse un taller sobre cómo coordinar equipos de trabajo para aprender qué elementos se deben considerar al repartir las tareas y de qué forma se puede dar seguimiento al avance.

Algo que se aprendió también en el campo laboral, fue el uso de *frameworks* de desarrollo.

En materias como Lenguajes de programación o Ingeniería de *software* podría dedicarse un tiempo a la investigación de cuáles son las tecnologías que más utilizan las empresas en la actualidad, sería conveniente los profesores fueran del campo laboral con experiencia en este rubro.

Durante este desarrollo pudo observarse que la formación no depende únicamente de los profesores y del plan de estudios de la Facultad, es también responsabilidad del ingeniero en formación preocuparse por investigar cuáles son las tendencias de la industria.

Continuando con la investigación realizada, se menciona que la hipótesis se comprobó afirmativamente, lo cual se demuestra principalmente con las gráficas *burn up* de cada sprint, las cuales muestran que ninguno de ellos sufrió retraso aun cuando el *product owner* considerara algún cambio al planteamiento original para el desarrollo de la aplicación de *Big Data*.

Con base en las investigaciones realizadas sobre qué es la mercadotecnia y cuál es su objetivo, puede observarse también que la solución aporta información que puede ser aprovechada por ésta área.

Los indicadores que muestra la solución se pueden utilizar para idear y aplicar estrategias de mercadotecnia para la adquisición de una mayor audiencia a través de canales de comunicación en línea además de analizar si dicho canal está aportando el valor esperado.

Los datos que procesa la solución pueden ser de gran volumen y a una alta velocidad debido a que el sitio *web* es un recurso que está disponible para cualquier persona que tenga conexión a internet.

Es por esto que los datos adquieren la propiedad de viralidad, ya que la ésta se entiende como el aumento exponencial en volumen generado de datos por unidad de tiempo. Se puede dar el caso en que un sitio *web* tenga mucho éxito y por lo tanto sea compartido y visitado por un gran número de personas que a su vez lo comparten con más usuarios.

Tienen también un cierto grado de viscosidad ya que el uso de los sitios *web* es un fenómeno que tiene un comportamiento estacional.

La variedad de los datos obtenidos por la solución puede entenderse de la siguiente manera: cada persona que utiliza internet lo hace a través de distintos dispositivos y de diferentes navegadores *web*, por lo tanto la adquisición y el registro de las visitas tiene su origen en distintas fuentes.

Los datos que se registran son veraces en la medida que la ingesta de la información no falle y al ser un componente que está embebido dentro del sitio *web* del cliente se puede considerar que el rastreo no depende de un usuario por lo que los datos tienen alto grado de veracidad.

Las decisiones que se toman a partir de la información producida por la solución agregan valor al negocio pues es posible identificar áreas de oportunidad en las estrategias de mercadotecnia aplicadas además de reforzar los casos exitosos.

Con base en todas estas características se determina que el proyecto dio como resultado una solución de *Big Data*, pues cumple con las características expuestas en el capítulo de antecedentes, donde se describen algunos modelos de 3, 4 y hasta 8 Vs en los que se basan este tipo de soluciones.

Finalmente se reconoce la importancia del trabajo en equipo, el cual es uno de los pilares fundamentales de la metodología que fue aplicada en proyecto, pues cada miembro del equipo estuvo comprometido y realizó sus tareas esforzándose por cumplir con el tiempo estimado y así evitar retrasos.

Trabajar en equipo significa más que simplemente repartir tareas y que cada elemento ejecute lo suyo, también es necesario coordinar a los elementos, identificar en cuáles son las fortalezas y debilidades, además de consolidar un objetivo para que los integrantes puedan apoyarse y sepan cuál es la meta a alcanzar.

FUENTES

Adell, J. (1992). UN GOPHER EN LA UNIVERSITAT JAUME I. Universidad Jaume I, España

Bakshi, K., (2010). Considerations for big data: Architecture and approach. En Aerospace Conference, (pp.1 - 7). IEEE Conference Publications.

Berners-Lee, T., J. Hendler y Lassila O. (2001), "The semantic web", Scientific American, 17 de mayo disponible en: [www.sciam.com/article.cfm?articleID = 00048144-10D2-1C70-84A9809EC588EF21](http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21) y numPag = 1 & catID = 2 Boston College, disponible en: [http://metaquest.bc.edu/V,Boston, MA](http://metaquest.bc.edu/V,Boston,MA).

Bojicic, I., Marjanovic, Z., Turajlic, N., Petrovic, M., Vuckovic, M., y Jovanovic, V. (2016). A comparative analysis of data warehouse data models. *2016 6Th International Conference On Computers Communications & Control (ICCCC)* (pp. 151-159).

Burby, J., Brown, A. et. al (2007). Web Analytics Definitions. *Web Analytics Association*

Chawda, R. K. y Thakur, G. (2016). Big Data Advanced Analytics Tools. *2016 Symposium on Colossal Data Analysis and Networking (CDAN)*.

Daniel, E. D. (1999). **Data Storage on Hard Magnetic Disks**. *Magnetic Recording: The First 100 Years*. pp. 270 – 299.

Dean, J. y Ghemawat, S. (2004). Sixth Symposium on Operating System Design and Implementation. Google, San Francisco, CA.

Deutsch, P. (2000). **Archie—A Darwinian Development Process**. *IEEE Internet Computing*, 4(1) pp. 69-71.

Elmasri, R. y Navathe, S. B. (2011) Fundamentos de sistemas de bases de datos. (5 Ed.).

Facultad de Ingeniería. (2010). Plan de Estudios de la carrera de Ingeniería en Computación. Universidad Nacional Autónoma de México.

Fowler, M. y Highsmith, J. (2001). El Manifiesto Ágil. Agosto 2001. Consultado en Febrero de 2017 de la página <http://agilemanifesto.org/iso/es/manifesto.html>

Frana, P.L. (2004). ***Before the web there was Gopher***. IEEE Annals of the History of Computing (pp. 20-41).

Gaffin, A. (1994). Everybody's Guide to the Internet (1 Ed.) (Pp.129-140).

Greenberg and L. Garber, "Searching for new search technologies", ***IEEE Computer***, 32(8):4-11 , Aug. 1999.

Grier, D. A. (s. f.). *The Origins of Statistical Computing*. Obtenido en Octubre 26 de 2016, de la página (<https://ww2.amstat.org/about/statisticiansinhistory/index.cfm?fuseaction=paperinfo&PaperID=4>)

Grier, D. A. (2000). **Agricultural Computing and the Context for John Atanasoff**. *IEEE Annals Of The History Of Computing*, 22(1), 48.

Hornby, T. (2013) **The Rise of Google: Beating Yahoo at Its Own Game**. *Low End Mac - Tech History*. Recuperado el 8 de Enero de 2017 del sitio (<http://lowendmac.com/2013/the-rise-of-google-beating-yahoo-at-its-own-game/>)

Inmon, W. H. (2002). *Building the Data Warehouse systems*. McGrawHill Cap 12 Data Warehouse and Executive Information System Fundamentals.

Jiménez, C. (s. f.). *Bases de datos*. Recuperado en Noviembre 28 de 2016, de la página (<http://wikifoundryattachments.com/YIJNuvIOmar+SwgXvN7F+A==448853>)

Jing, H., Haihong, E., Guan, L., & Jian, D. (2011). Survey on NoSQL database. *2011 6Th International Conference On Pervasive Computing & Applications (ICPCA)* (pp. 363-366).

Kaushik, A. (2009). **Chapter 2 The Optimal Strategy for Choosing Your Web Analytics Soul Mate**. *Web Analytics 2.0: The Art of Online Accountability and Science of Customer Centricity*. pp 15 – 34.

Kaushik, A. (2009). **Chapter 3 The Awsome World of Clickstream Analysis: Metrics.** *Web Analytics 2.0: The Art of Online Accountability and Science of Customer Centricity.* pp 35 – 74.

Khare, R. (1999). **WHO KILLED GOPHER? An Extensible Murder Mystery** IEEE Internet Computing Volume 3. (pp. 81 - 84).

Kimball, R. (2002). *The Data Warehouse toolkit: the complete guide to dimensional modeling* 2nd ed, USA. publicado por John Wiley & Sons inc.

Kotler, P. y Armstrong, G. (2013). *Fundamentos de marketing*, México 2013 PEARSON

Lacourly, N. (2010). **Capítulo 1: El pensamiento estadístico.** *Introducción a la Estadística.* p. 21

Lee, G. (1999) A Semantic Approach Adds Meaning to the Web.

Li, Q., Choi Yuk J. (2000). **WebReader: a Mechanism for Automating the Search and Collecting Information from the World Wide Web.** Web Information Systems Engineering.

Lindner, P. (1992). Frequently Asked Questions (FAQ) about Gopher. Documento electrónico. FTP boombox.micro.umn.edu. E-mail: Lindner@boombox.micro.umn.edu. Distribuido a través de grupo de News: alt.gopher.

Lindner, P. (1993). **Internet Gopher User's Guide.** University of Minnesota, USA.

Navarro, J. A. y Barrientos R. C. (2013) **¿Dónde quedó el cómputo científico? Avances y retrocesos de las herramientas computacionales en las ciencias biológicas.** *Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, 10.

Rodríguez, L., y Cervantes, O. (2015). **Small Step for Machines, Giant Leap for Mexico: A Local History of Computing.** *IEEE Annals Of The History Of Computing*, (4), 14.

Sadeh, T. y Walker, J. (2003). **Library portals: toward the semantic Web**. New Library World, Vol. 104. (pp. 11-19).

Satpathy, T. et. al. (2016). *A Guide to the SCRUM BOOK OF KNOWLEDGE (SBOK GUIDE)*, Estados Unidos 2016.

Schmarzo, B. (2013). *Big Data: Understanding How Data Powers Big Business*. Indianapolis, Indiana John Wiley & Sons Inc.

Schmarzo, B. (2013b). **Capítulo 2 Clase de historia sobre los Big Data**. *Big Data: El poder de los datos*. pp. 43 – 50.

Seymour T., Frantsvog D. y Kumar S. (2011). **History Of Search Engines**. International Journal of Management & Information Systems Volume 15, Number 4. (Pp47-58).

Sommerville, I. (2011). *Ingeniería de Software*. (6 Ed.) Pearson. Madrid, España.

Sruthika, S. y Tajunisha, N. (2015). A study on evolution of data analytics to big data analytics and its research scope. *2015 International Conference On Innovations In Information, Embedded & Communication Systems (ICIIECS)*, pp. 1-6.

Tamayo y Tamayo, M. (2004). *El proceso de la investigación científica*, (4ta ed). México: Llmusa

Trejo Medina, D. (2013). *Apuntes sobre Big Data*. México, CDMX.

Trejo Medina, D. (2011). *Apuntes de Ingeniería de Software*. México, CDMX, México: DanTM.

Ying, B. y Bhalla, S. (2012). **Introduction to Databases**. *Practical Database Programming with Visual Basic .NET*. pp. 10 – 12.