

PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA INFORMÁTICA

**CUSTOMER EXPERIENCE EN AGENCIA VIRTUAL
DE VIAJE HOSTELWORLD**

**VICTORIA MARIA CARVAJAL VIDELA
MARTIN ALONSO MUÑOZ FELDSTEDT**

INFORME FINAL DE PROYECTO
PARA OPTAR AL TÍTULO PROFESIONAL
DE INGENIERO DE EJECUCIÓN EN INFORMÁTICA

DICIEMBRE 2016

**PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA INFORMÁTICA**

CUSTOMER EXPERIENCE EN AGENCIA VIRTUAL DE VIAJE HOSTELWORLD

**VICTORIA MARIA CARVAJAL VIDELA
MARTIN ALONSO MUÑOZ FELDSTEDT**

**Profesor Guía: Cristian Alexandru Rusu
Profesor Co-Referente: Silvana Roncagliolo de la Horra**

Carrera: Ingeniería de Ejecución en Informática

Diciembre 2016

Dedicatoria

Trabajo dedicado a nuestros familiares y amigos, quienes
brindaron su apoyo en todo momento de la investigación.

Índice

Resumen	iv
Abstract	iv
Lista de Figuras	v
Lista de Tablas	vi
1 Introducción.....	1
2 Definición del Proyecto	2
2.1 Descripción del Problema	2
2.2 Objetivos	2
2.2.1 Objetivo General.....	2
2.2.2 Objetivos Específicos	2
2.3 Metodología de la Investigación.....	3
2.4 Plan de Trabajo	3
3 Marco Referencial.....	5
3.1 Experiencia del Cliente	5
3.2 Ciencia de los Servicios	7
3.3 Experiencia del Usuario.....	8
3.4 Usabilidad	9
3.5 Agencias de Viaje	11
3.6 Minería de Datos	12
3.7 Minería Web	14
4 Caso de Estudio	16
5 Trabajo Realizado	17
5.1 Herramienta de Extracción de Datos	17
5.2 Análisis de Datos.....	20
5.2.1 Matriz de Correlación.....	21
5.2.2 Prueba de Kruskal Wallis	24
5.2.2.1 De Acuerdo a Diferentes Capitales de América del Sur.....	24
5.2.2.2 De Acuerdo a Tipo de Viajero en Hostales de Santiago.....	25
5.2.2.3 De Acuerdo al Grupo de Edad en Hostales de La Paz.....	26
5.2.2.4 De Acuerdo al Mes de Hospedaje en Hostales de Buenos Aires.....	26
5.3 Comparación con Tripadvisor.com y Atrapalo.com.....	27

6	Conclusiones	29
7	Referencias.....	30
	Anexos.....	
	A: Análisis Región de Valparaíso	
	B: Manual de Usuario y Documentación de la Herramienta.	

Resumen

En el siguiente informe se presentan diferentes conceptos con tal de comprender y generar conocimiento útil, esto realizando un análisis a las valoraciones proporcionadas por los usuarios de la agencia virtual de viajes hostelworld, en la que los usuarios evalúan diferentes categorías de los servicios que brindan los hostales que se presentan en este sitio web.

Para abordar la problemática se deben entender conceptos tales como la interacción hombre máquina, la experiencia del cliente, usabilidad, las ciencias de los servicios y la minería de datos. Una vez explicado esto, se dará inicio a la extracción de comentarios existente en los hostales del sitio web con un script basado en PHP apoyado en diferentes librerías, y se analizarán las valoraciones dejadas por los usuarios con un programa estadístico para poder identificar tendencias.

Palabras-clave: experiencia del cliente, ciencia de los servicios, experiencia del usuario, usabilidad, minería de datos.

Abstract

The following report presents different concepts in order to understand and generate useful knowledge, this is an analysis of the valuations provided by users of the virtual travel agency hostelworld, in which users evaluate different categories of services that bridle the Hostels that are presented on this website.

In order to approach the problem we mean concepts such as human machine interaction, customer experience, usability, service sciences and data mining. Once this was explained, the extraction of comments was started in the hostels of the website with a script based on PHP supported in different libraries, and analyze the valuations left by the users with a statistical program to be able to identify trends.

Keywords: customer experience, service science, user experience, usability, data mining.

Lista de Figuras

Figura 3.1 Momentos de la Verdad [4].....	6
Figura 3.2 Factores de la experiencia del usuario.	9
Figura 3.3 Dimensiones de usabilidad.....	10
Figura 3.4 Modelo descriptivo y predictivo y sus respectivos modelos.....	13
Figura 3.5 Fases de la minería de uso web	15
Figura 5.1 Comentario de un usuario de HostelWorld.....	17
Figura 5.3 Procesamiento de Etiquetas.	19
Figura 5.4 Interfaz Herramienta Web.....	20
Figura 5.5 Archivo Excel Extraído.....	20

Lista de Tablas

Tabla 2.1 Actividades del plan de trabajo.	4
Tabla 5.2 Estructura de los datos extraídos.	18
Tabla 5.2 Matriz de correlación de Spearman Santiago	23
Tabla 5.3 Matriz de correlación de Spearman Lima	23
Tabla 5.4 Prueba de Kruskal Wallis sobre hostales de capitales de América del Sur	25
Tabla 5.5 Prueba de Kruskal Wallis de acuerdo al Tipo de Viajero	25
Tabla 5.6 Prueba de Kruskal Wallis de acuerdo al Grupo de Edad de los Viajeros.....	26
Tabla 5.7 Prueba de Kruskal Wallis de acuerdo al mes de hospedaje.....	27

1 Introducción

Hoy en día internet constituye una parte esencial para las diferentes empresas existentes en todo el mundo, como también para los usuarios, ya que prácticamente la mayoría de las personas acceden a esta tecnología con el fin de simplificar su vida o entretenerse. Las agencias de viaje son muy utilizadas últimamente ya que la mayor parte de ellas opera de forma online entregando un servicio que sirve para mostrar opiniones de los usuarios que han utilizado el servicio anteriormente, y de esta forma que esto sirve para los futuros usuarios de estos servicios.

El estudio de interacción con el cliente lleva años en desarrollo por parte de diferentes empresas, queriendo ganar la confianza en base a la emoción que experimentará él frente a los diferentes productos que se le ofrecerán. La experiencia del cliente es poder llegar a la sensibilidad y emoción más profunda de este, queriendo alcanzar el subconsciente racional frente a una marca durante un ciclo de vida determinado.

Actualmente, el diseño está más orientado a la comodidad y experiencia del usuario, que es una necesidad que estudian las empresas para captar mayor cantidad de usuarios. En este punto se suma además que la fiabilidad, eficiencia y eficacia son un punto clave para la satisfacción del usuario, en este campo es propio de las necesidades que se presentan por parte de las entidades hacia el cliente.

En el siguiente informe se presenta el marco referencial de la investigación de forma que permita comprender los conceptos involucrados en la misma. En la primera parte de este informe se dará a conocer la descripción del problema que se abordará, los diferentes objetivos que se pretenden alcanzar y la metodología que se llevará a cabo.

Se han investigado y dado a conocer los diferentes conceptos tales como la usabilidad, la experiencia del usuario, la ciencia de los servicios y la experiencia del cliente, así como también se abordan temas como las agencias de viaje y la minería de datos. Posteriormente se dará a conocer el caso de estudio con el que se trabajará a lo largo de todo el proyecto, el análisis que se entregará en base a los datos obtenidos y finalmente a partir de la información obtenida, proporcionar las conclusiones en base al análisis anterior tratando de alcanzar una solución a la problemática planteada.

2 Definición del Proyecto

2.1 Descripción del Problema

Las agencias virtuales de viaje son muy utilizadas hoy en día por millones de viajeros en el mundo, ya sea en busca de hoteles o vuelos. Han tenido un importante crecimiento a través de los últimos años e internet constituye un instrumento de vital importancia para la actividad de estas empresas [1]. Además no cabe duda que muchos de los usuarios de estas plataformas se basan en las opiniones disponibles en las distintas agencias para poder elegir donde será su estadía, por lo tanto las percepciones dejadas por los usuarios son muy importantes para la misma empresa como también para los futuros clientes.

Los servicios entregados por una empresa son de suma importancia para el futuro y el éxito de estas, por lo tanto evaluar la experiencia del cliente supone una mejoría al servicio entregado. Existen diversos métodos para evaluar la experiencia del consumidor, por ejemplo, ciertos métodos para evaluar la experiencia del usuario también sirven para evaluar ciertos aspectos de la experiencia del cliente [2].

En este caso se utilizará un enfoque cuantitativo de los datos disponibles de manera libre en el sitio web hostelworld.com a pesar de que se deberán extraer las opiniones y valoraciones de los usuarios del sitio web hostelworld.com para esto se ha desarrollado una herramienta de apoyo a la extracción, de manera de facilitar las tareas. Posteriormente se refinarán estas valoraciones para ser analizados con un programa estadístico y así poder identificar las diferentes tendencias.

2.2 Objetivos

A continuación se detallarán el objetivo general y los objetivos específicos del proyecto.

2.2.1 Objetivo General

Analizar las percepciones de los consumidores disponibles en el sitio web hostelworld.com.

2.2.2 Objetivos Específicos

- Comprender los conceptos de Customer Experience y sus métodos de evaluación.
- Identificar tendencias de las percepciones de los usuarios.
- Comprobar si las tendencias identificadas se aplican en otros contextos.
- Crear una herramienta de extracción de valoraciones del caso de estudio.

2.3 Metodología de la Investigación

Para la realización de este proyecto se utilizará como metodología de investigación el enfoque cuantitativo, en el cual se utiliza la recolección de datos para probar hipótesis, con base en la medición numérica y el análisis estadístico, con el fin de establecer patrones de comportamiento y probar teorías [3].

En este caso se utilizará un alcance descriptivo ya que en este se miden, evalúan o recolectan datos sobre diversos conceptos (variables), aspectos o dimensiones del fenómeno a investigar, y se busca identificar y comprobar tendencias dadas las opiniones de los usuarios de la agencia virtual de viajes.

A continuación se detallan ciertos hitos del proyecto de acuerdo al enfoque planteado anteriormente:

1. Análisis y estudio de material bibliográfico: Realizar una búsqueda en la literatura sobre Customer Experience y sus métodos de evaluación, como también sobre UX y Usabilidad.
2. Descripción del problema: Identificar el problema que se quiere resolver para posteriormente llevar a cabo una solución y entregar respuestas.
3. Definición de objetivos: Plantear el objetivo general y específico del proyecto.
4. Seleccionar caso de estudio: Utilizar un sitio web donde sea posible extraer los datos con sus diferentes dimensiones, de forma que estos sean de libre acceso para poder recolectarlos de forma automática.
5. Análisis estadístico de los datos: Se utilizará un software de estadística como SPSS u otro para obtener patrones e identificar tendencias.
6. Comprobación de tendencias: Se pretende evaluar si estas tendencias encontradas son aplicables a otros entornos como por ejemplo otros países.
7. Desarrollo de herramienta: Se deberá crear una herramienta de trabajo que pueda mostrar tendencias encontradas para su posterior interpretación. Esta se creará en la segunda parte del proyecto.

2.4 Plan de Trabajo

El plan de trabajo consiste en las asignaturas de proyecto 1 y proyecto 2, el primero realizado durante los meses de marzo a junio y el segundo desde agosto a diciembre. Para realizar este trabajo de la mejor forma posible se realizarán juntas de trabajo entre los integrantes 2 o más veces por semana, además de reuniones con el profesor guía de acuerdo a su disponibilidad, con el fin de aclarar dudas y mostrar avances. A continuación en la tabla 2.1 se muestra un detalle de las actividades realizadas durante cada mes.

Hasta el momento se encuentran realizadas todas las actividades propuestas para proyecto 1, las cuales se enfocaron en análisis de hostales de Valparaíso y Viña del Mar. Para proyecto 2, se ha creado una herramienta que apoya la extracción de datos, por lo tanto se ha expandido el análisis a diversas ciudades del país y a algunas capitales del Latinoamérica.

Tabla 2.1 Actividades del plan de trabajo.

Actividad /Mes	Mar	Abril	Mayo	Junio	Agosto	Sept	Oct	Nov
Proyecto 1	X	X	X	X				
Análisis y estudio del material bibliográfico	X	X	X					
Descripción del problema	X	X						
Definir objetivo general		X						
Definir objetivos específicos		X						
Elaboración marco referencial	X	X						
Selección caso de estudio		X	X					
Recolección de datos		X	X					
Refinación de los datos para análisis estadístico			X					
Análisis de datos			X	X				
Conclusiones				X				
Proyecto 2					X	X	X	X
Nueva recolección de datos					X			
Nuevo análisis de datos					X			
Formalizar propuesta						X		
Desarrollo de herramienta						X	X	
Refinación de propuesta							X	
Conclusiones								X

3 Marco Referencial

A continuación se describen los conceptos necesarios para comprender este proyecto.

3.1 Experiencia del Cliente

Son muchos los conceptos asociados a la disciplina de la experiencia del cliente, es fundamental entender como las percepciones desencadenan emociones y sentimientos en las organizaciones y cómo éstos tienen efectos directos en los resultados obtenidos [4]. Ya que es un campo relativamente nuevo, aún no existe una definición objetiva del concepto.

A través de diversos artículos se han considerado una serie de definiciones. Gentile, Spiller y Noci han definido la experiencia del cliente como una evolución entre una compañía y el cliente [5]. Laming y Mason la definen como las experiencias físicas y emocionales ocurridas a través de las interacciones con el producto y/o servicio ofrecido por una marca desde el primer punto de contacto hasta la etapa de post-consumo [6].

Lo claro está en que es un campo en el que las organizaciones hoy en día están interesadas, ya que en definitiva un cliente con una buena experiencia con un servicio o producto específico, estará fidelizado con este y además se encargará de recomendarlos, lo que se traduce como buenos beneficios para una empresa.

La experiencia del cliente se desarrolla a través de los puntos de contacto que vinculan a la organización con el cliente, tanto los que están bajo el control de la marca, como los que no. El cambio en los puntos de contacto más influyentes sobre el proceso de decisión de compra y el uso estratégico de las nuevas tecnologías, juegan un papel crítico en diferenciar su marca de la de sus competidores [7].

Una correcta experiencia del cliente debe basarse en tres pilares clave:

- Diseño de la experiencia adecuada para los clientes adecuados.
- Generación de valor añadido diferenciado por tipo de cliente en cada punto de contacto.
- Eficiencia: desarrollo de las capacidades organizativas necesarias para superar las expectativas de los clientes continuamente.

Cuando se habla del modelo de medición de la experiencia del usuario, hay que tener en cuenta los indicadores básicos, pero se necesita realmente gestionar y aprovechar la información y generar experiencias memorables. Se debe contar con modelos avanzados que indiquen un factor diferente a la satisfacción y que se alinean con los resultados de negocio.

La experiencia del cliente es un concepto abstracto que para ser medido debe ser desglosado en elementos más concretos y tangibles, también llamados puntos de contacto. Uno de estos elementos son lo que se denominan los “Momentos de la Verdad” -Moments of Truth, MOT’s en inglés [4].

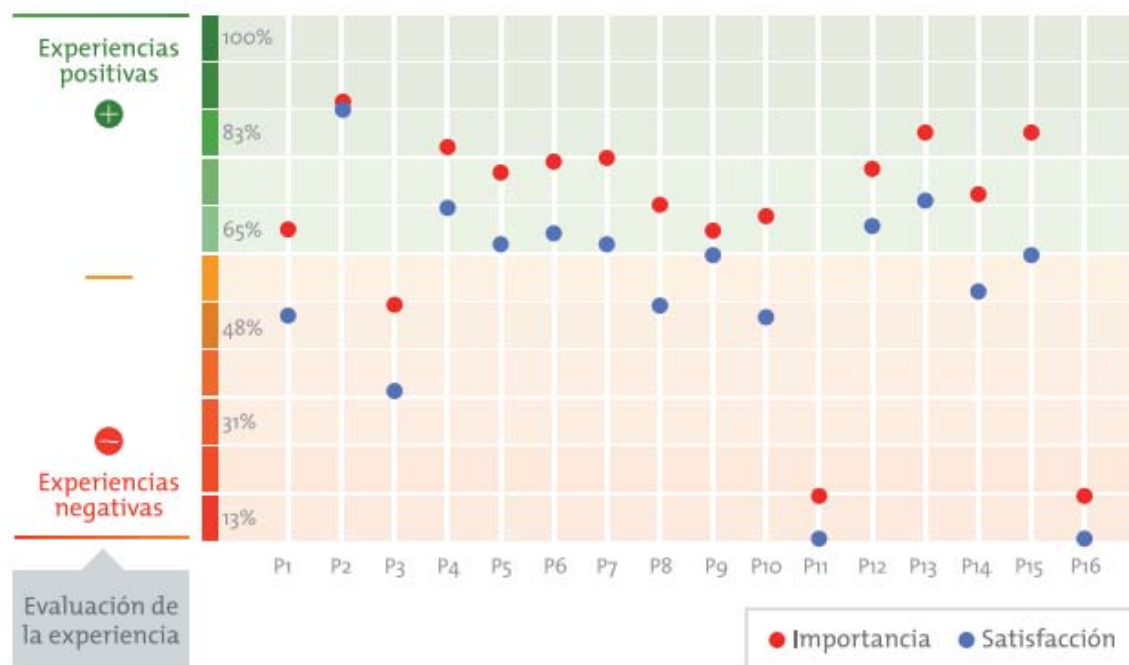


Figura 3.1 Momentos de la Verdad [4].

El mapa de la experiencia que se muestra en la figura 3.1, se basa en analizar la experiencia del cliente a lo largo de todo el ciclo de vida que tiene este con un servicio, comparando la expectativa del cliente con la experiencia percibida. Como se observa en la imagen, no todos los puntos de contacto tienen el mismo porcentaje de importancia, por ejemplo el punto 11, tiene un bajo porcentaje de satisfacción y al mismo tiempo bajo porcentaje de importancia.

Para analizar el ciclo de vida y los principales puntos de contacto. Se deberá realizar una encuesta a los clientes con los siguientes puntos:

- Importancia: ¿Cuál es la expectativa que tienen con la organización?
- Experiencia: ¿Cuál ha sido su experiencia con la organización?
- Generar los indicadores para cada punto de contacto. Es recomendable utilizar una escala numérica y concentrarse en los resultados de “Los momentos de la verdad”.
- Construir el mapa de la experiencia de forma gráfica comparando los resultados de importancia versus satisfacción, como se muestra en la figura 3.1.

Ya que no todos los puntos de contacto tienen la misma importancia para el cliente, es necesario trabajar en aquellos donde realmente es posible impactar la percepción de este y poder crear una buena experiencia. Es importante saber que la experiencia no solo está compuesta por los atributos físicos como por ejemplo, limpieza, entorno, calidad, etc., sino que también está conformada por la manera en que percibe y procesa la experiencia un cliente en específico, ya que no todos tienen las mismas características.

Para evaluar estos atributos es necesario apoyarse en dos herramientas que funcionan en forma conjunta, el análisis de correlación y las matrices de impacto. En el análisis de correlación, los modelos de regresión permiten comparar dos series de datos, donde el resultado obtenido muestra en qué medida se encuentran relacionados. Por otro lado las matrices de impacto son la representación gráfica de los indicadores en función de su valoración y de su índice de correlación, permiten visualizar que atributos son realmente importantes para su posterior mejora y por lo tanto para que el cliente tenga una mejor experiencia.

3.2 Ciencia de los Servicios

La Ciencia de los Servicios es una disciplina, actividad profesional y empresarial de triple enfoque ya que se trata de una ciencia porque crea conocimiento, gestión porque se caracteriza por poseer un valor empresarial e ingeniería por su contribución al desarrollo técnico y social [8].

Si bien una aceptación más amplia y completa sería la de “Ciencia, Gestión e Ingeniería de los Servicios” es la aplicación de disciplinas científicas, de gestión y de ingeniería a los trabajos que una organización (proveedora de los servicios) realiza para y con otra organización (la receptora de los mismos) [9]. En la definición anterior se entiende por “servicio” a aquello que tiene un valor económico pero no posee una consistencia material.

Los primeros en plantear este nuevo concepto fueron IBM, dándole importancia a esta nueva disciplina como una base para la innovación de servicios en el crecimiento económico [10]. Según la Universidad de Cambridge (2007), la evolución que ha tenido la investigación de esta área se puede dividir en seis periodos. Se detalla cada uno de estos a continuación.

- Antes de 1980: Comienzan a distinguirse los conceptos de marketing de producto y marketing de servicios, y también las operaciones con enfoque a producto o servicio.
- Entre 1980 y 1985: Existen algunas publicaciones sobre el área, y comienzan los primeros grupos de académicos y profesionales a desarrollar el tema.
- Entre 1985 y 1982: Se incrementa en forma significativa la literatura, incluyendo revistas de investigación, conferencias y libros.
- Entre 1993-2000: Se realizan investigaciones cuantitativas y se amplía el área de investigación.
- Desde el 2000 a hoy: Creación de un lenguaje propio con gran número de modelos de servicio. Aparece la iniciativa de IBM que implica al mundo académico y a la industria de servicios.
- Futuro: Crear comunidades que realicen aproximaciones desde distintas disciplinas y áreas de conocimiento sobre el tema de servicios, con enfoque en la Innovación de los Servicios, la Ingeniería, la Gestión y el Diseño.

3.3 Experiencia del Usuario

La experiencia del usuario (UX) es un término que se utiliza para describir el nivel de satisfacción de los usuarios cuando utilizan algún producto. Si es una buena experiencia, ellos estarán felices, si es una mala experiencia, los clientes no regresaran. Cuando se habla de la experiencia del usuario, representa la percepción dejada en la mente de un usuario después de una serie de interacciones entre la gente, dispositivos y eventos, y por lo tanto se encuentra relacionada con la experiencia del Cliente.

Algunos autores, como Knapp Bjerén (2003) definen la experiencia del usuario como; un conjunto de ideas, sensaciones y valoraciones del usuario resultado de la interacción con un producto; es resultado de los objetivos del usuario, las variables culturales y el diseño del interfaz” [11], especificando no sólo de qué fenómeno es resultante, sino también qué elementos la componen y qué factores intervienen en la interacción. En el contexto de la web, DNX (2005) definen la buena experiencia del usuario como un objetivo – “lo que se persigue es generar sensaciones y valoraciones de los usuarios hacia nuestro sitio web lo más agradables, positivas y satisfactorias posibles”, además de reseñar la “fidelidad del usuario” como consecuencia de alcanzar este objetivo [12].

La experiencia del usuario representa un cambio emergente del propio concepto de usabilidad, donde el objetivo no se centra solamente en mejorar el rendimiento de los usuarios en la interacción (eficiencia, eficacia y facilidad de aprendizaje), si no que busca resolver el problema estratégico de la utilidad del producto y el problema psicológico del placer y diversión de su uso.

Existen ciertos factores que componen la experiencia del usuario, el análisis de algunos de estos modelos que se descomponen de la experiencia del usuario, en diferentes variables que se condicionan y modelan, podemos nombrar el trabajo de Arhippainen y Tähti (2003). Los autores clasifican los diferentes factores en cinco grupos diferenciados: factores propios del usuario, factores sociales, culturales, del contexto de uso y propios del producto [13]. En la figura 3.2 se muestra lo explicado anteriormente.

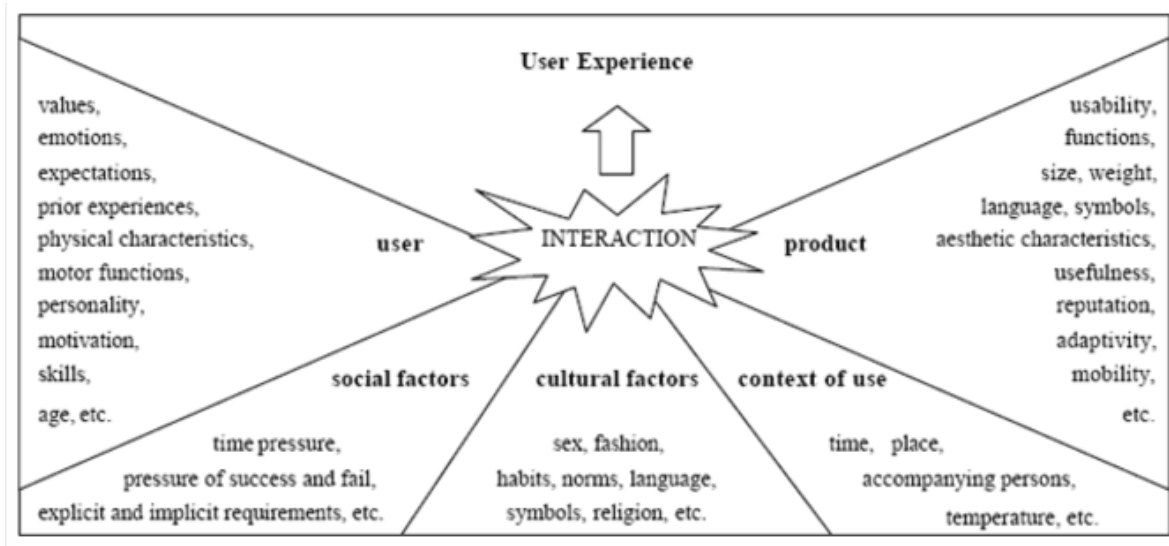


Figura 3.2 Factores de la experiencia del usuario.

La evaluación es una parte básica de un diseño, permite conocer el grado de cumplimiento de expectativas de los usuarios de un determinado sistema interactivo, así como si éste se adapta a su contexto social, físico y organizativo.

Cuando se habla de problemáticas de interacción con el usuario a través de la arquitectura web es necesario entender bien el concepto al momento de realizar diferentes análisis. Para la experiencia del usuario siempre será influida por expectativas y experiencias previas, y por tanto acondicionará expectativas y experiencias futuras, sumando que es un área de estudio multidisciplinario.

3.4 Usabilidad

La usabilidad web surgió a partir del nacimiento y desarrollo de internet como red de comunicación. Formalmente a partir del trabajo de Jakob Nielsen, considerado el padre de la usabilidad; es por esto que surgió en el ámbito de estudio Interacción persona-ordenador como una disciplina que busca que los usuarios se sientan cómodos al usar un software determinado.

Se define que la organización responsable de la estandarización ISO (International Standardisation Organization) propone dos definiciones del término usabilidad:

- El estándar ISO 9241-11 que forma parte de la serie ISO 9241, define la usabilidad como “la medida en la que un producto puede usar por determinados usuarios para conseguir objetivos específicos con efectividad, eficiencia y satisfacción en un contexto de uso especificado” [14].
- Algunos autores como Beltré Ferreras, en sus trabajos resalta que la definición del estándar ISO 9241-11 figura 3.1, contiene en su norma una visión sobre la aceptabilidad de un producto, la cual se fundamenta en [15]:

- o Eficacia (effectiveness): Representa la exactitud con la cual los usuarios alcanzan sus metas especificadas.
- o Eficiencia (efficiency): Los recursos gastados con relación a la certeza con la cual los usuarios logran las metas.
- o Satisfacción: la comodidad y la aceptabilidad del uso.



Figura 3.3 Dimensiones de usabilidad.

La usabilidad es la capacidad y facilidad para hacer accesibles la utilización de productos o servicios con la que una persona interactuara, por ejemplo, en la figura 3.3 se aprecia los beneficios de un trabajo que contiene las 3 actitudes. Por ejemplo, entre más compleja sea la aplicación y entre más trabajo requerirá al usuario poder utilizarla menor será el número de personas que podrán interactuar con ella, se reafirma lo anterior, si tiene la capacidad de atraer al usuario, es porque tiene calidad, y se puede afirmar que hay una técnica de usabilidad correctamente aplicada. Se debe considerar ciertos elementos extras que permitan a sus usuarios utilizar con cierto grado eficiencia, efectividad y satisfacción.

Es necesario entender que la usabilidad es una disciplina que está bastante asociada con la mejora de un diseño para, en consecuencia, crear una experiencia de usuario agradable o satisfactoria. Por esto que, aunque sea una forma de clarificar al usuario ciertas facilidades no está exenta de no dar una experiencia útil.

Para un sitio los objetivos que tienen que ser priorizados son los resultados deseados y al mismo tiempo que satisfaga al usuario, el grado de éxito de un sitio web depende del equilibrio que hay entre los objetivos de ciertas empresas y el de los usuarios; por esto no puede ser comprendida como una cualidad universal, todo producto, aplicación o sitio web, nace para satisfacer las necesidades de una audiencia específica. Por tanto, son esos productos lo que serán usados para una audiencia objetiva, no estará pensado para toda la población.

En el modelo de calidad de software la usabilidad se mantiene como una característica más del producto, pero que se llama operatividad, y tiene alguna diferencia respecto a la ISO 9126 [16], siendo sus sub-características:

- **Fácil de aprender:** capacidad para ser aprendido en su manejo
- **Fácil de entender:** capacidad del producto de ser entendida por nuevos usuarios en términos de su propósito y cómo puede ser usado en tareas específicas.
- **Fácil de usar:** capacidad del producto de ser operado y controlado en cualquier momento.
- **Atractivo:** capacidad del producto de ser atractivo para sus usuarios.
- **Conformidad:** capacidad del producto para adherirse a estándares, convenciones, guías de estilo o regulaciones relacionadas con la usabilidad.
- **Facilidad de ayuda:** hace referencia al grado con el que el producto ofrece ayuda a los usuarios cuando la necesitan, algo que en la ISO 9126 se consideraba implícita en la facilidad de aprender.
- **Accesibilidad técnica:** hace referencia a cómo el producto software puede ser usado por personas con discapacidad.

3.5 Agencias de Viaje

La gestión adecuada de la promoción de un destino turístico, cada vez es más indispensable para reducir riesgos en la toma de decisiones ya que se hacen más complejas e inciertas. Por eso el proceso de planificación estratégica y el diseño de estrategias de marketing exigen analizar variable de diversa índole, situaciones y comportamientos, tanto de entorno general como específico. Por lo tanto, se necesitan desarrollar metodologías de investigación de mercados que faciliten información objetiva y relevante de manera ágil y eficiente a los decisores e instituciones del sector turístico.

Los diferentes destinos turísticos han encontrado en internet la pareja ideal para romper fronteras con reducidos costes de promoción y la propia distribución. Para esto se dice que son múltiples los intentos para dejar un solo método de evaluación de sitios web. La opinión del usuario es un factor de suma importancia para la retroalimentación en el proceso de toma de decisiones.

Según algunos investigadores (Young A., Park y Ulrike Gretzel, 2008), los factores comunes de éxito más usuales analizados tanto en las webs de turismo son: (1) la Calidad de Información; (2) la Facilidad de Uso; (3) Seguridad /privacidad; (4) La Apariencia Visual; (5) Personalización; (6) Sensibilidad; (7) Interactividad; (8) Confianza; Y, (9) Cumplimiento [17]. Es una visión muy amplia de las dimensiones actualmente pertinentes, esto se debe tomar como un punto de inicio para futuras investigaciones, y eso se basa netamente en las características evolutivas de este campo.

Las evaluaciones de los sitios web son de uso reducido, si estos resultados no son interpretados en el contexto de metas específicas. Para esto también se habla de “data mining”, que su objetivo principal es trabajar con grandes volúmenes de datos aplicando

diferentes técnicas adecuadas para analizar y extraer el mayor conocimiento, para que la información descubierta sea más comprensible para las diferentes empresas.

3.6 Minería de Datos

A lo largo de la vida de una empresa se van acumulando grandes cantidades de datos que son almacenados, otros son usados y unos cuantos se perderán por falta de actualización o por cambios de políticas de manejo de datos. Ahora con el desarrollo de los sistemas de cómputo, las empresas tienen la capacidad de almacenar y acceder, en archivos o bases de datos, grandes cantidades de datos históricos sobre las operaciones diarias de sus negocios; información que en ese momento fue usada para satisfacer necesidades propias de la empresa o como soporte de decisiones.

El concepto es bien amplio, pero se destaca por ser un proceso que tiene como propósito descubrir, extraer y almacenar información relevante de amplias bases de datos, a través de programas de búsqueda e identificación de patrones y relaciones globales, tendencias, desviaciones y otros indicadores aparentemente caóticos que tienen una explicación que pueden descubrirse mediante diversas técnicas de esta herramienta.

Se define la minería de datos como el proceso de extraer conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos. Es decir, la tarea fundamental de la minería de datos es encontrar modelos inteligibles, es decir, comprensibles a partir de los datos [18].

El objetivo fundamental es aprovechar el valor de la información localizada y usar los patrones preestablecidos para que los directivos tengan un mejor conocimiento de sus negocios y puedan tomar decisiones más confiables.

La minería de datos auxilia a los usuarios empresariales en el procesamiento de reservas de datos para descubrir relaciones de las que, en algunos casos, anteriormente ni siquiera se sospechaba. La información obtenida a través de la minería de datos ayuda a los usuarios a elegir cursos de acción y a definir estrategias competitivas, porque conocen información que solo ellos pueden emplear.

Cuando se habla de modelos de minería de datos se les dice a la forma en que representamos los resultados que obtenemos, los mismos pueden ser predictivos y descriptivos.

Algunos autores definen estos dos modelos como: “Los modelos predictivos permiten estimar valores desconocidos o predecir los valores futuros sobre las variables de interés. Es decir tomando campos o variables de la base de datos estimando las variables a determinar su objetivo”. Algunos de los modelos predictivos son, clasificación y regresión. Un ejemplo de modelo predictivo es estimar la demanda de un nuevo producto en función del gasto en publicidad.

Los modelos descriptivos identifican patrones entre los datos existentes, nos sirven para predecir y mostrar un resumen de los datos. Algunos de los modelos descriptivos son, agrupamiento, reglas de asociación y el análisis correlacional.

La técnica que se ha utilizado para este proyecto y determinar ciertos tipos de patrones se llama “reglas de asociación” que de por sí es bastante utilizada en el rubro, esta consiste en determinar patrones en los conjuntos de datos en los que ocurren transacciones de datos.

Con esta técnica, pueden encontrarse relaciones sin que exista intervención alguna por parte de algún operador. El descubrimiento de estas reglas ayuda a las organizaciones dedicadas al comercio electrónico a definir estrategias de mercado efectivas. El aprendizaje de reglas de asociación se divide normalmente en dos fases:

1. Extracción de los conjuntos de ítems que cumplen con la cobertura requerida a partir de los datos.
2. Generación de las reglas a partir de estos documentos.

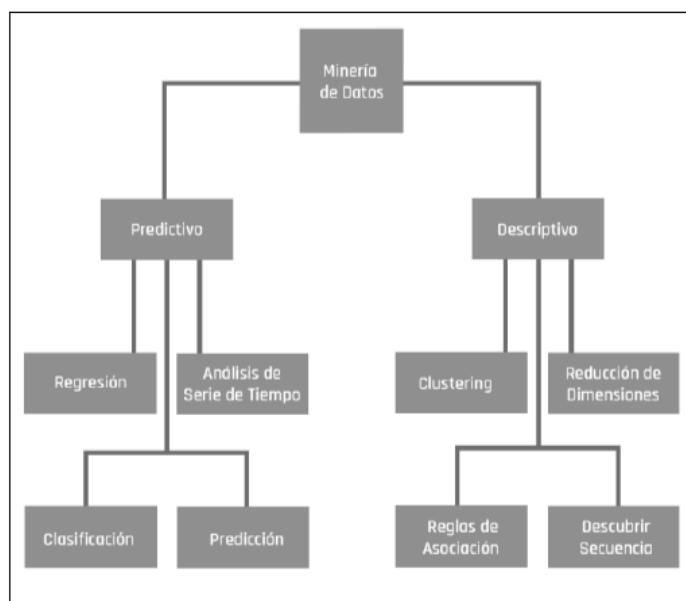


Figura 3.4 Modelo descriptivo y predictivo y sus respectivos modelos.

Actualmente el valor de la información cada vez aumenta al de punto en transformarse en un activo estratégico para la competitividad de una empresa, su unidad y consistencia son importantes, pues de estas características depende una buena parte de la confiabilidad de la información seleccionada para tomar decisiones, a medida en la que captan datos de sus operaciones cotidianas tendrán la oportunidad de correlacionarlos y hacer descubrimientos que le ayuden a identificar posibles clientes, puntos de ventas, fraudes, entre otros.

3.7 Minería Web

Hoy en día internet constituye un factor de gran importancia, y ante su rápido crecimiento surgen nuevas competencias entre las organizaciones con el fin de mejorar la calidad de los servicios entregados en los sitios web. La minería web es una herramienta utilizada con el fin de descubrir información útil para el hallazgo de nuevos conocimientos, empleando la información entregada por los sitios.

Se ha definido como el empleo de las técnicas de la minería de datos para descubrir y extraer información automáticamente de la web. Algunos de sus campos de aplicación son los motores de búsqueda y el comercio electrónico. Existen 3 clases fundamentales que se detallarán a continuación.

- **Minería Web de Contenido:** Intenta extraer información relevante sobre el contenido que posee el sitio web, con el fin de mejorar la organización de este y así poder perfeccionar el proceso de acceso y recuperación de la información.
- **Minería web de Estructura:** Enfocada en conocer la organización de la web, su estructura y cómo se puede navegar en ella.
- **Minería de uso Web:** Su objetivo es poder extraer patrones de uso de la web por parte de los usuarios, utilizando los registros de los sucesos y eventos. Este tipo tiene 2 objetivos principales:
 - a. Extraer patrones generales con el fin de mejorar el acceso para los usuarios y facilitar el uso de estos.
 - b. Extraer perfiles de usuarios de acuerdo a su comportamiento en la web para posteriormente entregar una atención personalizada. [19]

Se dará una explicación más exhaustiva de la técnica de minería de uso web, por ser la utilizada en este proyecto explicando su funcionamiento y su uso.

La minería de uso Web presenta cuatro fases fundamentales, las cuales se muestran en la figura 3.5.



Figura 3.5 Fases de la minería de uso web

- Recolección de datos- búsqueda. Consiste en la recuperación automática de la información relevante para su posterior procesamiento.
- Procesamiento de los datos. Una vez recuperados los documentos, se ordenan y se preparan para la próxima etapa; se utilizan herramientas para obtener información valiosa en forma automática.
- Descubrimiento de patrones. Existen múltiples técnicas, aplicables al descubrimiento de patrones. Entre ellas, para el agrupamiento y clasificación, para el establecimiento de reglas de asociación y el hallazgo de secuencias frecuentes.
- Análisis de patrones. Comprende la interpretación y validación de los patrones. [19].

4 Caso de Estudio

El caso de estudio elegido para este proyecto corresponde a la agencia virtual de viajes HostelWorld, la cual ofrece reservas a diferentes hostales a través del mundo. Opera en 19 idiomas y posee más de ocho millones de opiniones a través de 33.000 propiedades en más de 170 países. Su público objetivo son los viajeros menores a 35 años ya que el 90% de los usuarios está bajo ese rango de edad. Los usuarios de esta plataforma no son sus turistas promedios; solo anhelan experiencias, no únicamente recuerdos y quieren estar en el corazón de la acción. Quieren conocer el mundo, y es la naturaleza social de los albergues que ayudaran a facilitar las aventuras de sus viajes [20].

Las dimensiones requeridas por HostelWorld para las críticas de sus usuarios corresponden a ubicación, atmósfera, relación calidad-precio, limpieza, seguridad, servicios y personal. A continuación, se entregará un detalle de cada una.

- **Ubicación:** Categoría muy importante para los turistas antes de viajar ya que influye si queda lejos o cerca de los aeropuertos y además de los atractivos turísticos que se encuentren cercanos al hostel.
- **Atmósfera:** Se refiere a la convivencia que se tiene con los demás huéspedes de este y que tan bueno puede ser la convivencia dentro del hostel, por un viaje más a gusto.
- **Relación Calidad-Precio:** Para muchos turistas saber que se paga un precio para cierta calidad es importante.
- **Limpieza:** Calidad de limpieza que presenta el lugar. Esto se debe a algo de obviedad, una persona busca un lugar donde sentirse reconfortante, la limpieza y orden es fundamental para dar esta sensación.
- **Seguridad:** Uno de los puntos importante que toman las personas, sin contar como será la seguridad del lugar visitado, es la seguridad del hostel.
- **Servicios:** Se refiere a la cantidad de servicios ofrecidos por un cierto hostel, ya sea si tiene desayuno incluido y otros servicios básicos como lo son el internet y la televisión.
- **Personal:** Se evalúa el servicio entregado por el personal del hostel, ya que es muy importante como son atendidos los turistas los clientes.

Una vez finalizada la estadía en el hostel, el sitio presenta la encuesta de calidad al cliente, el cual debe evaluar cada una de las 7 dimensiones con una valoración que va desde el 2 al 10. El sitio muestra una calificación de cada hostel la cual es calculada en base al porcentaje de valoraciones que han dejado los usuarios durante los últimos 12 meses.

5 Trabajo Realizado

En primer lugar, para poder ver si era factible la extracción de datos en el sitio web HostelWorld, se procedió a ver la estructura de sus comentarios. En la figura 5.1 se muestra una opinión de un usuario del sitio. Al crear un perfil en el sitio web viene configurada la opción de tener un perfil como usuario anónimo o en caso de que el usuario quiera puede mostrar su nombre de manera pública, también se muestra la fecha del comentario, la valoración general promedio y las valoraciones de las 7 dimensiones que califica el usuario del sitio web, las cuales fueron mencionadas anteriormente.



Figura 5.1 Comentario de un usuario de HostelWorld.

5.1 Herramienta de Extracción de Datos

Para realizar la extracción de datos, se utilizó la librería Curl de PHP, la cual tiene como funcionalidad realizar scrapping web, que es extraer datos desde un sitio web. Además, se utilizó la librería PHPEXcel para transformar los datos obtenidos en una tabla de Excel.

En primera instancia la herramienta al recibir los datos que el usuario ingresa mediante un sistema de combo box, procesa los datos mediante una función enviada por Ajax para realizar el envío en tiempo real sin recargar. Una vez recibida la información en otro script de PHP, se arma la url para su posterior extracción, mediante las librerías mencionadas anteriormente, estas son procesadas en una función que extraerá la información requerida (figura 5.2 – estructura de los datos extraídos), las cuales enviará página por página hasta llegar al final de la línea de comentarios, extrayendo hasta el último que esté disponible en el sitio.

Tabla 5.2 Estructura de los datos extraídos.

Etiqueta HTML	Descripción
<code><h1 class="hwta-property-name" data-name=" "></h1></code>	Contiene el nombre del hostel.
<code><li class="reviewername blue"> </code>	Contiene el nombre del usuario.
<code><div class="reviewtext translate" id="review8624012"> </div></code>	Contiene el comentario que deja el usuario sobre el hostel.
<code><li class="reviewerdetails"> </code>	Contiene el país, grupo de viaje o género en caso de que sea solo una persona y por último el rango de edad del usuario.
<code><div class="textrating "></div></code>	Contiene la valoración promedio de las calificaciones que ha realizado el usuario.
<code><ul class="ratingbreakdown no-bullet"> Calidad- Precio Seguridad Ubicación Servicios </code>	Clasificación de las dimensiones Calidad-Precio, Seguridad Ubicación y Servicios.
<code><ul class="ratingbreakdown no-bullet"> Personal Atmósfera Limpieza </code>	Clasificación de las dimensiones Personal, Atmósfera y Limpieza.

Además, esta información que se extrae mediante la etiqueta se trabajará de forma simultánea extrayendo un comentario completo a la vez, mientras se trabaja la herramienta para agilizar el proceso. Cabe destacar que estas etiquetas son procesadas por las librerías para realizar la subida de los comentarios al archivo Excel, se aprecia en la imagen como cada una de las etiquetas son guardadas en una variable para luego entrar en proceso de extracción. Figura 5.3 procesamiento de etiquetas.

```

//Se extrae: Nombre de los comentarios
$nombreComentario = $html->find('li.reviewername');

if( $nombreComentario ){
    //Se extrae: El Pais, Genero, Edad
    $paisGeneroEdad = $html->find('li.reviewerdetails');
    //Se extrae: El comentario
    $comentario = $html->find('div.reviewtext');
    //Se extrae: las valoraciones por comentario
    $valoracionComentario = $html->find('ul.ratingbreakdown > li');
    //Se extrae: El nombre del hotel
    $nombreHotel = $html->find('h1.hwta-property-name');
    //Se extrae: La valoración general
    $valoracionGeneral = $html->find('div.textrating');
    //Se extrae: La fecha del comentario
    $fechaComentario = $html->find('span.reviewdate');
}

```

Figura 5.3 Procesamiento de Etiquetas.

En la figura 5.4 se observa la imagen de la interfaz principal de la herramienta que presenta 3 combo box donde se ingresarán el país, ciudad y hostel, en la cual se puede extraer tanto por la ciudad como por hostel por separado, una vez enviada la información se iniciará a generar un archivo Excel con toda la información desde la fecha del comentario hasta las valoraciones (figura 5.5), una vez que se ha terminado de crear el archivo Excel y se ha descargado automáticamente, hay que refinar los datos de la forma más conveniente que se requiera para su análisis. Cabe destacar que muchos comentarios, tanto nombre o tipo de viajero, están en blanco ya los comentarios antiguos no tienen todos los campos completos, además de valoraciones que posiblemente no estén completamente ingresadas y en este caso se visualizará un “0”.

Se quiere destacar que la aplicación tiene ciertas limitaciones, se quiere clarificar que está habilitada solamente para países de Latinoamérica, y esto es debido a que los link del sitio no son totalmente genéricos para rescatar todas ciudades y hostales de cada uno de los países, siendo un trabajo que podría tomar más tiempo del debido, en la figura 5.5 se podrá exponer un ejemplo:

Para entender bien, el script que se ha creado para la extracción de todos los hostales de cada ciudad funciona con la url, esta fue diseñada para trabajar con todos los países de Latinoamérica, el problema con otro continente, es que la url deja de ser genérica y cambia dependiendo del estado, distrito, etc. Por esto mismo rediseñar el script puede tomar más tiempo. A continuación, se muestra la url genérica con Chile, y la otra url que es la de Estados Unidos.

- <http://www.spanish.hostelworld.com/Albergues/Valparaiso/Chile>
- <http://www.spanish.hostelworld.com/alojamiento/zona/Massachusetts/Estados-Unidos>

EXTRAER DATOS NUNCA FUE TAN FÁCIL

VIVE LA EXPERIENCIA CON NOSOTROS!

Pais:

Ciudad:

Hostal:

LIMPIAR TEMPORALES

EXTRAER DATOS

CANCELAR

Figura 5.4 Interfaz Herramienta Web.

Fecha	Nombre	Pais	Género	Edad	Hostal	Ciudad	Valoración G	Calidad-Prec	Seguridad	Ubicación	Servicios	Personal	Atmósfera	Limpieza	Comentario
30nov2016	Anónimo	El Salvador	Pareja	25-30	Art-Factory-f	Buenos-Aires	9,7	10	10	10	8	10	10	10	Very nice p
27nov2016	vblanger	EstadosUnid	Mujer	18-24	Art-Factory-f	Buenos-Aires	8,9	8	10	10	10	10	6	8	The hostel
24nov2016	Anónimo	Países Bajos	Hombre	18-24	Art-Factory-f	Buenos-Aires	9,4	10	10	10	8	10	10	10	It was a re
24nov2016	badamson94	Inglaterra	Todo el grupo	18-24	Art-Factory-f	Buenos-Aires	8	8	10	10	8	8	8	6	Nice hoste
19nov2016	Anónimo	Brasil	Hombre	31-40	Art-Factory-f	Buenos-Aires	9,4	10	10	8	10	8	10	10	Mui bueno
19nov2016	Anónimo	Argentina	Mujer	25-30	Art-Factory-f	Buenos-Aires	10	10	10	10	10	10	10	10	Excelente p
15nov2016	chessiesarge	Inglaterra	Mujer	18-24	Art-Factory-f	Buenos-Aires	8,3	6	8	10	10	8	10	8	The Art Fac
15nov2016	Anónimo	Ecuador	Hombre	25-30	Art-Factory-f	Buenos-Aires	8,9	10	10	10	10	8	10	4	Overall it v
11nov2016	Anónimo	Canadá	Mujer	18-24	Art-Factory-f	Buenos-Aires	10	10	10	10	10	10	10	10	Amazing ho
8nov2016	Anónimo	Australia	Hombre	25-30	Art-Factory-f	Buenos-Aires	8	8	8	8	8	10	4	8	A really nic
5nov2016	rhysegbuehr	Australia	Hombre	18-24	Art-Factory-f	Buenos-Aires	8	8	8	8	8	8	8	8	Good host
4nov2016	Anónimo	Australia	Pareja	25-30	Art-Factory-f	Buenos-Aires	9,4	10	10	10	8	8	10	10	Really like
2nov2016	Anónimo	Países Bajos	Mujer	25-30	Art-Factory-f	Buenos-Aires	9,4	10	10	10	10	8	10	10	I really like
1nov2016	Anónimo	Argentina	Hombre	31-40	Art-Factory-f	Buenos-Aires	9,4	10	10	8	10	10	10	10	Great locat
30oct2016	Anónimo	Francia	Mujer	25-30	Art-Factory-f	Buenos-Aires	7,1	8	10	10	10	4	6	6	Very noisy
26oct2016	Anónimo	Colombia	Hombre	31-40	Art-Factory-f	Buenos-Aires	4,3	2	10	2	2	2	2	2	La verdad c
26oct2016	Anónimo	Colombia	Mujer	18-24	Art-Factory-f	Buenos-Aires	10	10	10	10	10	10	10	10	Es un excel
20oct2016	Anónimo	Noruega	Mujer	25-30	Art-Factory-f	Buenos-Aires	10	10	10	10	10	10	10	10	I really enj
16oct2016	Anónimo	Suiza	Hombre	25-30	Art-Factory-f	Buenos-Aires	8	8	8	10	6	6	6	8	Es war mei
12oct2016	Anónimo	EstadosUnid	Mujer	18-24	Art-Factory-f	Buenos-Aires	8,3	8	6	10	8	10	10	6	Super frier
7oct2016	Anónimo	Canadá	Hombre	18-24	Art-Factory-f	Buenos-Aires	5,4	2	10	8	2	2	4	10	The 6 man

Figura 5.5 Archivo Excel Extraído.

5.2 Análisis de Datos

Cuando se realiza un estudio de investigación, se pretende generalmente inferir o generalizar resultados de una muestra a una población. Se estudia en particular a un reducido número de individuos a los que tenemos acceso con la idea de poder generalizar los hallazgos a la población de la cual esa muestra procede. Este proceso de inferencia se efectúa por medio de métodos estadísticos basados en la probabilidad [21].

Las opiniones que posee el sitio web HostelWorld corresponden a datos cuantitativos y cualitativos, para este análisis se utilizaron los datos cuantitativos (valoraciones) entregadas por los usuarios, los cuales evalúan las siguientes dimensiones.

- D0: Calidad-Precio
- D1: Seguridad

- D2: Ubicación
- D3: Servicios
- D4: Personal
- D5: Atmósfera
- D6: Limpieza

A pesar de que se cuenta con la dimensión Valoración General, la cual fue extraída pero no analizada junto a las demás dimensiones ya que se trata de un promedio calculado por el sitio web de acuerdo a la valoración de todas las dimensiones entregadas por el usuario. Las valoraciones cuantitativas dadas por el usuario se muestran a continuación junto con las puntuaciones que estos pueden indicar.

- 2: Malo
- 4: Tolerable
- 6: Bueno
- 8: Muy bueno
- 10: Excelente

Se analizaron datos extraídos correspondientes a hostales de diferentes ciudades del país y también de países extranjeros. En primer lugar, el análisis sólo abarcaba a las ciudades de Viña del Mar y Valparaíso, con datos extraídos en abril de 2016, los resultados se encuentran disponibles en el Anexo A.

Con el fin de actualizar los datos se realizó una nueva extracción en el mes de diciembre 2016, y se extrajeron comentarios de las ciudades de Valparaíso, Viña del Mar y Santiago, correspondientes a Chile y también capitales de otros países de Latinoamérica, elegidas aleatoriamente, como Buenos Aires, Lima, Asunción y La Paz, estas últimas con el fin de comparar a Chile con otros países de América del Sur.

Para los análisis se contó con un total de 42.108 comentarios de diferentes hostales, cabe destacar que sólo fueron considerados los hostales que contaban con más de 5 valoraciones y una vez obtenidos estos datos, se realizó una etapa de refinación de los datos, considerando solo los que tenían evaluadas todas las dimensiones requeridas y los campos completos, ya sea su rango de edad o país de procedencia, entre otros, por lo tanto la cantidad de comentarios analizados fue de 28.794.

5.2.1 Matriz de Correlación

Se realizó el test de Spearman el cual se basa en los rangos de los datos en lugar de hacerlo en los valores reales. Resulta apropiada para datos ordinales, o los de intervalo que no satisfagan el supuesto de normalidad. Los valores del coeficiente van desde -1 a 1. El signo del coeficiente indica la dirección de la relación y el valor absoluto del coeficiente de correlación indica la fuerza de la relación entre las variables. Los valores absolutos mayores indican que la relación es mayor.

El coeficiente de correlación posee las siguientes características:

- El valor del coeficiente de correlación es independiente de cualquier unidad usada para medir las variables.
- El valor del coeficiente de correlación se altera de forma importante ante la presencia de un valor extremo, como sucede con la desviación típica. Ante estas situaciones conviene realizar una transformación de datos que cambia la escala de medición y modera el efecto de valores extremos (como la transformación logarítmica).
- El coeficiente de correlación mide solo la relación con una línea recta. Dos variables pueden tener una relación curvilínea fuerte, a pesar de que su correlación sea pequeña. Por tanto cuando analicemos las relaciones entre dos variables debemos representarlas gráficamente y posteriormente calcular el coeficiente de correlación.
- El coeficiente de correlación no se debe extrapolar más allá del rango de valores observado de las variables a estudio ya que la relación existente entre X e Y puede cambiar fuera de dicho rango.
- La correlación no implica causalidad. La causalidad es un juicio de valor que requiere más información que un simple valor cuantitativo de un coeficiente de correlación.

Se explicará el rango en que oscila la variable ρ , siendo valores entre -1 y 1, los cuales en el caso de ser positivos se habla de una correlación directa y en el caso de resultar negativo, se trataría de una correlación inversa. A continuación, se detalla la escala utilizada para evaluar que tan fuerte es la correlación:

- 0 - 0,19: Muy Baja
- 0,20 - 0,39: Baja
- 0,40 - 0,59: Moderada
- 0,60 - 0,79: Alta
- 0,80 – 1,00: Muy Alta

Se planteó la siguiente hipótesis con el fin de comprobar que las dimensiones analizadas estuvieran correlacionadas o no.

H_0 : $\rho = 0$, las dimensiones D_m y D_n son independientes.

H_1 : $\rho \neq 0$, las dimensiones D_m y D_n son dependientes.

Tabla 5.2 Matriz de correlación de Spearman Santiago

	D0: Calidad- Precio	D1: Seguridad	D2: Ubicación	D3: Servicios	D4: Personal	D5: Atmósfera	D6: Limpieza
D0	1	0.484	0.335	0.648	0.557	0.604	0.602
D1		1	0.339	0.523	0.459	0.426	0.537
D2			1	0.321	0.332	0.316	0.297
D3				1	0.517	0.571	0.634
D4					1	0.612	0.489
D5						1	0.511
D6							1

Como se observa en la tabla 5.2, matriz de correlación correspondiente a Santiago, todas las dimensiones están correlacionadas positivamente. En todos los casos las correlaciones son significativas ya que el valor p es menor al nivel de significancia elegido ($\alpha=0.05$). Se puede apreciar que existen correlaciones muy variadas entre todas las dimensiones.

- La dimensión Calidad-Precio (D0), está altamente correlacionada con las dimensiones D3, D5 y D6, pero moderadamente correlacionada con D1 y D4, y débilmente con D2.
- La Seguridad (D1), está débilmente correlacionada con Ubicación, pero moderadamente correlacionada con Servicios, Personal, Atmósfera y Limpieza.
- Ubicación (D2), está moderadamente correlacionadas con Servicio (D3), Personal (D4), Atmósfera (D5) y altamente correlacionada con Limpieza (D6)
- Servicios (D3), está moderadamente correlacionada con D4, D5 y D6
- D4 (Personal), está fuertemente correlacionada con Atmósfera (D5) y moderadamente correlacionada con Limpieza (D6).

Tabla 5.3 Matriz de correlación de Spearman Lima

	D0: Calidad- Precio	D1: Seguridad	D2: Ubicación	D3: Servicios	D4: Personal	D5: Atmósfera	D6: Limpieza
D0	1	0.447	0.391	0.668	0.551	0.613	0.625
D1		1	0.395	0.495	0.468	0.460	0.515
D2			1	0.365	0.350	0.415	0.341
D3				1	0.512	0.612	0.656
D4					1	0.585	0.500
D5						1	0.577
D6							1

Como se observa en la tabla 5.3, matriz de correlación correspondiente a la ciudad de Lima, todas las dimensiones están correlacionadas positivamente. En todos los casos las correlaciones son significativas ya que el valor p es menor al nivel de significancia elegido

($\alpha=0.05$). Se puede apreciar que existen correlaciones muy variadas entre todas las dimensiones.

- La dimensión Calidad-Precio (D0), está fuertemente correlacionada con las dimensiones, D3, D5 y D6 y moderadamente correlacionada con D1 y D4 y débilmente con D2.
- La Seguridad (D1), está moderadamente correlacionada con D3, D4, D5 y D6. Y débilmente con D2.
- Ubicación (D2), está débilmente correlacionadas con Servicio (D3), Personal (D4) y Limpieza (D6), pero moderadamente con Atmósfera (D5).
- Servicios (D3), está fuertemente correlacionada con D5 y D6, pero moderadamente correlacionada con D4.
- D4 (Personal), está moderadamente correlacionada con D5 y D6.

5.2.2 Prueba de Kruskal Wallis

Se utilizó el test de Kruskal Wallis para comprobar si existen diferencias entre las opiniones de distintos viajeros ya sea comparándolos de acuerdo a su rango de edad, grupo de viajero, entre otros. Para todos los casos se utilizó el valor $p \leq 0.05$ como regla de decisión.

5.2.2.1 De Acuerdo a Diferentes Capitales de América del Sur

Los datos analizados corresponden a hostales de las ciudades de Santiago, Lima, Asunción, La Paz y Buenos Aires. Se realizó el test para comprobar si existen diferencias entre las opiniones de las distintas ciudades. A continuación, se detallan la cantidad de hostales y valoraciones por cada una.

- Santiago (44 hostales, 9127 valoraciones)
- Lima (26 hostales, 5778 valoraciones)
- Asunción (8 hostales, 495 valoraciones)
- La Paz (20 hostales, 3990 valoraciones)
- Buenos Aires (24 hostales, 5949 valoraciones)

Se utilizó el test con el fin de comprobar o rechazar alguna de las siguientes hipótesis:

H_0 = No hay diferencias significativas entre las opiniones de los viajeros hospedados en los hostales de diferentes capitales.

H_1 = Hay diferencias significativas entre las opiniones de los viajeros hospedados en los hostales de diferentes capitales.

Tabla 5.4 Prueba de Kruskal Wallis sobre hostales de capitales de América del Sur

	D0: Calidad- Precio	D1: Seguridad	D2: Ubicación	D3: Servicios	D4: Personal	D5: Atmósfera	D6: Limpieza
p-valor	0,000	0,000	0,000	0,000	0,000	0,000	0,000

En la tabla 5.4 se observa que existen diferencias significativas entre todas las dimensiones, se deduce de esto que, al ser ciudades de distintos países, y con un enfoque de viajero diferente, fueran los resultados esperados. Por lo tanto, con estos datos no se podría obtener un patrón o tendencia y sería necesario seguir realizando otros tipos de estudios.

5.2.2.2 De Acuerdo a Tipo de Viajero en Hostales de Santiago

Los viajeros están clasificados por Hostelworld en 6 tipos, los cuales están descritos a continuación, junto con la cantidad de valoraciones obtenidas por cada grupo.

- Mujer (3377 valoraciones)
- Hombre (3806 valoraciones)
- Grupo Mixto (288 valoraciones)
- Grupo Mujeres (216 valoraciones)
- Grupo Hombres (125 valoraciones)
- Pareja (1315 valoraciones)

El test fue realizado para comprobar alguna de las siguientes hipótesis:

H_0 = No hay diferencias significativas entre las opiniones de los distintos tipos de viajeros.

H_1 = Hay diferencias significativas entre las opiniones de los distintos tipos de viajeros.

Tabla 5.5 Prueba de Kruskal Wallis de acuerdo al Tipo de Viajero

	D0: Calidad- Precio	D1: Seguridad	D2: Ubicación	D3: Servicios	D4: Personal	D5: Atmósfera	D6: Limpieza
p-valor	0,000	0,000	0,000	0,000	0,000	0,000	0,000

Como se observa en la tabla 5.5, de acuerdo al valor p elegido existen diferencias significativas sobre todas las dimensiones analizadas. Anteriormente se había realizado el mismo test en la quinta región (Anexo A.2) y los resultados fueron distintos, ya que en esta oportunidad, se encontró que los tipos de viajeros perciben de forma distinta todas las dimensiones. Por lo tanto, no se pueden encontrar patrones o tendencias con este tipo de prueba.

5.2.2.3 De Acuerdo al Grupo de Edad en Hostales de La Paz

Los datos analizados corresponden a la ciudad de La Paz, Bolivia ya que el sitio web separa a los viajeros en 4 rangos de edad, los cuales se detallan a continuación junto con la cantidad de hostales y valoraciones obtenidas por cada uno.

- 18-24 años (1103 valoraciones)
- 25-30 años (1901 valoraciones)
- 31-40 años (723 valoraciones)
- 41+ años (263 valoraciones)

Se crearon las siguientes hipótesis con el fin de comprobar si existen diferencias entre las opiniones.

H_0 = No hay diferencias significativas entre las opiniones de los viajeros de acuerdo a su grupo de edad.

H_1 = Hay diferencias significativas entre las opiniones de los viajeros de acuerdo a su grupo de edad.

Tabla 5.6 Prueba de Kruskal Wallis de acuerdo al Grupo de Edad de los Viajeros

	D0: Calidad- Precio	D1: Seguridad	D2: Ubicación	D3: Servicios	D4: Personal	D5: Atmósfera	D6: Limpieza
p-valor	0,089	0,303	0,003	0,126	0,000	0,000	0,277

En la tabla 5.6 se aprecia que Calidad-Precio (D0), Seguridad (D1), Servicios (D3) y Limpieza (D6) son percibidas de una manera similar por los distintos viajeros por lo tanto se deduce que existe un patrón, sin embargo, la Ubicación (D2), Personal (D4) y Atmósfera (D5) arrojan diferencias entre las opiniones. También se observa que la mayoría de los viajeros son menores de 30 años y por lo tanto se espera que estas opiniones difieran por ejemplo de los viajeros de más de 41 años.

5.2.2.4 De Acuerdo al Mes de Hospedaje en Hostales de Buenos Aires

Los comentarios fueron separados de acuerdo a la fecha en que se realizó la valoración de los usuarios ya que esta se extrae junto a todos los datos. Cabe mencionar que el sitio web invita al usuario a dejar una retroalimentación una vez terminada su estadía en el hostel, y este dispone del plazo de 1 semana para realizarlo, por lo tanto, la fecha de estadía con la del comentario no varían en más de 7 días. En esta oportunidad se analizó a la ciudad de Buenos Aires y la cantidad de comentarios obtenidos se detallan a continuación.

- Enero (637 valoraciones)
- Febrero (604 valoraciones)
- Marzo (470 valoraciones)

- Abril (466 valoraciones)
- Mayo (372 valoraciones)
- Junio (333 valoraciones)
- Julio (510 valoraciones)
- Agosto (417 valoraciones)
- Septiembre (390 valoraciones)
- Octubre (471 valoraciones)
- Noviembre (615 valoraciones)
- Diciembre (563 valoraciones)

El test fue realizado para comprobar alguna de las siguientes hipótesis:

H_0 = No hay diferencias significativas entre las opiniones de los viajeros de acuerdo al mes de su estadía.

H_1 = Hay diferencias significativas entre las opiniones de los viajeros de acuerdo al mes de su estadía.

Tabla 5.7 Prueba de Kruskal Wallis de acuerdo al mes de hospedaje

	D0: Calidad- Precio	D1: Seguridad	D2: Ubicación	D3: Servicios	D4: Personal	D5: Atmósfera	D6: Limpieza
p-valor	0,000	0,000	0,000	0,000	0,000	0,000	0,000

Como se observa en la tabla 5.7, nuevamente se encontraron diferencias significativas en todas las dimensiones analizadas. Con esto se puede asumir que no existe un patrón entre estos usuarios.

5.3 Comparación con Tripadvisor.com y Atrapalo.com

En base al análisis realizado a HostelWorld en contraste a la investigación aplicada a los casos de estudio TripAdvisor [22] y Atrápalo, se realizaron análisis en Chile y Latinoamérica pudiéndose hacer una comparación con algunas de estas ciudades, obteniéndose los siguientes resultados:

Para la ciudad de Santiago las comparaciones fueron las siguientes:

Similitudes:

- Para el test de Spearman con Atrápalo fueron muy similares los resultados, obteniendo una fluctuación entre moderada y altamente correlacionada. Destacando que la ubicación para los viajeros se mantiene uniforme y la evalúan de forma similar.

- En el test de Spearman con TripAdvisor los valores fluctúan tanto en moderado como altamente correlacionado, obteniendo resultados bastantes similares.
- Para el análisis de Krukal-Wallis, los datos obtenidos en los 3 casos de estudio, todos presentaron diferencias significativas en cada una de las dimensiones, por lo tanto, los usuarios perciben cada dimension de forma diferente.

Diferencias:

- Es importante destacar que la dimensión D2: Ubicación es la que menos correlacionada está en comparación para la ciudad de Santiago.

Para la ciudad de Lima las comparaciones fueron las siguientes:

Similitudes:

- Para el test de Spearman, los valores fueron bastante similares, ya que para Tripadvisor se obtuvieron valores que fluctúan entre bajo y alto, se entiende que los usuarios valorizan de forma uniforme.

Para la región de Valparaíso las comparaciones fueron las siguientes:

Similitudes:

- Para el test de Krukal-Wallis, el valor p en Atrápalo, se encontraron diferencias significativas en varias dimensiones, dando como resultado una similitud con HostelWorld y sus usuarios.

6 Conclusiones

A modo de conclusión se puede señalar que, a pesar de no tener una definición exacta de la experiencia del cliente, las percepciones de estos son una base importante al momento de la toma de decisiones, entendiendo que la calidad y diseño de cada sistema, software o aplicación, son parte de un estudio que se centró en las necesidades de los clientes, para otorgarles una buena experiencia.

Otro punto importante a tomar en cuenta, es que son muchas las tendencias que se pueden sacar al estar manejando grandes cantidades de datos, son estas tendencias lo que ayudan a las organizaciones a entender donde se debe desarrollar de manera más eficiente y óptima, y tener en claro cuáles son los puntos más importantes para los usuarios. En este trabajo se han dado a conocer los diferentes conceptos que permiten entender como el usuario se desenvuelve en varios aspectos de las dimensiones que se estudiaron, para así poder entender donde se vincula más con el cliente.

De acuerdo a los análisis realizados se pudo comprobar con el test de Spearman y los test de Kruskal-Wallis, que las dimensiones evaluadas por los usuarios se encuentran correlacionadas, y que, separando a los usuarios de acuerdo a distintos grupos según los datos encontrados en el sitio, en algunos casos se encontraron diversas diferencias entre las opiniones sobre los distintos grupos generados.

Además, para la prueba de Kruskal-Wallis, los resultados de algunas ciudades como Santiago o Valparaíso, fueron valores que variaron, pero aun así obtuvieron una diferencia significativa, cabe destacar que la ubicación en varios análisis, fue la que se mantuvo sin diferencia significativa, pudiéndose inferir que los usuarios valorizan de forma muy similar la ubicación tomando un rango de valores uniforme.

En general, para la Test de Spearman como también para el análisis de Kruskal-Wallis, los valores fueron bastante simétricos tanto para Chile como algunas capitales de Latinoamérica, en mayor parte por la ubicación, personal y atmósfera, tomando en cuenta que para estos hostales la mayoría de los visitantes son personas jóvenes.

Como trabajo futuro se podría perfeccionar la herramienta, para poder alcanzar más destinos en el mundo, mejorar ciertas funcionalidades para una extracción más completa, y para el análisis, se espera poder alcanzar análisis entre los nuevos destinos para tratar de encontrar nuevos hallazgos significativos.

7 Referencias

- [1] Suárez-Álvarez, L., Río-Lanza, A.B., Vázquez-Casielles, R.: Las agencias de viaje virtuales: ¿Cómo analizar la calidad de e-servicio y sus efectos sobre la satisfacción del cliente? *Universia Business Review*, 24, pp. 122-143 (2009)
- [2] Rusu, V., Rusu, C., Guzman, D., Espinoza, D., Rojas, D., Roncagliolo, S., Quiñones, D.: Assessing the Customer eXperience Based on Quantitative Data: Virtual Travel Agencies. pp. 499-508 (2016)
- [3] Hernandez, R., Fernandez-Collado, C., Baptista, P.: *Metodología de la Investigación*. Editorial Mexicana pp. 276-346 (2016).
- [4] Alfaro, E., Velilla, J., Brunetta, H., Molina, C.: Customer Experience. Una vision multidimensional del marketing de experiences. *THD Coated*. pp. 035-090 (2015)
- [5] Gentile, C., Spiller, N., Noci, G.: How to Sustain the Customer Experience: An Overview of Experience Components that Co-create Value With the Customer. *European Management Journal*, 25 (5), pp. 395–410 (2007)
- [6] Laming, C., Mason, K.: Customer experience – An analysis of the concept and its performance in airline brands. *Research in Transportation Business & Management*, 10, pp. 15-25 (2014)
- [7] Thinkandsell.com: Experiencia del Cliente. Revisado por última vez el 18 de abril del 2016.
- [8] Valero, S.: Ciencia de Servicios como herramienta de innovación y creación de valor en tiempos de crisis. Congreso de Ingeniería de Organización Barcelona-Terrassa. pp. 1895-1899 (2009).
- [9] Martinez, A., Lázaro, P.; La Ciencia de los Servicios: Un nuevo enfoque para la innovación en compañías de servicios. *Universia Business Review* (15) pp. 120-127 (2007)
- [10] Takahiko, A.: What is Service Science? Research Report (2005)
- [11] Knapp Bjerén, A. La Experiencia del Usuario. En: Knapp Bjerén, A. (coord.). *La Experiencia del Usuario*. Madrid: Anaya Multimedia, 2003, ISBN 84-415-1044-X. pp. 126-176 (2003).
- [12] McClelland, I. 'User Experience' Design: A new form of design practice takes shape. CHI 2005, April 2–7, Portland, Oregon, USA, pp. 1096-1097. (2005).
- [13] Arhippainen, L., Tähti, M. Empirical Evaluation of User Experience in Two Adaptative Mobile Application Prototypes. *Proceedings of the 2nd International Conference on Mobile and Ubiquitous Multimedia*, 10–12 December (2003)
- [14] ISO 9241-11. (1998). Ergonomic requirements for office work with visual display terminals (VDT)s - Part 11 Guidance on usability.
- [15] Béltran Ferreras, Tesis doctoral: “Aplicación de usabilidad al proceso de desarrollo de páginas web”. Universidad Politecnica de Madrid. pp. 009-056. (2008)
- [16] ISO/IEC 9126-1:2001 Software engineering -- Product quality -- Part 1: Quality model.
- [17] Young, A; Park and Ulrike Gretzely: “Success Factors for Destination Marketing Web Sites: A Qualitative Meta-Analysis”, *Journal of Travel Research*, vol 46, pp. 46-63 (2007).

- [18] Hernández, J., Ramírez, M.J., Ramírez, C.: Introducción a la Minería de Datos. Pearson España ISBN: 84 205 4091 9. pp. 137-198. (2004).
- [19] Fuentes Reyes, S.; Ruiz Lobaina, M.: Minería Web: un recurso insoslayable para el profesional de la información. ACIMED, 16 (4). pp. 001-010. (2007)
- [20] Hostelworldgroup.com: Company Overview. Revisado por última vez el 19 de abril del 2016.
- [21] Pita Fernández, S. Uso de la estadística y la epidemiología en atención primaria. En: Gil VF, Merino J, Orozco D, Quirce F. Manual de metodología de trabajo en atención primaria. Universidad de Alicante. Madrid, Jarpyo Editores, S.A.; 115-161 (1997).
- [22] D. Espinoza y D. Rojas: Minería de opiniones de usuarios en sitios web orientados al servicio. Pontificia Universidad Católica de Valparaíso. Julio 2016.

Anexos

En este anexo, se presenta información complementaria a los análisis realizados. Se incluyen diversos análisis realizados a las valoraciones extraídas de la Región de Valparaíso, con el fin de comparar con los nuevos análisis realizados. Cabe mencionar que estos análisis fueron realizados con los datos extraídos en el mes de abril, año 2016.

A: Análisis Región de Valparaíso

A.1. Prueba de Kruskal Wallis

Se utilizó el test de Kruskal Wallis para comprobar si existen diferencias entre las opiniones de distintos grupos de viajeros ya sea comparándolos de acuerdo a su rango de edad, grupo de viaje, entre otros. Para todos los casos se utilizó el valor- $p \leq 0.05$ como regla de decisión.

A.2. De Acuerdo al Grupo de Viaje

Los viajeros están clasificados por HostelWorld en 6 tipos, los cuales están descritos a continuación, junto con la cantidad de valoraciones obtenidas por cada tipo.

- Mujer (1017 valoraciones)
- Hombre (988 valoraciones)
- Grupo Mixto (147 valoraciones)
- Grupo Mujeres (90 valoraciones)
- Grupo Hombres (30 valoraciones)
- Pareja (542 valoraciones)

Se utilizó el test con el fin de comprobar o rechazar alguna de las siguientes hipótesis:

H_0 = No hay diferencias significativas entre las opiniones de los distintos tipos de viajeros.

H_1 = Hay diferencias significativas entre las opiniones de los distintos tipos de viajeros.

Tabla A1.1 Prueba de Kruskal Wallis de acuerdo al grupo de viaje

	D0: Calidad- Precio	D1: Seguridad	D2: Ubicación	D3: Servicios	D4: Personal	D5: Atmósfera	D6: Limpieza
p-valor	0,005	0,022	0,582	0,540	0,008	0,028	0,096

Según el test y como se muestra en la tabla A1.1, existen diferencias significativas en las opiniones de los usuarios, sobre las dimensiones de Calidad-Precio (D0), Seguridad (D1), Personal (D4) y Atmósfera (D5).

A.2. Grupo de Edad

Los usuarios de HostelWorld se caracterizan por ser jóvenes y la mayoría está entre los 18 a 30 años de edad, el sitio web los separa en 4 grupos los cuales se detallan a continuación, junto con la cantidad de valoraciones obtenidas por cada grupo.

- 18-24 años (786 valoraciones)
- 25-30 años (1282 valoraciones)
- 31-40 años (567 valoraciones)
- 41+ años (179 valoraciones)

Se realizó el test con el fin de comprobar alguna de las siguientes hipótesis:

H_0 = No hay diferencias significativas entre las opiniones de los viajeros de acuerdo a su edad.

H_1 = Hay diferencias significativas entre las opiniones de los viajeros de acuerdo a su edad.

Tabla A1.2 Prueba de Kruskal Wallis de acuerdo al grupo de edad

	D0: Calidad- Precio	D1: Seguridad	D2: Ubicación	D3: Servicios	D4: Personal	D5: Atmósfera	D6: Limpieza
p-valor	0,401	0,054	0,681	0,422	0,122	0,079	0,276

Como se observa en la tabla A1.3, ningún valor es menor a 0,05 por lo tanto no existen diferencias significativas entre las opiniones de los usuarios de acuerdo a su edad.

A.3. De acuerdo al Continente de Procedencia

Ya que en las valoraciones se extrae el país de procedencia de los usuarios que dejan sus opiniones, estos fueron separados en grupos de acuerdo al continente de procedencia, se realizó de esta forma, ya que la mayoría de los usuarios son extranjeros y no se hizo la comparación por país ya que en algunos casos hay países con sólo 1 comentario. A continuación, se muestran los grupos creados de acuerdo a las nacionalidades de los usuarios.

- Latinoamérica (539 valoraciones)
- Norteamérica (558 valoraciones)
- Oceanía (308 valoraciones)
- Asia (66 valoraciones)
- Europa (1327 valoraciones)
- África (16 valoraciones)

Se crearon las siguientes hipótesis con el fin de comprobar si existen diferencias entre las opiniones de acuerdo al continente de los usuarios.

H_0 = No hay diferencias significativas entre las opiniones de los viajeros de acuerdo a su nacionalidad.

H_1 = Hay diferencias significativas entre las opiniones de los viajeros de acuerdo a su nacionalidad.

Tabla A1.3 Prueba de Kruskal Wallis de acuerdo a continente de procedencia

	D0: Calidad- Precio	D1: Seguridad	D2: Ubicación	D3: Servicios	D4: Personal	D5: Atmósfera	D6: Limpieza
p-valor	0,008	0,024	0,000	0,000	0,000	0,000	0,000

Como se observa en la tabla A1.3, existen diferencias en todas las dimensiones analizadas, con este resultado se rechaza la hipótesis nula, y se acepta H_1 , dejando clara la diferencia de opiniones entre los distintos continentes de los viajeros de HostelWorld.

A.4. De Acuerdo al Mes de Hospedaje

Los comentarios fueron separados de acuerdo a la fecha en que se realizó la valoración de los usuarios, cabe mencionar que el sitio web invita al usuario a dejar una retroalimentación una vez terminada su estadía en el hostel, y este dispone del plazo de 1 semana para realizarlo.

- Enero (420 valoraciones)
- Febrero (352 valoraciones)
- Marzo (392 valoraciones)
- Abril (281 valoraciones)
- Mayo (149 valoraciones)
- Junio (95 valoraciones)
- Julio (137 valoraciones)
- Agosto (116 valoraciones)
- Septiembre (139 valoraciones)
- Octubre (160 valoraciones)
- Noviembre (216 valoraciones)
- Diciembre (367 valoraciones)

El test fue realizado para comprobar alguna de las siguientes hipótesis:

H_0 = No hay diferencias significativas entre las opiniones de los viajeros de acuerdo al mes de su estadía.

H_1 = Hay diferencias significativas entre las opiniones de los viajeros de acuerdo al mes de su estadía.

Tabla A1.4 Prueba de Kruskal Wallis de acuerdo al mes de hospedaje

	D0: Calidad- Precio	D1: Seguridad	D2: Ubicación	D3: Servicios	D4: Personal	D5: Atmósfera	D6: Limpieza
p-valor	0,000	0,058	0,007	0,000	0,007	0,241	0,022

Como se observa en la tabla A1.4, de acuerdo al valor p elegido existen diferencias significativas sobre las dimensiones Calidad-Precio (D0), Ubicación (D2), Servicios (D3), Personal (D4) y Limpieza (D6).

B: Manual de Usuario y Documentación de la Herramienta.

Se ha desarrollado una aplicación para la extracción de datos del sitio “HostelWord.com”, para esto se ha desarrollado un manual de uso. En el presente manual se expondrán 2 tipos de manuales para los usuarios, estos siendo un usuario promedio y documentación de la herramienta que se expone el funcionamiento interno de la herramienta para futuros estudios. Para el manual de usuario promedio se expondrá la información lo más detallada posible con una explicación sencilla, en este manual de usuario se pretende guiar y facilitar mediante una explicación paso a paso de cómo funciona la herramienta propuesta incluyendo una serie de fotografías y explicaciones de estas.

Para la documentación de la herramienta se explicarán los tipos de tecnologías que se ocuparon desde un nivel grafico como a nivel de código, además del servidor en el cual se hicieron todas las pruebas necesario para su funcionamiento. Se añadirá un detalle con fotografías del código explicando las funciones que se ocuparon, y las etiquetas HTML que fueron necesarias, además se explicara del uso de la tecnología Ajax para la gran mayoría de las funciones y la ayuda de ciertas librerías que fueron necesarias incluir para el tratamiento de la información que se extraía del sitio web.

B.1. Manual de Usuario Promedio

Cuando se inicia la aplicación en la pantalla inicial, figura B1.1, se puede observar 3 tipos de categorías, País, Ciudad y Hostal. En este se puede extraer datos tanto por hostel como por ciudad completa como se observa en la figura B1.2.

EXTRAER NUNCA FUE TAN FÁCIL

VIVE LA EXPERIENCIA CON NOSOTROS!

País:

Ciudad:

Hostel:

Figura B1.1 Pantalla inicial.

En la imagen a continuación se observa que cuando se rellena el campo “País” se abre la casilla de “Ciudad” y en la figura 1.3 se observa que al ingresar el campo “ciudad”, podemos iniciar a extraer los datos de todos los hostales de esa ciudad seleccionada.

País:

Ciudad:

Hostel:

Figura B1.2 Selección categorías Ciudad.

País:

Ciudad:

Hostel:

LIMPIAR TEMPORALES **EXTRAER DATOS** **CANCELAR**

Figura B1.3 Selección categorías Hostel.

Una vez que se hayan completado todos los campos a selección y esté todo correcto, se dispondrá a extraer los datos para su debido análisis. Como se observa en la figura B1.4.



Figura B1.4 Datos completados.

Una vez empezado la extracción, la aplicación comenzará a trabajar e iniciará a extraer todos los datos necesarios de la ciudad con sus debidas valoraciones y datos del usuario, este proceso puede tomar aproximadamente entre 1 – 5 minutos dependiendo de la cantidad de comentarios existentes, mostrando el avance de cada hostel si fuera una ciudad completa, si no es el caso, solamente se mostrará por pantalla “Extrayendo Datos”.

Una vez completada la operación de extracción se visualizará un archivo en descarga de forma automática (Figura B1.5).

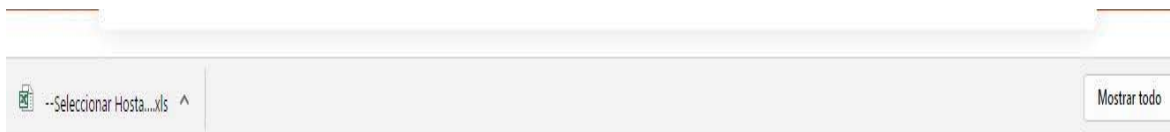


Figura B1.5 Descarga de hostel.

Una vez abierto el archivo Excel se observará que está dividido en 16 categorías que se podrán filtrar y manejar los datos de la forma más conveniente posible como sea necesario, estas categorías son:

- Fecha
- Nombre
- País
- Genero
- Edad
- Hostel
- Ciudad
- Valoración-General
- Calidad-Precio
- Seguridad
- Ubicación
- Servicio
- Personal
- Atmosfera
- Limpieza
- Comentario

Fecha	Nombre	País	Género	Edad	Hostal	Ciudad	Valoración G	Calidad-Prec	Seguridad	Ubicación	Servicios	Personal	Atmósfera	Limpieza	Comentario
30nov2016	Anónimo	El Salvador	Pareja	25-30	Art-Factory-f	Buenos-Aires	9,7	10	10	8	10	10	10	10	Very nice p
27nov2016	vblanger	EstadosUnid	Mujer	18-24	Art-Factory-f	Buenos-Aires	8,9	8	10	10	10	6	8	10	The hostel
24nov2016	Anónimo	Países Bajos	Hombre	18-24	Art-Factory-f	Buenos-Aires	9,4	10	10	8	10	10	10	8	It was a rei
24nov2016	badamson94	Inglaterra	Todo el grupo	18-24	Art-Factory-f	Buenos-Aires	8	8	10	8	8	8	6	8	Nice hoste
19nov2016	Anónimo	Brasil	Hombre	31-40	Art-Factory-f	Buenos-Aires	9,4	10	8	10	8	10	10	10	Mui bueno
19nov2016	Anónimo	Argentina	Mujer	25-30	Art-Factory-f	Buenos-Aires	10	10	10	10	10	10	10	10	Excelente p
15nov2016	chessiesarge	Inglaterra	Mujer	18-24	Art-Factory-f	Buenos-Aires	8,3	6	8	10	8	10	8	8	The Art Fax
15nov2016	Anónimo	Ecuador	Hombre	25-30	Art-Factory-f	Buenos-Aires	8,9	10	10	10	8	10	4	10	Overall it v
11nov2016	Anónimo	Canadá	Mujer	18-24	Art-Factory-f	Buenos-Aires	10	10	10	10	10	10	10	10	Amazing h
8nov2016	Anónimo	Australia	Hombre	25-30	Art-Factory-f	Buenos-Aires	8	8	8	8	10	4	8	10	A really nic
5nov2016	rhysegbuehr	Australia	Hombre	18-24	Art-Factory-f	Buenos-Aires	8	8	8	8	8	8	8	8	Good host
4nov2016	Anónimo	Australia	Pareja	25-30	Art-Factory-f	Buenos-Aires	9,4	10	10	8	8	10	10	10	Really like
2nov2016	Anónimo	Países Bajos	Mujer	25-30	Art-Factory-f	Buenos-Aires	9,4	10	10	10	8	10	10	8	I really like
1nov2016	Anónimo	Argentina	Hombre	31-40	Art-Factory-f	Buenos-Aires	9,4	10	8	10	10	10	10	8	Great locat
30oct2016	Anónimo	Francia	Mujer	25-30	Art-Factory-f	Buenos-Aires	7,1	8	10	10	4	6	6	6	Very noisy
26oct2016	Anónimo	Colombia	Hombre	31-40	Art-Factory-f	Buenos-Aires	4,3	2	10	2	2	2	2	10	La verdad c
26oct2016	Anónimo	Colombia	Mujer	18-24	Art-Factory-f	Buenos-Aires	10	10	10	10	10	10	10	10	Es un excel
20oct2016	Anónimo	Noruega	Mujer	25-30	Art-Factory-f	Buenos-Aires	10	10	10	10	10	10	10	10	I really enj
16oct2016	Anónimo	Suiza	Hombre	25-30	Art-Factory-f	Buenos-Aires	8	8	8	10	6	6	8	10	Es war mei
12oct2016	Anónimo	EstadosUnid	Mujer	18-24	Art-Factory-f	Buenos-Aires	8,3	8	6	10	8	10	10	6	Super frier
7oct2016	Anónimo	Canadá	Hombre	18-24	Art-Factory-f	Buenos-Aires	5,4	2	10	8	2	2	4	10	The 6 man

Figura B1.6 Excel Extraído.

Es necesario tomar en cuenta que el Excel que se extrae, en algunos casos los campos extraídos no estarán totalmente completados, ya que dependiendo del usuario puede que no rellene ciertos valores, y como la aplicación descarga y extrae todos los comentarios de dicho hostel puede que tenga valores vacíos, en el caso de que alguna dimensión no sea rellena por el usuario este campo se visualizara con un “0”. Ver figura B1.7.

19dic2010	Anónimo	Inglaterra			Hostal-Ca: Valparaíso	10	0	10	10	0	10	10	10	I loved staying in Casa Valparaíso The atmospl
17dic2010	Anónimo	Inglaterra			Hostal-Ca: Valparaíso	9,3	0	10	10	0	10	8	10	Staff were extremely helpful and welcoming i
15dic2010	Anónimo	Irlanda			Hostal-Ca: Valparaíso	9	0	8	10	0	10	10	8	Free breakfast of fresh fruit eggs and bread w
12dic2010	Anónimo	Canadá			Hostal-Ca: Valparaíso	8,7	0	6	10	0	10	10	8	Very nice place Great host good recommend
8dic2010	Anónimo	Alemania			Hostal-Ca: Valparaíso	9,7	0	10	10	0	10	8	10	good location very quite place extremely helpf
5dic2010	Anónimo	Canadá	Hombre	41+	Hostal-Ca: Valparaíso	9,7	0	10	10	0	10	10	8	Hostal Casa Valparaíso is an excellent choice i
25nov2010	Anónimo	Canadá	Grupomix	25-30	Hostal-Ca: Valparaíso	10	0	10	10	0	10	10	10	seriously an AMAZING hostel staff was so incr
12nov2010	Anónimo	Australia	Mujer	18-24	Hostal-Ca: Valparaíso	5,7	0	6	10	0	4	4	2	Was a lovely hostel however the staff did not
6nov2010	tanelle79	Australia			Hostal-Ca: Valparaíso	7,3	0	8	8	0	6	6	8	good location bit of a climb up stairs to get the

Figura B1.7 – Valoraciones vacíos.

Ahora si bien se quiere extraer los datos, pero de un solo hostel, el proceso es prácticamente el mismo, con la diferencia en que se ingresan todos los datos, como se observa en la imagen a continuación.

Pais:

Argentina

Ciudad:

La-Plata

Hostal:

El-Intercultural-Hostel

LIMPIAR TEMPORALES

EXTRAER DATOS

CANCELAR

Figura B1.8 – Campos completos.

El proceso es el mismo que el anterior, se procede a clicar el botón extraer, una vez que ese proceso se ha terminado se descargará el Excel con el hostel seleccionado y el usuario podrá trabajar los datos como él lo requiera.

La aplicación además permite cancelar la extracción si el usuario lo desee, una vez cancelado el proceso se alzará un mensaje emergente. Figura B1.9.



Figura B1.9 – proceso terminado.

Además, la aplicación permite eliminar todos los archivos temporales existentes de la carpeta origen, los cuales son los que se van almacenando cuando se van extrayendo tanto hostales como ciudades completas, existen 2 mensajes emergentes que se observan en la figura B1.11 y en la figura B1.12 en caso de que no existan archivos temporales.



Figura B1.11 – Pestaña limpiar temporales.



Figura B1.12 – Pestaña limpiar, error.

Por último, la aplicación permite actualizar la base de datos, este botón se encuentra en la esquina superior derecha de la aplicación. Esta funcionalidad sirve para ir mejorando y agregando como eliminando hostales, en el caso de que un hostel ya no exista no se encuentre en la base de datos, como también si se agrega un nuevo hostel al sitio.

Cuando se procede a actualizar la base de datos, el tiempo promedio puede ser desde 10 a 15 minutos dependiendo tanto de la conexión como factores del mismo sitio web de Hostelworld.

El proceso puede ser cancelado en la mitad de su actualización si el usuario lo deseara. Se observa en la imagen a continuación como se visualizaría todo lo mencionado anteriormente.



Figura B1.13 – Actualizar base de datos.

B.2. Documentación de la Herramienta.

En la documentación se explicarán diferentes funciones de forma explícita, dando una visión más avanzada de cómo funciona internamente la herramienta de extracción del sitio HostelWorld.com.

Principalmente como se mencionó anteriormente se explicará inicialmente el php de index de la herramienta, y las etiquetas que se utilizaron. Cabe destacar que para la interfaz se utilizó “Bootstrap” es un framework o conjunto de herramientas de Código abierto para diseño de sitios y aplicaciones web. Contiene plantillas de diseño con tipografía, formularios, botones, cuadros, menús de navegación y otros elementos de diseño basado en HTML y CSS, así como, extensiones de JavaScript opcionales adicionales.

En la figura B2.1 se aprecia diferentes etiquetas “div” que contienen una clase, ese tipo de clase con equivalente al bootstrap para darle un diseño más elegante e interactivo para el usuario, que además fueron modificados algunos valores para que se pudiera adaptar mejor a la herramienta y las dimensiones.

Se aprecia 3 select que equivales a una etiqueta HTML la cual contiene un id cada una de ellas las cuales se utilizaran luego para otro tipo de tecnología que se explicará más adelante en este manual, en esta se rescata la elección del usuario la cual más adelante esta información será procesada y trabajada en otro php. Por último, se observa al final del código un Button el cual realizara la acción y llamado para realizar este procesamiento de información la cual está ligado a otra llamada que se presentará cuando se haga click en “Extraer Datos”, la id llamada “enviarUrl” será la encargada de realizar el llamado a la función.

```
14 <div class="container">
15   <div class="col-md-8 text-center">
16     <p class="textoIndex">País:</p>
17     <select class="form-control" name="" id="paises"></select>
18   </div>
19   <div class="col-md-8 text-center">
20     <p class="textoIndex">Ciudad:</p>
21     <select class="form-control" name="" id="ciudadesSelect"></select>
22   </div>
23   <div class="col-md-8 text-center">
24     <p class="textoIndex">Hostel:</p>
25     <select class="form-control" name="" id="hostalesSelect"></select>
26   </div>
27   <div class="col-md-12 col-md-offset-9">
28     <br>
29     <button class="form-submit btn btn-special" id="enviarUrl">Extraer Datos</button>
30     <h2><a class="mdl-button" id="cargando">Extrayendo Datos  </a></h2>
31   </div>
32 </div>
```

Figura B2.1 Índex etiquetas principales.

Antes del procesamiento de la información enviada por el usuario, en cada uno de los select abra un combo box que se desplegará como se explicó en el manual anterior la cual esta información es enviada desde esta parte del código. Lo que se está utilizando aquí es una tecnología llamada Ajax.

Ajax es un acrónimo de Asynchronous JavaScript And XML (JavaScript asíncrono y XML), es una técnica de desarrollo web para crear aplicaciones interactivas o RIA (Rich Internet Applications). Estas aplicaciones se ejecutan en el cliente, es decir, en el navegador de los usuarios mientras se mantiene la comunicación asíncrona con el servidor en segundo plano. De esta forma es posible realizar cambios sobre las páginas sin necesidad de recargarlas, mejorando la interactividad, velocidad y usabilidad en las aplicaciones.

En la figura B2.2 se aprecia diferentes variables, en Ajax presenta un tipo que es Post que la forma en que llegaran los datos al php de autocomplete, que básicamente ese php está conectada con la base de datos donde contendrá todos los países, ciudades y hostales que presenta el sitio. Con la variable data se enviarán y recibirá la información. La primera acción que se realiza es asignar las ciudades y hostel a las variables, luego la variable options se asigna los países que se reciben de la base de datos, posterior a esto en la función llamada append que insertará el nombre de los países en la etiqueta options que es la que se mostrará en el select del combo box.

```
109 $.ajax({
110     type: "POST",
111     contentType: "application/json; charset=utf-8",
112     url: "autocomplete.php",
113     data: "{}",
114     dataType: "json",
115     success: function (data) {
116         //console.log(data);
117         ciudades = data.ciudad;
118         hostales = data.hostal;
119         var options = $("#países");
120         $.each(data.pais, function() {
121             options.append("<option />").val(this.id_pais).text(this.nombre));
122         });
123     },
124     error: function (result) {
125         //console.log(result);
126     }
127 });
```

Figura B2.2 – Ajax principal del autocompletado.

Para la figura B2.3 se aprecia diferentes funciones con países, ciudad y hostel. En este código cada función cumple un rol para cada uno de los select, para el primer caso países tiene 2 variables que es id_pais donde se almacenara el id del país seleccionado de la etiqueta html, luego una variable que es un array para almacenar las ciudades de ese país. Posterior a esto, en la función que sigue se hace una comparación entre el ítem seleccionado con el id del país para corroborar, y si todo es correcto se agregaran al array el id de la ciudad con sus nombres.

Cabe destacar que la función change que se aprecia, es para cuando se haya seleccionado el país se habilitara la casilla siguiente para la siguiente elección, y eso pasa exactamente lo mismo con CiudadesSelect que se realiza exactamente lo mismo que en países. En cambio, para Hostaselect que presenta un empty que funciona para dar acceso a la elección de la ciudad y hostel una vez seleccionado el país, de esta forma el usuario no tendrá acceso al siguiente combo box sin haber elegido antes la casilla anterior.

```

129 $( "#países" ).change(function() {
130     var id_pais = $("#países").val();
131     var ciudadesPais = [];
132
133     $.each(ciudades, function(i, item) {
134         if( item.id_pais == id_pais ){
135             ciudadesPais.push({
136                 id_ciudad : item.id_ciudad,
137                 nombre    : item.nombre
138             })
139         }
140     });
141
142     $("#hostalesSelect").empty();
143     var options = $("#ciudadesSelect");
144     options.empty();
145     options.append($("<option />").val(-1).text("--Seleccionar Ciudad--"));
146     $.each(ciudadesPais, function() {
147         options.append($("<option />").val(this.id_ciudad).text(this.nombre));
148     });
149
150 });
151
152 $( "#ciudadesSelect" ).change(function() {
153     var id_ciudad = $("#ciudadesSelect").val();
154     var hostalesCiudades = [];
155
156     $.each(hostales, function(i, item) {
157         if( item.id_ciudad == id_ciudad ){
158             hostalesCiudades.push({
159                 id_link : item.id_link,
160                 nombre  : item.nombre
161             })
162         }
163     });
164
165     var options = $("#hostalesSelect");
166     options.empty();
167     options.append($("<option />").val(-1).text("--Seleccionar Hostal--"));
168     $.each(hostalesCiudades, function() {
169         options.append($("<option />").val(this.id_link).text(this.nombre));
170     });
171
172 });
173

```

Figura B2.3— Ajax selección de combo box.

Se ha hecho una explicación previa antes del inicio de javascript para dar entendimiento de cómo funcionan el combo box, ahora se iniciará a dar explicación de cómo se procesa la información elegida por el usuario. En esta parte se observan (figura B2.4) diferentes variables que son las responsables para asignar los datos rescatados por el usuario. Posteriormente en “#enviarUrl” es la función que funcionará al momento en que se click por parte el usuario y la información se iniciará a rescatar como se aprecia en la imagen.

```

126 <script type="text/javascript">
127     $( document ).ready(function() {
128         console.log( "ready!" );
129
130         var ciudadSeleccionada;
131         var hostalselect;
132         var id_hostalSeleccionado;
133         var ciudades;
134         var hostales;
135         var eliminar;
136
137
138         // $("#barra").hide();
139         $("#cargando").hide();
140         $("#cargandoBd").hide();
141
142         $( "#enviarUrl" ).click(function() {
143
144             //var url = $("#countries").val();
145             ciudadSeleccionada = $("#ciudadesSelect option:selected").text();
146             hostalselect = $("#hostalesSelect option:selected").text();
147             id_hostalSeleccionado = $("#hostalesSelect").val();
148
149             ciudadSeleccionada = arreglarPalabras(ciudadSeleccionada);
150             hostalselect = arreglarPalabras(hostalselect);
151             console.log("nombre ciudad: " + ciudadSeleccionada);
152             console.log("nombre hostel: " + hostalselect);
153             console.log("id hostel: " + id_hostalSeleccionado);
154

```

Figura B2.4 Inicio JavaScript.

Ahora una vez que las variables ya están con la información requerida se iniciará Ajax para iniciar el envío de toda la información al php que se encuentra en la carpeta "php/generarExcel.php", este archivo procesará la información para generar el resultado final que es el Excel. Lo que realiza ese Ajax básicamente es asignar a las variables el nombre del hostel, de la ciudad y el id del hostel seleccionado para luego enviar esta información. Más adelante en este manual se explicará dónde y cómo se procesa esta información en el php antes mencionado. Una vez la información haya sido tratada correctamente Ajax descargará de forma automática el Excel una vez haya completado de extraer toda la información con la función `Windows.location.href`, además de asignarle el nombre del hostel seleccionado o país si fuera el caso.

```

var llamada = $.ajax({
    type: "POST",
    url: "php/generarExcel.php",
    data: {
        //url:url,
        nombre_hostal : hostalselect,
        nombre_ciudad : ciudadSeleccionada,
        id_hostal      : id_hostalSeleccionado
    },
    beforeSend: function(e){
        $("#barra").show();
        $("#cargando").show();
        $('#Info').fadeOut(1000);
    },
    success: function(data){
        $("#barra").hide();
        console.log(data);
        if(data == 1){
            window.location.href = 'php/' + ciudadSeleccionada + '.xls';
            $("#cargando").hide();

        }else{
            window.location.href = 'php/' + hostalselect + '.xls';
            //window.open('php/' + hostalselect + '.xls', '_blank' );
            $("#cargando").hide();
        }
    },
    error: function(error){
        console.log(error.responseText);
        if (error.responseText.search("(") != -1){
            alert("Hostal no tiene comentarios");
        }else{
            window.location.href = 'php/' + hostalselect + '.xls';
            //window.open('php/' + hostalselect + '.xls', '_blank' );
        }

        $("#cargando").hide();
    }
});

```

Figura B2.5 Ajax del envío de datos.

En esta parte del manual se iniciará a explicar el php que realiza las funciones de procesamiento de la información del usuario. Como se observa en la figura B2.6, se inicializan las variables a ocupar y se asignan los nombres de la ciudad, del hostal y su respectivo id si fuera el caso de que fuera un hostal, en caso contrario se verifica si es un país o hostal, en el if donde se verifica si presenta el texto (' --Selecccionar Hostal-- '), si la consulta es correcta se hace una consulta a la base de datos rescatando la id de la ciudad para luego realizar una segunda consulta para rescatar los nombres e ids de los hostales asignados a esa ciudad.

```

function programa(){
    ini_set('display_errors', 1);
    ini_set('display_startup_errors', 1);
    error_reporting(E_ALL);
    include_once('../conexion.php');
    include('PHPExcel.php');

    $result1 = array();
    $result2 = array();
    $resultado2 = array();
    $asd=[];
    $final=[];
    $lista_simple=[];
    $nombre_ciudad = $_POST['nombre_ciudad'];
    $id_hostal = $_POST['id_hostal'];
    $nombre_hostal = $_POST['nombre_hostal'];
    $div = 0;

    if($nombre_hostal == '--Seleccionar Hostal--'){

        $query = "SELECT id_ciudad from ciudad where nombre='$nombre_ciudad' ";
        foreach($conn->query($query) as $row ) {
            $id_pa = $row[0];
        }

        $query1 = "SELECT nombre,id_link from hostel where id_ciudad='$id_pa' ";

        foreach($conn->query($query1) as $row1 ){
            $result1[]=$row1['nombre'];
            $result2[]=$row1['id_link'];
        }
    }
}

```

Figura B2.6 Inicio función php Excel.

En la imagen a continuación B2.7, se inicia el for para ir recorriendo la cantidad de hostales que presenta esa ciudad, luego se limpia las palabras erróneas con la función “ArreglarPalabras”, la cual cambia palabras que le falten acentos o que presenten simbología no acorde a la requerida. Posteriormente se arma el link para iniciar la extracción de datos. En la siguiente función se inicia con el for de las iteraciones, estas corresponden a las páginas de comentarios presentadas en el sitio, y en el caso de que lleguen hasta el límite de las paginas antes de que terminen las 500 iteraciones se corta con la variable Global. Más Adelante se explicará la función obtenerDatos. Posterior al termino de las iteraciones se almacenan todos los comentarios de un hostel en la variable final, que además se almacenaran en lista_simple para limpiar la variable final para recibir los nuevos comentarios del siguiente hostel.

```

for($i=0; $i<count($result2); $i++){

$corte = 0;

$result1[$i] = arreglarPalabras($result1[$i]);

$link_a="http://www.spanish.hostelworld.com/hosteldetails.php/" . $result1[$i] . "/" . $nombre_ciudad . "/" . $result2[$i] . "/comentarios

$resultado2 = explode("page=", $link_a);

for ($a=1; $a < 500 ; $a++) {
    $link = $resultado2[0].'page=' . $a.substr($resultado2[1], 1);
    array_push($asd, obtenerDatos($link,$a,$nombre_ciudad,$result1[$i]) );

    if( $GLOBALS["corte"] == 1 ){
        $a=500;
    }
}

if($GLOBALS["corte"] != 2){

foreach ($asd as $paginaComentario) {
    foreach ($paginaComentario as $comentario) {
        array_push($final,$comentario);
    }
}

foreach ($final as $rows) {
    array_push($lista_simple,$rows);
}

while(count($final)){
    array_pop($final);
    reset($final);
    array_pop($asd);
    reset($asd);
}

}

}

$GLOBALS["div"] == 1;

```

Figura B2.7 función ciudad completa.

En el caso de que el usuario haya elegido un hostel, en vez de una ciudad completa el proceso es casi el mismo con la diferencia de que no es necesario realizar alguna consulta a la base de datos, simplemente están todos los datos para extraer toda la información requerida, como se observa en la imagen B2.8.

```

}
else
{
    $corte = 0;
    $link_a="http://www.spanish.hostelworld.com/hosteldetails.php/" . $nombre_hostal . "/" . $nombre_ciudad . "/" . $id_hostal . "/comentarios

    $resultado2 = explode("page=", $link_a);

    for ($a=1; $a < 500 ; $a++) {
        $link = $resultado2[0].'page=' . $a.substr($resultado2[1], 1);
        array_push($asd, obtenerDatos($link,$a,$nombre_ciudad,$nombre_hostal) );

        if( $GLOBALS["corte"] == 1 ){
            $a=500;
        }
    }

    foreach ($asd as $paginaComentario) {
        foreach ($paginaComentario as $comentario) {
            array_push($lista_simple,$comentario);
        }
    }

    $GLOBALS["div"] == 2;
}
}

```

Figura B2.8 función hostel completa.

En la función llamado Obtener datos, es donde se realizará la parte más importante de la extracción. En esta parte es donde se rescata la información del sitio para luego ser traspasado

al Excel, cuando se van recorriendo las páginas de comentarios del sitio, aquí llegan una por una y se van extrayendo los comentarios. En la función llega el número de página que se está extrayendo y el link, en este se procesa el link con la función:

- curl_init - Inicia una nueva sesión CURL.
- curl_setopt - Define opciones para nuestra sesión CURL.
- curl_exec - Ejecuta la petición http

Luego se extrae el nombre del comentario, el género, país, etc. Esto se realiza gracias a la etiqueta que presenta el sitio en donde se muestra la información, cabe destacar que esta parte del código es donde se realiza la extracción para luego ir siendo almacenada en el Excel donde se aprecia en la figura B2.9.

```
function obtenerDatos($link,$iteracion){
    //$output = fopen("output.csv", "a") or die("Unable to open file!");
    $ch = curl_init();
    curl_setopt( $ch, CURLOPT_URL, $link);
    curl_setopt( $ch, CURLOPT_RETURNTRANSFER, true);

    $contenido = curl_exec($ch);
    curl_close($ch);    // Closing cURL

    //se guarda el contenido del sitio en la variable $html...
    $html = new simple_html_dom();
    $html->load($contenido);

    //Se extrae: Nombre de los comentarios
    $nombreComentario = $html->find('li.reviewername');

    if( $nombreComentario ){
        //Se extrae: El Pais,Genero,Edad
        $paisGeneroEdad = $html->find('li.reviewerdetails');
        //Se extrae: El comentario
        $comentario = $html->find('div.reviewtext');
        //Se extrae: las valoraciones por comentario
        $valoracionComentario = $html->find('ul.ratingbreakdown > li');
        //Se extrae: El nombre del hotel
        $nombreHotel = $html->find('h1.hwta-property-name');
        //Se extrae: La valoración general
        $valoracionGeneral = $html->find('div.textrating');
        //Se extrae: La fecha del comentario
        $fechaComentario = $html->find('span.reviewdate');
```

Figura B2.9 Extrayendo información del sitio.

Para esta parte del código, lo que realiza cada parte de este código es ir agrupando la información extraída anteriormente y guardándola en un array, cada uno de los comentarios de la página enviada, además se agrupa por los nombres representativos de cada extracción.

```

for ($i=0; $i < count($nombreComentario); $i++) {
    for ($j=$count; $j < $count+7; $j++) { //for x las valoraciones
        $valoracionComentarioCsv = $valoracionComentarioCsv.$valoracionComentario[$j]."##";
    }

    $count+=7;
    $auxiliar = explode(",",$paisGeneroEdad[$i]);
    if($auxiliar[0]) $pais = $auxiliar[0]; else $pais=" ";
    if($auxiliar[1]) $genero = $auxiliar[1]; else $genero=" ";
    if($auxiliar[2]) $edad = $auxiliar[2]; else $edad=" ";
    $lineaCsv = str_replace(" ", "", strip_tags($fechaComentario[$i]) )."##".
        str_replace(" ", "", strip_tags($nombreComentario[$i]) )."##".
        str_replace(" ", "", strip_tags($pais) )."##".
        str_replace(" ", "", strip_tags($genero) )."##".
        str_replace(" ", "", strip_tags($edad) )."##".
        strip_tags($nombreHotel[0])."##".
        strip_tags($valoracionGeneral[$i])."##".
        str_replace(" ", "", strip_tags($valoracionComentarioCsv) ).
        //str_replace(" ", "", strip_tags($comentario[$i]) )."##";
        preg_replace("/[^\w]*/", "", strip_tags($comentario[$i]) )."##";

    array_push( $lineaCsvArray,explode("##", $lineaCsv) );
    //array_push( $lineaCsvArray,$lineaCsv );
    $valoracionComentarioCsv="";
}
}
}
else{
    $GLOBALS["corte"] = 1;
}

return $lineaCsvArray;
}

```

Figura B2.10 Agrupando datos extraídos.

Por ultimo cuando se envía el array con todos los comentarios de la página extraído, es enviado a esta parte del código, en donde se van agrupando y rellenando el Excel. Una vez se haya completado toda la operación, antes de volver al index, se retorna tanto como un “1” como un “2” dependiendo si es un hostel o una ciudad para poder asignar correctamente el nombre de al Excel.

```

print_r($GLOBALS["corte"]);
print_r($lista_simple);

$doc = new PHPEXcel();
$doc->setActiveSheetIndex(0);
$doc->getActiveSheet()->fromArray($lista_simple, null, 'A1');

header('Content-Type: application/vnd.openxmlformats-officedocument.spreadsheetml.sheet');
header('Content-Disposition: attachment;filename="your_name.xls"');
header('Cache-Control: max-age=0');

// Do your stuff here
$writer = PHPEXcel_IOFactory::createWriter($doc, 'Excel5');
if($GLOBALS['div'] == 1 )
{
    $writer->save($nombre_ciudad.'.xls');
    return 1;
}
else
{
    $writer->save($nombre_hostal.'.xls');
    return 2;
}
}

```

Figura B2.10 Excel final.

Se iniciará a explicar la función de limpiar temporales. Como se observa en a figura B2.11, cuando se ha clickeado en el botón limpiar se da inicio a la función, en el cual solamente hace una limpieza veloz en otro script llamado “php/eliminar.php”, en este script solamente elimina todos los Excel residuales luego de alguna extracción, en cuanto ha tenido éxito se muestra un mensaje por pantalla principal indicando cual fue el resultado. En la figura B2.12 se puede observar cómo se realiza la función de limpieza.

```

//Eliminacion de archivos temporales
$( "#limpiar" ).click(function() {

    $.ajax({
        type: "POST",
        url: "php/eliminar.php",
        data: "{}",

        success: function(data){
            console.log(data);
            $("#cargando").hide();
            $('#Info').fadeIn(1000).html(data);
            $('#Info').fadeOut(1000);
        },
    }); //FIN AJAXx

});
// *****
```

Figura B2.11 Ajax limpieza de temporales.

```

<?php

if( !array_map('unlink', glob("*.xls"))){
    echo "Error, no existen archivos para eliminar";
}else{
    echo "Elimacion Exitosa!";
}

?>
```

Figura B2.12 Script limpieza de temporales.

Cuando el usuario a clickeado el botón de actualizar la base de datos, se da inicio a la función como se observa en la imagen B2.13. En el Ajax se envía la información a “php/rescatarHosteles.php”, una vez haya finalizado y haya enviado la información de vuelta, se muestra un mensaje por pantalla y se aborta la función del Ajax, y en caso de que se cancele bd.abort() realiza la parada de toda la función. En la figura B1.23 se explicará el script de cómo se actualiza la base de datos.

```

$( "#actualizarBd" ).click(function() {
    $("#cargandoBd").show();
    var bd = $.ajax({
        type: "POST",
        contentType: "application/json; charset=utf-8",
        url: "php/rescatarHosteles.php",
        data: "{}",
        dataType: "json",
        success: function (data) {
            $('#Info').fadeOut(6000).html("Actualizacion completada!");
            $("#cargandoBd").hide();
            bd.abort();
        },
        error: function (result) {
            console.log(result);
            $("#cargandoBd").hide();
        }
    });
    $( "#cancelax" ).click(function() {
        $("#cargandoBd").hide();
        bd.abort();
        location.reload(true);
    });
});
});

```

Figura B2.13 Ajax actualización BD.

En las primeras 2 consultas que se visualizan en la imagen, se eliminan todas las ciudades y hostales de todos los países, posterior a eso se seleccionan los países de latinoamerica para su nueva asignación de ciudad y hostales. Luego en el for se inicia la función para formar el link de la extracción y se llama con la etiqueta de la librería de html con el link formado y se asigna a \$html.

Luego en la figura B2.14, en la continuación de la función, se asigna en \$ciudadpais el nombre con la clase del sitio web, luego se recorre cada ciudad de ese país y se va insertando a la base de datos, una vez terminado esto se procede a buscar e ingresar los hostales de cada ciudad seleccionada, para este caso, se vuelve a formar el link que se asigna a \$hotelespais, además hay 2 if que es en caso de que fuera un caso especial, ya que no todos los link son genéricos, pueden ir variando dependiendo del país o ciudad. Una vez verificado bien todos los casos posibles, se asigna a \$hostelespaisnombre la clase del sitio, que sería el nombre del hostel. Se recorre cada uno de los hostales ingresándolos a la base de datos hasta llegar hasta el último encontrado.

Una vez finalizado, se da inicio nuevamente a la función para ir rellenando cada uno de los hostales de cada ciudad, ya que va rellenando una ciudad a ciudad, este proceso puede tomar tiempo. Una vez finalizada la actualización de la base de datos, se mostrará un mensaje por pantalla indicando que ha terminado con éxito.

```

$queryEliminarHostal = "DELETE FROM hostal;";
$conn->query($queryEliminarHostal);

$queryEliminarCiudad = "DELETE FROM ciudad;";
$conn->query($queryEliminarCiudad);

$query = "SELECT id_pais,nombreEsp
          FROM pais
          WHERE id_pais = 52
          OR id_pais = 55
          OR id_pais = 59
          OR id_pais = 53
          OR id_pais = 60
          OR id_pais = 58
          OR id_pais = 54
          OR id_pais = 56
          OR id_pais = 61
          OR id_pais = 57
          OR id_pais = 83
          OR id_pais = 73
          OR id_pais = 81
          OR id_pais = 72
          OR id_pais = 82
          OR id_pais = 87;";

foreach ($conn->query($query) as $row) {
    $link    = "http://www.spanish.hostelworld.com/alojamiento/" . $row[1];
    $pais    = $row[1];
    $id_pais = $row[0];

    // Create DOM from URL or file
    $html = file_get_html($link);

```

Figura B2.14 Parte 1 script base de datos.

```

if($html){
    $ciudadesPais = $html->find('.otherlocations ul li a');
    //echo "Pais: ".$pais." Ciudad: ".$ciudadesPais;
    //$hotelesPais = $html->find('span.propname a');

    // Recorrer cada ciudad
    foreach($ciudadesPais as $ciudad){
        $ciudadHotel = $ciudad->innertext;
        $ciudadHotel = mysql_real_escape_string(preg_replace('#[ -]+#', '-', $ciudadHotel));
        //Insertar todas las ciudades a la BD...

        $queryInsert = "INSERT INTO ciudad (id_pais,nombre)
                        VALUES ($id_pais,'$ciudadHotel');";

        $conn->query($queryInsert);
        $id_ciudad = $conn->lastInsertId();

        //recorrer hoteles por ciudad...
        //echo "http://www.spanish.hostelworld.com/alojamiento/" . $ciudadHotel . "/" . $pais . "<br>";
        $hotelesPais = file_get_html("http://www.spanish.hostelworld.com/Albergues/" . $ciudadHotel . "/" . $pais . "?ShowAll=1");

        if( $ciudadHotel == "Rio-de-Janeiro" )
            $hotelesPais = file_get_html("http://www.spanish.hostelworld.com/Albergues/Rio-de-Janeiro/Brasil");

        if( $ciudadHotel == "Buenos-Aires" )
            $hotelesPais = file_get_html("http://www.spanish.hostelworld.com/Albergues/Buenos-Aires/Argentina");

        if($hotelesPais){
            $hotelesPaisNombres = $hotelesPais->find('.resultheader h2 a');
            foreach( $hotelesPaisNombres as $hotel){
                //echo $hotel->innertext."<br>";
                $nombreHotel = $hotel->innertext;
                $nombreHotel = mysql_real_escape_string(preg_replace('#[ -]+#', '-', $nombreHotel));
                $link        = $hotel->href;
                $id_link     = end(explode('/', $link));
                $queryInsert = "INSERT INTO hostal (id_ciudad,nombre,id_link)
                                VALUES ($id_ciudad,'$nombreHotel',$id_link);";

                $conn->query($queryInsert);
            }
        }
    }
}

```

Figura B2.15 Parte 2 script base de datos.