

# Caracterización de contenido de sitios web turísticos mediante scraping y minería web para contribuir a la satisfacción de turista

## Characterization of content of tourism websites through web scraping and web mining to contribute to tourist satisfaction

VILLABONA, Natividad<sup>1</sup>

GARCÉS, David J.<sup>2</sup>

MARTELO, Raúl J.<sup>3</sup>

### Resumen

El objetivo fue caracterizar el contenido de sitios web de agencias turísticas a través de Scraping y Minería Web para contribuir a la satisfacción de turistas. La investigación fue cualitativa, porque no se realizaron mediciones numéricas y las variables utilizadas no necesitaron procesos estadísticos para su análisis. Como resultado se obtuvo que las técnicas Scraping y minería web permiten a las agencias turísticas reducir dificultades considerando aspectos técnicos y publicitarios, y contribuir a la satisfacción de la experiencia turística.

**Palabras clave:** agencia de turismo, sitios web turísticos, minería web, scraping

### Abstract

The objective was to characterize the content of tourist agency websites through Scraping and web mining to contribute to the satisfaction of tourists. The research was qualitative because no numerical measurements were made and the variables used did not require statistical processes for their analysis. As a result, it was obtained that Scraping and web mining techniques allow tourist agencies to reduce difficulties considering technical and advertising aspects, and contribute to the satisfaction of the tourist experience.

**key words:** tourism agency, tourism websites, web mining, scraping

---

## 1. Introducción

La revolución digital ha transformado la manera de presentar, distribuir, vender y comprar productos turísticos, y ha mejorado la forma de planear los viajes (Pinto & Castro, 2019). Las agencias turísticas se han unido a esta tendencia con el propósito de optar modelos de negocios basados en el entorno digital (Saura, Reyes-Menendez & Palos-Sanchez, 2020). Las agencias de turismo representan un agente intermedio entre servicios turísticos y turistas potenciales (Al-Maliki & Mohammed, 2019), las cuales hacen uso de publicidad para presentar sus servicios a un público específico y ser más competitiva en el mercado (Souza, Mendes-Filho & Buhalis, 2019). La

---

<sup>1</sup> Docente. Facultad de Ciencias Económicas. Universidad de Cartagena. nvillabonag@unicartagena.edu.co

<sup>2</sup> Estudiante. Facultad de Ingeniería. Universidad de Cartagena. dgarcesc1@unicartagena.edu.co

<sup>3</sup> Docente. Facultad de Ingeniería. Universidad de Cartagena. rmartelog1@unicartagena.edu.co

publicidad turística permite mejorar la promoción de la marca de un destino y promueve el desarrollo económico y social del mismo (Zhao, 2019).

Del mismo modo, existen varias formas de hacer publicidad turística online como redes sociales, blogs y anuncios online (Newman, Chang, Walters & Wills, 2016). Las herramientas anteriores pueden ser diseñadas para captar la atención de los usuarios e incrementar el atractivo de un destino (Moro, Ramos, Esmerado, & Jalali, 2019). Asimismo, los sitios web son de gran importancia para lograr este objetivo (Domínguez, González, & Darcy, 2018). Un sitio web es una agrupación de páginas web interconectadas que comparten un nombre de dominio, el cual puede ser creado y mantenido por un individuo, grupo, empresa u organización para una diversidad de propósitos (Jackson, 2019). Adicionalmente, un sitio web permite mejorar la interactividad entre empresas turísticas, empleados, profesionales y clientes (Adeola & Evans, 2019).

Por otro lado, los consumidores de servicios turísticos juegan un papel importante en la interacción con un sitio web, debido a que pueden ser: revisores, leer opiniones de otros usuarios; o colaboradores, escribir sus propias opiniones (Davis, & Agrawal, 2018). Por lo anterior, mantener un sitio web actualizado y optimizado es de vital importancia, debido a que permite incrementar la satisfacción de los clientes que utilizan el sitio (Odiko, Ogutu, Yabs & Omar, 2018). No obstante, los sitios web son propensos a la desactualización de la información, contenidos duplicados e irrelevantes (Qian, Law, & Wei, 2019), enlaces rotos, saturación visual, lentitud de carga y descarga (Chong & Law, 2019). Por tanto, existen varias técnicas que permiten conocer el estado de un sitio web, entre las que se encuentran el web scraping y la minería web.

Web Scraping es un proceso automatizado que permite la extracción de datos y su análisis para obtener información relevante de un objetivo particular (Indra & Dinesh, 2019). La técnica de web scraping se caracteriza por varios aspectos como comparar los precios de sitios online, monitoreo de datos, detección de cambios en un sitio web, investigación y la integración de datos web (Singrodia, Mitra & Paul, 2019). Por otra parte, la minería web es un proceso de extracción y descubrimiento de patrones, tendencias y reglas a partir de información no estructurada en Internet, como documentos de texto, archivos HTML, correos electrónicos y comentarios (Tang, Siu, Choy, Ho, Lam & Tsang, 2019; Verma & Kumar, 2019). La minería web permite identificar información relevante presente en un sitio web (Kaur & Garg, (2019).

Por lo anterior, la implementación de las técnicas web scraping y minería web en sitios web turísticos, es fundamental para retener y fidelizar a los turistas (Al-Abdullatif & Kotb, 2014; Damodharan & Ravichandran, 2019). Por esta razón, el objetivo de esta investigación radica en caracterizar el contenido sitios web de agencias turísticas a través de Scraping y minería web para contribuir a la satisfacción de turistas. Esto se realiza con el fin de que las agencias de turismo conozcan el potencial de las técnicas web scraping y minería web para autoevaluar aspectos técnicos y publicitarios de un sitio web, lo cual permite reducir falencias y contribuir a la satisfacción del turista en la búsqueda de información, precisión de resultados y facilidad de uso.

---

## **2. Marco teórico**

### **2.1. Agencia turística**

Una agencia turística es una empresa asociada al turismo e interesada en la organización y elaboración de productos turísticos (Frenzel, 2017), con el propósito de satisfacer las necesidades de personas interesadas en vivir nuevas experiencias fuera de su lugar de origen (Singh & Ranjan, 2019) y conocer nuevas culturas (Lin, Nguyen & Yeng, 2019). Las agencias turísticas han centrado sus estrategias en mejorar la calidad de servicio, con el objetivo de obtener una ventaja competitiva en el mercado, lo cual permite mayor satisfacción del consumidor, mayores ganancias y fidelización de clientes (Kuo, Chang, Cheng & Lai, 2013).

## 2.2. Sitio web

Un sitio web es un conjunto de archivos y páginas web electrónicos relacionadas con tema específico, el cual integra una página de bienvenida con un nombre de dominio y dirección en Internet (Zhang, Chau, Xu & Liu, 2019). De igual manera, el objetivo principal de un sitio web es entregar mensajes fácilmente a los usuarios y mantener el registro de todas las ofertas, avisos y otros informes relacionados con una actividad particular (Bhowmik, Roy, Biswas, Roy, & Roy, 2019). Por otro lado, el diseño de un sitio web debe ser fácil de usar, donde los usuarios puedan entender el contenido de maneja rápida, por tanto, en la construcción se debe tener en cuenta la perspectiva de los clientes (Yang & Ha, 2019).

## 2.3. Dominio web

Un dominio web es un nombre único que determina un sitio web en Internet, donde un navegador utiliza un dominio para dirigirse a un sitio web (Estira, Espinosa, Guzman, & Orlanda-Ventayen, 2018). Asimismo, un dominio se forma de dos partes principales: nombre del sitio web y extensión de dominio (Nuñez, Moquillaza & Paz, 2019). Del mismo modo, un dominio web se usa para proporcionar interfaz y componentes funcionales de una aplicación web (Kuchhal, 2018).

## 2.4. Minería web

La minería web es una metodología que utiliza la minería de datos para recuperar y extraer información de la estructura, del contenido y las relaciones de un sitio web y de los registros de navegación de los usuarios (Rita, Rita & Oliveira, 2018). Este proceso se realiza a través del desarrollo de algoritmos y métodos especializados que permiten obtener información relevante de los sitios web y efectuar comercio electrónico con el propósito de marketing personalizado (Damodharan & Ravichandran, 2019).

## 2.5. Web Scraping

Web scraping es un procedimiento de rastreo automático de un sitio web con el fin de recopilar datos a través de un script implementado para realizar esta actividad (Canito, Ramos, Moro & Rita, 2018). El Web Scraping proporciona varias ventajas con respecto a las encuestas, debido a que los usuarios realizan comentarios voluntariamente en los sitios web, información que puede ser recuperada y analizada con un propósito determinado (Schuckert, Liu & Law, 2016; Darmawiguna, Pradnyana & Santyadiputra, 2019). Por lo anterior, el volumen de información que se puede obtener es grande (Amalia, & Angriani, 2019).

---

## 3. Metodología

Esta investigación fue cualitativa con diseño no experimental, transeccional, descriptivo. Cualitativa, porque no se manipuló la información consultada ni se procesaron datos medibles (Patten & Newhart, 2018). No experimental, debido a que las variables no fueron manipuladas deliberadamente y se observó el fenómeno en su ambiente natural para analizarlo ((Hernández, Fernández, & Baptista, 2014). En este caso, se analizaron los sitios web turísticos. Transeccional, porque se observaron las variables en un solo punto en el tiempo y su interacción en ese momento (Ávila, 2006). Finalmente, descriptivo porque detalla el estado, características, factores y procedimientos de las variables de estudio, tal como se presentan naturalmente (Bermúdez y Rodríguez, 2013).

De acuerdo a lo anterior, se identificó la variable Contenido de un sitio web turístico, con su respectivas dimensiones e indicadores, como se observa en la Tabla 1.

**Tabla 1**  
Variable, dimensiones e indicadores

Variable	Dimensión	Indicador
Contenido de un sitio web turístico	Actualización	Fecha última actualización
		Contenidos duplicados
	Adaptabilidad al usuario	Idiomas
		Tipos de moneda
	Variedad de presentación	Tipos de formatos
		Uso correcto de formatos
	Precisión	Errores ortográficos
		Palabras claves
		Promedio palabras por página
	Autoridad	Verificabilidad de datos de la agencia
		Uso de Redes sociales
		Políticas de tratamiento de datos

Fuente: Autores

### 3.1. Población y muestra

La Asociación Colombiana de Agencias de Viajes y Turismo (ANATO), proporciona un registro de 77 agencias de viajes certificadas en Colombia y aliadas a ANATO, de las cuales 80,5% cuenta con un sitio web y con Registro Nacional de Turismo (RNT) activo. De estas agencias, se eligieron 5 por muestreo de conveniencia, debido a que se seleccionaron deliberadamente (Beins & McCarthy, 2012), donde se tuvo en cuenta 3 aspectos que debían cumplir los sitios web turísticos: primero, ofrecer paquetes o planes turísticos de al menos cuatro tipos de turismo distintos (etnográfico, ecológico, deportivo, aventura, etc.); segundo, ofrecer paquetes o planes turísticos que abarquen un servicio completo de experiencia turística (vuelo, hotel, visitas, etc.); tercero, tener presencia nacional e internacional.

### 3.2. Técnicas de recolección de información

Como técnica de recolección de información se aplicó la revisión documental, la cual facilita la recuperación, análisis e interpretación de datos de otros autores (Chacón, Herrera & Villabona, 2013). Por tanto, se consultaron bases de datos indexadas, artículos científicos, repositorios universitarios y sitios web. Además, se utilizaron herramientas informáticas que permitieron obtener información de los sitios web turísticos colombianos con un propósito específico, las cuales se muestran en la Tabla 2.

**Tabla 2**  
Herramientas informáticas utilizadas

Nombre	Propósito	Enlace
Think with Google	Prueba de velocidad	<a href="https://www.thinkwithgoogle.com/">https://www.thinkwithgoogle.com/</a>
Woorank	Análisis de tráfico, puesto, mapa de sitios	<a href="https://www.woorank.com/">https://www.woorank.com/</a>
Nibbler silktide	Análisis SEO, datos de tiempo, tráfico, palabras claves, enlaces, errores gramaticales	<a href="https://nibbler.silktide.com/">https://nibbler.silktide.com/</a>
Alexa	Análisis de palabras claves y sitios web relacionados	<a href="https://www.alexa.com/siteinfo">https://www.alexa.com/siteinfo</a>
Siteliner	Contenido duplicado	<a href="http://siteliner.com">siteliner.com</a>
Httpstatus	Estado de respuesta de las páginas internas	<a href="https://httpstatus.io/">https://httpstatus.io/</a>
Web Scraper	Extractor de información	<a href="https://webscraper.io/">https://webscraper.io/</a>
Cyotek	Minería de archivos	<a href="http://cyotek.com/cyotek-webcopy">cyotek.com/cyotek-webcopy</a>

Fuente: Autores

### 3.3. Análisis de los datos

El análisis de datos estuvo basado en la caracterización de contenido de sitios web turísticos colombianos, por lo cual se aplicaron las técnicas análisis documental, web scraping y minería web. La primera, permitió conocer y detallar los elementos y significados de documentos relacionados con sitios web turísticos. La segunda, permitió utilizar programas informáticos para extraer información de sitios web. La última, permitió el descubrimiento de patrones a partir de la estructura, contenido y registros de sitios web turísticos.

## 4. Resultados

Los hallazgos de esta investigación están basados en el contenido de sitios web de agencias turísticas colombianas. Por tanto, los resultados obtenidos fueron analizados de un punto de vista general, con respecto a las dimensiones e indicadores, como se muestra en la Tabla 3.

**Tabla 3**  
Análisis de la variable Contenido de un sitio web turístico

Dimensión	Indicador	Sitio 1	Sitio 2	Sitio 3	Sitio 4	Sitio 5
Actualización	Fecha última actualización (días)	5	49	20	17	42
	Contenidos duplicados	3%	26%	25%	29%	29%
Adaptabilidad al usuario	Idiomas	1	1	1	1	2
	Tipos de monedas	14	1	1	2	1
Variedad de presentación	Tipos de formatos	4	4	3	4	3
	Uso correcto de formatos	No	Si	Si	Si	No
Precisión	Errores ortográficos	4	6	3	1	1
	Palabras claves	5	6	7	7	6
	Promedio palabras por página	1104	157	680	1110	448
Autoridad	Verificabilidad de datos de la agencia	Sí	Sí	Sí	Sí	Sí
	Uso de Redes sociales	Sí	Sí	Sí	Sí	Sí
	Políticas de tratamiento de datos	Sí	Sí	Sí	Sí	Sí

Fuente: Autores

#### 4.1. Dimensión: Actualización

Se refleja que el contenido de sitios web turísticos colombianos se modificaron, en promedio, 26 días antes de la aplicación de la técnica minería web y el porcentaje de contenido duplicado está alrededor del 22%. Según Andrade y Sobata (2020), plantean que el contenido de un sitio web turístico debe mantenerse en actualización constante con el propósito de mantener al turista informado de los productos que ofrece la agencia, donde el sitio web con información actualiza puede aumentar la sensación de que el turista potencial adquiera un producto turístico. Asimismo, Benevolo y Spinelli (2019), plantean que las agencias deben mantener sitios web actualizados con respecto a tarifas, eventos, espectáculos y fiestas con el fin de aumentar el flujo de turistas.

#### 4.2. Dimensión: Adaptabilidad al usuario

Se evidencia que los sitios web turísticos presentan información a los usuarios en el idioma español. Además, uno (1) de cada cinco (5) sitios web presenta contenido a los usuarios en dos (2) idiomas: español e inglés. Asimismo, el 20% de la muestra usa herramientas de traducción de texto plano del contenido web, lo que permite tener acceso a decenas de idiomas. Adicionalmente, el 100% de los sitios utilizan el peso como moneda principal y solo el 20% utiliza otros tipos de monedas como dólar, peso mexicano, euro, entre otros. Según Rodas-Silva, Galindo, García-Gutiérrez y Benavides, (2019), plantea que los sitios web deben tener acceso a un

complemento de traducción con el propósito de ofrecer contenido adaptable al idioma que conoce el turista, lo cual facilita la adquisición del producto turístico.

#### **4.3. Dimensión: Variedad de presentación**

Se refleja que las agencias turísticas se apoyan en sitios web para presentar contenido a los usuarios a través de texto plano, imágenes, videos y en formato de documentos portátil (PDF), donde las imágenes son las que más se utilizan. No obstante, solo el 60% de los sitios hace uso correcto de los tipos de formatos para mostrar información relevante a los usuarios. De acuerdo con Herrero, San Martín y Hernández (2015), las agencias de turismo deben tener presencia activa en los sitios web y en constante revisión para respaldar su credibilidad a través de la publicación de contenido relevante a los turistas en diferentes formatos.

#### **4.4. Dimensión: Precisión**

Se muestra que los sitios web turísticos colombianos, presentan en promedio tres (3) errores ortográficos por página, entre los que se encuentran: el uso inadecuado de tildes, confusiones entre s, c y z, entre otros. De acuerdo con Cooper, Diab y Beeson (2019), la presencia de errores ortográficos en sitios web, afecta la calidad de la agencia y la percepción del turista, lo cual disminuye las opciones de compras de paquetes turísticos. Por otro lado, en promedio se utilizan 6 palabras claves por sitios web, donde las palabras más comunes son: vuelos baratos, hoteles, itinerario, viajes, reserva, destinos, entre otras. No obstante, algunos sitios web utilizan palabras para ser hallados por los buscadores que no tienen relación con lo que ofrecen, palabras como política, privacidad, generales, entre otros. Del mismo modo, se identificó que cada sitio web presenta en promedio 700 palabras por página.

#### **4.5. Dimensión: Autoridad**

Se evidencia que los sitios web enfocados al turismo, tienen 100% disponible el código Registro Nacional de Turismo (RNT), sellos de certificación e información legal disponible en página principal. Del mismo modo, estos sitios hacen uso de redes sociales como Facebook, Youtube y Twitter. Además, el 60% de ellos usan Instagram; el 40% usa LinkedIn y Pinterest; y el 20% usa TripAdvisor. No obstante, el 40% de los sitios web turísticos colombianos usan la red social Google Plus, la cual fue cerrada por los propietarios en abril de 2019. Según Guerreiro, Viegas y Guerreiro (2019), estudios han demostrado el impacto de las redes sociales en el turismo, donde la mayoría de turistas usan redes sociales cuando planifican su viaje, buscan información de destinos y comparten sus experiencias, lo cual permite a las agencias de turismo tener presencia en redes sociales con el propósito de compartir experiencias de sus clientes y mostrar a los nuevos posibles clientes contenido atractivo.

#### **4.6. Resumen de las dimensiones**

Con base en lo anterior, se refleja que el contenido de sitios web turísticos colombianos, presenta información desactualizada, poco relevante y con excesivo contenido duplicado. Según Rinka y Pratt (2018), el contenido publicado en sitios web turístico debe ser potencialmente atractivo e influyente con el propósito de fidelizar y retener a los turistas, lo cual permite incrementar la satisfacción de turistas que usan el sitio web. Del mismo modo, la mayoría de las agencias disponen sus sitios en un solo idioma y un tipo de moneda para ofrecer productos a los turistas. Las agencias utilizan material didáctico como imágenes, videos y PDF, en sus sitios para mostrar información relevante a turistas, donde solo el 60% de las agencias hace uso correcto de los formatos. Los sitios presentan varios errores ortográficos por página y usan palabras claves que no tienen relación con el turismo. Las agencias presentan fortaleza con respecto a la verificación del código RNT y el uso de redes sociales.

---

## 5. Conclusiones

Con base en los resultados se concluye que: las agencias turísticas colombianas presentan falencias con respecto a la actualización de información, contenidos duplicados y errores ortográficos en sus sitios web, así como poca disponibilidad de idiomas y monedas para atraer turistas internacionales. Del mismo modo, el contenido de sitios web turísticos colombianos presenta falencias para contribuir a la satisfacción de la experiencia del turista en la búsqueda de información, obtención de respuestas, precisión de resultados y facilidad de uso. Adicionalmente, las técnicas de Web Scraping y minería web son fundamentales para hacer marketing digital en sitios web, debido a que permiten a las agencias turísticas autoevaluar aspectos técnicos, publicitarios, optimización de SEO, corrección de enlaces caídos o rotos y eliminación de contenidos duplicados de un sitio web. Las agencias presentan fortalezas con respecto a la verificación del código RNT, políticas de tratamiento de datos y al uso de redes sociales.

---

## Referencias bibliográficas

- Adeola, O., & Evans, O. (2019). Digital tourism: mobile phones, internet and tourism in Africa. *Tourism Recreation Research*, 44(2), 190-202.
- Al-Abdullatif, W., & Kotb, Y. (2014). Using online hotel customer reviews to improve the booking process. *International Journal of Computer Applications*, 97(16).
- Al-Maliki, S., & Mohammed, A. (2019). The investment opportunities for building smartphone applications for tourist cities in Saudi Arabia. *Expert Systems in Finance: Smart Financial Applications in Big Data Environments*, 112.
- Amalia, A., & Angriani, R. (2019). North Sumatera tourism spots searching system based on semantic web. In *Journal of Physics: Conference Series* (Vol. 1235, No. 1, p. 012099). IOP Publishing.
- Andrade, J., & Sobata, M. F. (2020). Most Important Contents in Travel Blogs: A Perspective from Brazilian Tourists. In *Advances in Tourism, Technology and Smart Systems* (pp. 77-91). Springer, Singapore.
- Ávila Baray, H. L. (2006). *Introducción a la metodología de la investigación*. Edición electrónica.
- Beins, B., & McCarthy, M. (2012). *Research Methods and Statistics*. New Jersey: Pearson.
- Benevolo, C. & Spinelli, R. (2019). The use of websites by Mediterranean tourist ports. *Journal of Hospitality and Tourism Technology*, 10(2), 190 - 204.
- Bermúdez, L. T., & Rodríguez, L. F. (2013). *Investigación en la gestión empresarial*. Ecoe Ediciones
- Bhowmik, D., Roy, M., Biswas, D., Roy, S., & Roy, S. (2019). Converting and Developing Live Web Site into a Web Content Management System. In *Advances in Communication, Cloud, and Big Data*, 79 - 87.
- Canito, J., Ramos, P., Moro, S., & Rita, P. (2018). Unfolding the relations between companies and technologies under the Big Data umbrella. *Computers in Industry*, 99, 1 - 8.
- Chacón, J. W. B., Herrera, J. C. B., & Villabona, M. R. (2013). Revisión y análisis documental para estado del arte: una propuesta metodológica desde el contexto de la sistematización de experiencias educativas. *Investigación Bibliotecológica: archivonomía, bibliotecología e información*, 27(61), 83 - 105.
- Chong, S., & Law, R. (2019). Review of studies on airline website evaluation. *Journal of Travel & Tourism Marketing*, 36(1), 60-75.

- Cooper, A., Diab, D., & Beeson, K. (2019). Why Spelling Errors Matter: Online Company Reviews and Organizational Attraction. *Corporate Reputation Review*, 1-10.
- Damodharan, P., & Ravichandran, C. (2019). Applicability Evaluation of Web Mining in Healthcare E-Commerce towards Business Success and a derived Cournot Model. *Journal of medical systems*, 43(8), 1 - 10.
- Darmawiguna, I., Pradnyana, G., & Santyadiputra, G. (2019). The Development of Integrated Bali Tourism Information Portal using Web Scrapping and Clustering Methods. In *Journal of Physics: Conference Series*, 1165(1), 1 - 10.
- Davis, J. M., & Agrawal, D. (2018). Understanding the role of interpersonal identification in online review evaluation: An information processing perspective. *International Journal of Information Management*, 38(1), 140-149.
- Domínguez, T., González, E., & Darcy, S. (2018). Website accessibility in the tourism industry: an analysis of official national tourism organization websites around the world. *Disability and rehabilitation*, 40(24), 2895 - 2906.
- Estira, K., Espinosa, N., Guzman, M., & Orlanda-Ventayen, C. (2018). Tourism and Product Advertising Practices of Top Tourist Destination in one Philippine Province. *Southeast Asian Journal of Science and Technology*, 3(1), 1 - 5.
- Frenzel, F. (2017). Tourist agency as valorisation Making Dharavi into a tourist attraction. *Annals of Tourism Research*, 66, 159 - 169.
- Guerreiro, C., Viegas, M., & Guerreiro, M. (2019). Social Networks and Digital Influencers: Their Role in Customer Decision Journey in Tourism. *Journal of Spatial and Organizational Dynamics*, 7(3), 240 - 260.
- Hernández, R., Fernández, C., & Baptista, P. (2014). *Metodología de la investigación*. México: Mc Graw Hill Education.
- Herrero, Á., San Martín, H. & Hernández, J. (2015). How online search behavior is influenced by user-generated content on review websites and hotel interactive websites. *International Journal of Contemporary Hospitality Management*, 27(7), 1573 - 1597.
- Indra, E., & Dinesh, T. (2019). Designing Android Gaming News & Information Application Using Java-Based Web Scrapping Technique. In *Journal of Physics: Conference Series*, 1230(1). IOP Publishing.
- Jackson, A. T. (2019). Remembering Jim Crow, again—critical representations of African American experiences of travel and leisure at US National Park Sites. *International Journal of Heritage Studies*, 25(7), 671-688.
- Kaur, J., & Garg, K. (2019). Efficient Management of Web Data by Applying Web Mining Pre-processing Methodologies. In *Software Engineering* (pp. 115-122). Springer, Singapore.
- Kuchhal, M. (2018). Web applications having end-user data isolated in the local domain. United States Patent Application.
- Kuo, N., Chang, K., Cheng, Y., & Lai, C. (2013). How service quality affects customer loyalty in the travel agency: The effects of customer satisfaction, service recovery, and perceived value. *Asia Pacific Journal of Tourism Research*, 18(7), 803-822.
- Lin, Y., Nguyen, T., & Yeng, L. (2019). Reliability evaluation of a multi-state air transportation network meeting multiple travel demands. *Annals of Operations Research*, 277(1), 63 -82.



- Moro, S., Ramos, P., Esmerado, J., & Jalali, S. (2019). Can we trace back hotel online reviews' characteristics using gamification features? *International Journal of Information Management*, 44, 88 - 95.
- Newman, R., Chang, V., Walters, R. J., & Wills, G. B. (2016). Web 2.0—The past and the future. *International Journal of Information Management*, 36(4), 591 - 598.
- Nuñez, A., Moquillaza, A., & Paz, F. (2019). Web Accessibility Evaluation Methods: A Systematic Review. In *International Conference on Human-Computer Interaction*, 226 - 237.
- Odiko, S., Ogutu, M., Yabs, J., & Omar, M. (2018). International Marketing Strategy, Information Communication Technology and Performance of Tour Firms in Kenya. *Journal of Information and Technology*, 2(1), 39 - 64.
- Patten, M., & Newhart, M. (2018). *Understanding research methods: An Overview of the Essentials*, Tenth Edition. New York: Taylor & Francis.
- Pinto, I., & Castro, C. (2019). Online travel agencies: factors influencing tourist purchase decision. *Tourism & Management Studies*, 15(2), 7 - 20.
- Qian, J., Law, R., & Wei, J. (2019). Knowledge mapping in travel website studies: A scientometric review. *Scandinavian Journal of Hospitality and Tourism*, 19(2), 192 - 209.
- Rinka, X. & Pratt, S. (2018). Social media influencers as endorsers to promote travel destinations: an application of self-congruence theory to the Chinese Generation Y. *Journal of Travel & Tourism Marketing*, 35(7), 958 - 972.
- Rita, P., Rita, N., & Oliveira, C. (2018). Data science for hospitality and tourism. *Worldwide Hospitality and Tourism Themes*, 10(6), 717 - 725.
- Rodas-Silva, J., Galindo, J. A., García-Gutiérrez, J., & Benavides, D. (2019). Selection of software product line implementation components using recommender systems: An application to Wordpress. *IEEE Access*, 7, 69226-69245.
- Saura, J., Reyes-Menendez, A., & Palos-Sanchez, P. (2020). The Digital Tourism Business: A Systematic Review of Essential Digital Marketing Strategies and Trends. In *Digital Marketing Strategies for Tourism, Hospitality, and Airline Industries*, 1 - 22.
- Schuckert, M., Liu, X., & Law, R. (2016). Stars, votes, and badges: How online badges affect hotel reviewers. *Journal of Travel & Tourism Marketing*, 33(4), 440 - 452.
- Singh, S., & Ranjan, R. (2019). Online Travel Portal and their effect on Travel Agency: A Study on outbound visitors of Varanasi. *International Journal of Research and Analytical Reviews (IJRAR)*, 6(2), 387 - 393.
- Singrodia, V., Mitra, A., & Paul, S. (2019). A Review on Web Scrapping and its Applications. In *2019 International Conference on Computer Communication and Informatics (ICCCI)*, 1 - 6. IEEE.
- Souza, J., Mendes-Filho, L., & Buhalis, D. (2019). Evaluating the effectiveness of tourist advertising to improve the competitiveness of destinations. *E-Tourism Economics*, 1 - 20.
- Tang, V., Siu, P., Choy, K., Ho, G., Lam, H., & Tsang, Y. (2019). A web mining-based case adaptation model for quality assurance of pharmaceutical warehouses. *International Journal of Logistics Research and Applications*, 22(4), 325 - 348.

- Verma, K., & Kumar, A. (2019). An Efficient Way of Handling Large Scale Web News Data using Data Mining Techniques. Proceedings of 2nd International Conference on Advanced Computing and Software Engineering (ICACSE) 2019.
- Yang, Y., & Ha, J. (2019). Research on User Mental Model Based on Website Classification System. In 2nd International Conference on Social Science, Public Health and Education (SSPHE 2018).
- Zhang, Y., Chau, K. T., Xu, J., & Liu, C. (2019). An investigation into the attractiveness level of commercial website for visitors: a case study on 66rpg website. In Proceedings of the 3rd International Conference on Cryptography, Security and Privacy (pp. 99-102). ACM.
- Zhao, C. (2019). Research on the "Implanted" Model of Tourism Advertising in City Brand Promotion under the Strategy of Guangdong-Hong Kong-Macao Greater Bay Area. In 2018 International Symposium on Social Science and Management Innovation (SSMI 2018). Advances in Economics, Business and Management Research, 68, 687 - 689.