

1. EL USO TECNOLÓGICO DESDE LAS PÁGINAS WEB 2.0 EN EL SECTOR HOTELERO DE PARANÁ (PROVINCIA DE ENTRE RÍOS)

Graciela Mingo¹
y Mariano Testaseca²

Fecha de recepción: 07/01/2020
Fecha de aceptación: 26/04/2021



Introducción

Aquí se sistematiza parte de la información obtenida en un proyecto de investigación titulado "Estrategias comerciales de los hoteles de la ciudad de Paraná a partir de la WEB 2.0 y la aparición de las Agencias de Viajes Virtuales (OTA – *On Line Travel Agency*)". En particular, se trata de la información recabada para un hotel de cinco estrellas a través de dos de las páginas Web 2.0 que son: *Booking* y *TripAdvisor*. La técnica en la interpretación de lo extraído de dichas páginas consistió en el análisis de contenido automatizado tanto de la interactividad lograda como de las opiniones que los clientes sobre dicho establecimiento hotelero al haber pernoctado en él. Se analizan a su vez las respuestas dadas desde el hotel sobre las opiniones vertidas por los clientes.

Hemos partido de entender que estudios sobre las nuevas tecnologías -y en especial la minería de datos- ocupan un lugar importante para muchas disciplinas, tanto en el desarrollo en la gestión de

¹ Universidad Autónoma de Entre Ríos. Dirección de contacto: gmbevilac@gmail.com

² Universidad Nacional del Litoral; Dirección de contacto: mtestasecca@gmail.com

Las organizaciones, de las técnicas implementadas y en la sistematización de los datos recabados a través de las páginas digitales, como son dos sitios virtuales. El reto para el equipo de investigación (cuya composición es interdisciplinaria) se presentó como una opción de aprendizaje continuo que se traslució en las decisiones que se fueron tomando en el desarrollo de la propuesta. Por ello el lector encontrará primero ideas básicas en referencia a las técnicas utilizadas al describir el pase de una metodología tradicional al uso de una densidad de datos posible de lograr gracias al apoyo tecnológico, y luego por cada página Web 2.0 se presentan diferentes procesamientos que muestran las opiniones vertidas por los clientes, la interactividad lograda y lo que se destaca cómo más significativo en los servicios que pone a disposición de los clientes el hotel. El período de análisis fue del 1 de octubre de 2017 al 30 de setiembre de 2018.

Del análisis de contenido tradicional al análisis de contenido automatizado

La técnica que se ha empleado para esta indagación ha sido el análisis de contenido automatizado. Es una técnica de análisis que posibilita el ejercicio analítico documental, cuyo objetivo es obtener información de cierto conjunto de datos, generalmente surgidos de documentos, prensa escrita, transcripciones de entrevistas, textos o grabaciones (Moreiro González, 2002; Krippendorff, 1990), y en este caso de páginas web.

La técnica en su modo tradicional, tuvo su origen con rigor científico en EE.UU., fundamentalmente en los departamentos de ciencias políticas, jugando un papel muy importante en el desarrollo ante los problemas puestos de manifiesto en la Segunda Guerra Mundial. Con el transcurrir del tiempo el análisis de contenidos se ha servido de técnicas específicas que mejoraron su alcance, aplicándose en marcos metodológicos de investigación cuantitativos, cualitativos y mixtos, mientras se "emplea un amplio rango de técnicas analíticas para generar descubrimientos y ponerlos en contexto" (Domas White y Marsh, 2006: 22).

Con el avance de la sociedad de la información, el interés de los científicos sociales por aplicar técnicas de análisis de contenido automatizadas (minería de textos) para dar respuestas a preguntas clásicas de investigación tales como: "quién habla, a quién lo dice, por qué lo dice, cómo y con qué efecto" ha ido creciendo.

Específicamente al llevar a cabo la cuantificación se trata de registrar las frecuencias del contenido manifiesto (en documentos, noticias, entrevistas) configurando el *corpus* y luego realizando la interpretación, escudriñando el contenido latente (Mingo, 2009) que extrae de cada uno de ellos. Este interés subyace aunque aumenta la cantidad de datos y aparezcan nuevas formas de explorar los contenidos en diarios, revistas, publicaciones *online* y bases de datos.

Al referirse al análisis de contenido clásico, Cea D'Ancona (1996) menciona las características de la objetividad, sistematicidad y cuantificación de los contenidos manifiestos de la comunicación, y agrega que su uso permite las inferencias válidas y replicables que se pueden aplicar a un contexto. Ya en el análisis de contenido tradicional, la contabilización de la cantidad de palabras, o la frecuencia de su aparición en los textos, es la forma habitual de presentar la información cuantificables: generalmente, habrá de limitarse a aspectos formales y al contenido manifiesto (referidos a la extensión dedicada a un tema, tapa, peso, tamaño, etc.) (Mayntz et al., 1993). Este planteo sigue vigente de acuerdo al tamaño de los contenidos observados; y si son documentos, discursos o pequeños textos, el giro se da a partir de la conformación de grandes *corpus* de datos con la aparición del Big Data. El giro no solo es mecánico; lo es epistémico, otorgando una nueva relevancia y llevando a una nueva forma de pensar, escudriñar e interpretar la información.

Con el avance de las tecnologías las formas de lograr sistematizar datos han madurado, al conformarse una nueva disciplina denominada Humanidades Digitales, es decir, la aplicación de novedosos algoritmos y métodos computacionales para la conformación de instrumentos de observación y luego de interpretación.

Como en los instrumentos de observación, se requiere de un "input" de datos, cuyo *corpus* suministra la información que el investigador va a analizar. Desde la lógica del procesamiento y la recolección de las fuentes de información -en cantidades abundantes de datos textuales (estructurados, semi-estructurados o sin estructura)- se procede por medio de procedimientos que van desde la recolección manual (texto por texto) y su digitalización, hasta los más automatizados (como *web scrapping*), para una posterior sistematización en planillas o base de datos (relacional, no relacional u orientada a grafos).

Aspectos metodológicos específicos

De este modo, los instrumentos de observación -un conjunto seleccionado de métodos computacionales- son aplicados sobre el *corpus* (en este caso las opiniones de los clientes y las respuestas desde los hoteles en las páginas web 2.0). Esto posibilitó la generación del *corpus* de la investigación, luego de lo cual se puso en práctica el análisis de contenido (automatizado), obteniéndose así ciertos patrones e información relevante y novedosa para la investigación en curso. Con más precisión aclaramos que este *corpus* está conformado por los comentarios ingresados en la web 2.0, por los huéspedes de cada uno de los hoteles seleccionados de la ciudad de Paraná, y las respuestas dadas por el personal de los hoteles, información extraída de los sitios *TripAdvisor* y *Booking* durante el período: 1 de octubre de 2017 al 30 de septiembre de 2018.

Con referencia a las páginas web 2.0 seleccionadas, es interesante acotar que *Booking* es una plataforma que ofrece los hoteles a través de su red y que a muchos hoteleros le resulta una garantía para tener sus plazas ocupadas. *TripAdvisor* nació como un sitio de opiniones sobre los hoteles y sus valores en los mercados, luego creció añadiendo un comparador de precios compitiendo con *Trivago* y continuó a través de las OTAs negociando con los hoteles (Llibre, 2015). Por ende, nuestras unidades de observación son las Páginas web 2.0 de los sitios *TripAdvisor* y *Booking* y las unidades de análisis son los hoteles seleccionados acorde a los comentarios que plasman las opiniones de los huéspedes que han pernoctado en ellos y las respuestas recibidas.

El análisis por la cantidad de información encontrada de los comentarios sobre los hoteles es facilitado por la aplicación de técnicas de minería de datos o textos, técnicas que se vienen desarrollando en paralelo al surgimiento de las redes sociales. Los pasos que configuran el proceso específico han viabilizado el proceso y permitido algunos patrones sobre las opiniones de los usuarios o clientes que en un ambiente digital compartido expresan. Se produce así conocimiento que antes no era factible encontrar, debido a la variedad y complejidad de la información que se permite a través del medio digital en donde queda plasmada la misma y que hoy es factible procesar (Mendoza, 2011).

En la sistematización de la información de las opiniones vertidas por los clientes respecto de sus estadías en los hoteles de la ciudad de Paraná, se trabajó con una planilla de cálculo en Excel. En la

misma se incorporaron los comentarios u opiniones de los clientes que pernoctaron en cada uno de los hoteles seleccionados de la ciudad de Paraná. Una vez realizada la recolección y obtenida nuestra población o corpus, construimos un instrumento de observación, para el cual, se realizó una selección de algoritmos a emplear. Entendemos a los algoritmos, como un grupo finito de instrucciones o reglas establecidas y organizadas de manera lógica y ordenada linealmente que permiten solucionar un determinado problema. De los algoritmos existentes de minería de textos, hemos optado por los siguientes para su presentación, y dar así argumentos válidos a los objetivos que nos trazamos con la investigación a través del análisis de:

- 1) *Gráficos de líneas*: contabilizando los términos, permiten graficar la frecuencia de repetición de palabras.
- 2) *Nubes de palabras*: permiten graficar y visualizar la frecuencia de repetición de las palabras de un modo más "preciso por su distribución y tamaño en la nube, facilitando una rápida visualización de lo que surge como más significativo las expresiones vertidas por los huéspedes. Es una representación visual de las palabras que conforman un texto, en donde el tamaño es mayor para las palabras que aparecen con más frecuencia" (Marketing Digital, 2017).
- 3) *Palabras clave por método RAKE*: identifica las palabras y frases clave de cada texto. El algoritmo RAKE es de tipo estadístico. Tiene la finalidad de identificar y extraer palabras clave multipalabra de documentos digitales.
- 4) *Coocurrencias*: con graduaciones en color, permite enfatizar la utilización conjunta de términos según las expresiones dadas por los huéspedes en cada uno de los hoteles seleccionados.

Cuando se trabajó las nubes de palabras para el presente estudio, no habían sido unificados los términos plurales y los singulares. Es decir, encontraremos los términos "habitación" y "habitaciones" tal cual han sido empleados por los usuarios. Para poder unificar ambos en un único término, debimos con antelación haberse empleado un analizador sintáctico (*tagging*) sobre cada término para diferenciar términos singulares de plurales y luego aplicar la técnica de *stemming*, para

obtener las raíces de las palabras plurales y así unificarlos. Esto no se realizó porque lo describimos al momento de la interpretación de las palabras, si se pudieron re-trabajar los términos que estaban con acento o sin ellos.

Otra acción a considerar en el proceso previo a emitir los cuatro gráficos obtenidos por la aplicación de algoritmos de minería de textos es la realización de un "preprocesamiento", encargado de eliminar signos de puntuación, transformar números a letras, eliminar espacios sobrantes, etc., a los efectos de obtener una lista de términos preparada para ser analizada por el algoritmo seleccionado.

Además, los algoritmos existentes de *Machine Learning* fueron considerados de manera previa: *Machine Learning* refiere al empleo de técnicas avanzadas de búsqueda de patrones y correlaciones entre los datos, cuando no se pueden obtener las mismas de forma procedural o expresadas mediante reglas. El algoritmo *Topic Modelling*³, a su vez, fue descartado por no contar con un *corpus* de tamaño considerable. El mismo es utilizado para determinar temáticas dentro de un conjunto de documentos. Es decir, si tomásemos un centenar de publicaciones científicas, este algoritmo podría agrupar palabras, del *corpus* de publicaciones dadas, que respondan al mismo tema. Este algoritmo es más bien empleado para revelar el contenido de diversos documentos e indicar la probabilidad de pertenencia de los documentos a los grupos encontrados. Y en segundo lugar, los algoritmos de clasificación (*k-means*) también fueron descartados por no ser relevantes para los objetivos de la presente investigación. Los algoritmos de clasificación aplicados a textos permiten realizar agrupaciones como particiones de los documentos dados como entrada. Estos algoritmos son empleados cuando el objetivo perseguido es encontrar documentos similares dentro de un *corpus* amplio de documentos.

Hasta aquí se han explicitado las decisiones metodológicas que se fueron tomando con el avance del trabajo. A continuación se presenta el trabajo efectuado para uno de los hoteles seleccionados, y la interactividad cliente–hotel, a partir de las opiniones que del hotel vierten en las páginas los clientes.

³<http://www.cs.columbia.edu/~blei/topicmodeling.html>

Opiniones de los clientes en las páginas Web2.0 *Trip Advisory* y *Booking*, hotel de 5 estrellas de la ciudad de Paraná

La clasificación de la información partió de la transcripción de los comentarios u opiniones de las páginas a la hoja de Excel. Luego se realizó desde la minería de textos una síntesis. En el caso *TripAdvisor* sobre un *corpus* de 51 comentarios en y en *Booking* sobre un total de 339 comentarios. Los comentarios extraídos fueron los enviados por huéspedes a los hoteles⁴ seleccionados para su análisis. Dentro de los comentarios, experiencias y opiniones de las personas se deducen las expresiones que implican calificativos positivos y otras de tipos negativos. El sustento de esta conjetura se basó en las palabras que utilizaron para expresar sus sentimientos con respecto a la estadía y el servicio recibido.

Página Web 2.0 - *Trip Advisor* y *Booking*

Con el sentido de exponer el proceso realizado por el equipo de investigación, se seleccionó uno de los hoteles para mostrar la sistematización de la información trabajada con minería de datos. En primera instancia se procesó lo extraído del sitio web *Tripadvisor* que recibió un total de 51 posts (comentarios realizados por los clientes), de los cuales desde el hotel se realizaron un total de 33 respuestas, lo que representa un porcentaje total del 64.71 % de respuestas. Esto permitió decir que el hotel está relativamente presente en la página web e interactúa con bastante frecuencia con los usuarios.

Del total de 51 posts, 29 pudieron ser calificados de carácter positivo por los términos empleados por los usuarios, 4 fueron calificados de carácter negativo, y los otros, que sólo describen características del hotel sin dejar entrever apreciaciones específicas, se los consideró neutros. Esto demuestra que en su mayoría los comentarios con respecto al hotel son buenos.

⁴ Al ser información pública la que se obtiene de las páginas, lo mismo decidimos no aclarar el nombre del hotel con que se trabajó el *corpus*.

Cuadro 1: Análisis de los comentarios extraídos de los clientes según las páginas web 2.0 *TripAdvisor* y *Booking* de un hotel 5 estrellas. Ciudad de Paraná (período 1 de octubre 2017 a 30 de setiembre 2018)

Howard de 5 estrellas	Trip Advisor	Booking
Total de posts	51	339
Total de posts respondidos	33	2
Porcentaje respuestas	64,71	0,59
Posts (+)	29	233
Posts(-)	4	101
Hombre	27	182
Mujer	24	106
N/A	0	51
Porcentaje usuario hombre	52,94	53,69
Porcentaje usuario mujer	47,06	31,27
Porcentaje usuario N/A	0	15,04

(Elaboración propia en base a las páginas web 2.0 consultadas)

Páginas Web 2.0 *TripAdvisor*

Para el análisis de este sitio web otra variable considerada fue el sexo del usuario o cliente que realizó el post. Y lo constatado es una diferencia mínima, como se observa en el Cuadro 1, 27 comentarios fueron realizados por hombres lo que representa un 52.94% y 24 realizados por mujeres, lo que dan 47.06 % de comentarios realizado.

La otra información trabajada sobre los usuarios, aparte de lo referido al sexo⁵, es importante aclarar que al tener posibilidades de seleccionar en dos ítems, existen cuentas anónimas de las cuales no se puede extraer información de ningún tipo. Respeto a la nacionalidad que es otro punto posible de procesar, la mayoría que escribieron los posts son de origen argentino, y unos pocos pertenecían a usuarios de los países como de Uruguay y de España.

Nubes de palabras

Como se expresó en la parte metodológica para la representación gráfica recurrimos a la nube de palabras, esta herramienta deja traslucir a los clientes o usuarios a través de sus comentarios, y permite, a partir de las palabras repetidas, y por ende se obtuvo la preponderancia de éstas en la

⁵ El atributo para calificar ha sido por sexo y lo disponible en las plataformas relevadas admite solamente los estados: Hombre y Mujer

generar el gráfico de coocurrencia. En el Gráfico 3 se puede apreciar con mayor claridad la relación entre los principales términos o palabras claves utilizadas en los posts y la vinculación lograda entre sí. Este tipo de representación es utilizado para reforzar el análisis estadístico de texto.

El Hotel de 5 estrella ha sido apreciado por sus huéspedes, sobre todo, por su vista al río Paraná; y sus habitaciones las cuales son valoradas y catalogadas como excelentes. Entre las características generales percibidas por los clientes en su mayoría mencionan de tipo positivo, y esto se puede leer en términos o calificativos que dan ideas de una percepción buena expresada en los comentarios. Entre ellas se reconocen palabras que lo califican como amplio, cómodo, lindo, excelente entre otros modos utilizados en los comentarios.

Página Web 2.0 *Booking*

En el estudio se tuvo en cuenta el otro sitio web 2.0 llamado *Booking*, sobre el cual se efectuó el mismo proceso antes mencionado; en Excel se cargaron las opiniones y luego la información la convertimos en los gráficos. El período de análisis fue el mismo y como se registra en el Cuadro 1, este sitio recibió un total mayor de posts, se computaron 339 posts (comentarios realizados por los clientes) y por el contrario a la página antes analizada, al momento de interactuar desde el establecimiento hotelero, sólo se encontraron escasas respuestas. Lo que representa un porcentaje total del 0,59 % de respuestas: esto permite entender que la presencia e interacción del Hotel 5 estrellas en la página web de *Booking* ha sido casi nula.

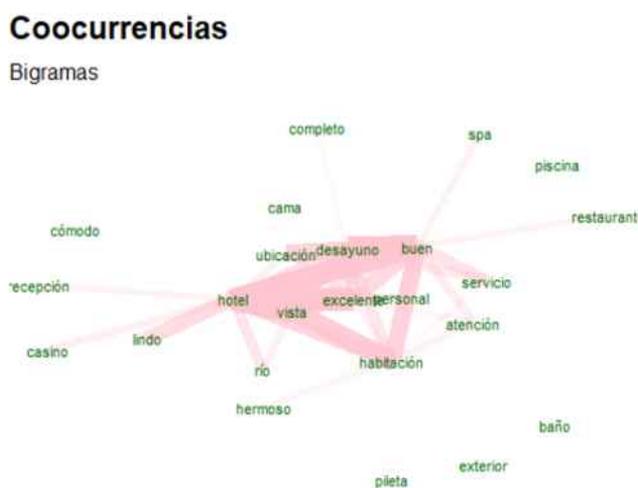
Dentro de los comentarios, experiencias y opiniones de los clientes encontramos comentarios que aluden a visiones positivas del hotel y otras no tantas. Las expresiones fueron a través de palabras que presentaban sus sentimientos con respecto a la estadía y el servicio brindado. En este caso del total de 339 posts, 233 (69%) fueron consideradas de carácter positivo y 101 (39%) fueron de carácter negativo. Esto demuestra que por la frecuencia y reiteración de palabras los comentarios relevados hacia el hotel revelan una percepción en el que prevalecen los comentarios positivos.

Aquí también se trabajó para distinguir a los clientes por el género del que escribe el post, de los 186 comentarios se detectaron que fueron realizados por hombres, lo que representa un 53,69%. Luego hay 106 realizados por mujeres, lo que representa un 31,27% y el resto de los comentarios han

Coocurrencia de palabras

A partir del análisis del bigrama de coocurrencias interrelacionamos las opiniones de los clientes del Hotel, que en este caso refleja la relación directa entre el término *hotel* y *vista*, sumado a calificativos como *lindo* y *excelente*. A su vez, mencionaron una relación positiva entre las palabras *habitación*, *ubicación* y *desayuno*; nuevamente aparece la *vista*, el *río* y el *spa*. Otra de las relaciones destacada está en referencia a los servicios y modos de trabajar del personal destacándose la *atención* y *recepción*, que pueden asociarse a los términos *buen* y *excelente*.

Gráfico 3: Ilustración Nube de palabras Howard Jonshon Mayorazgo (elaboración propia).



A modo de cierre

En esta exposición en la que hemos recortado y seleccionado una parte de nuestro trabajo, sintetizamos el manejo y proceso posible de plasmar desde la minería de datos, parte del análisis de contenido computacional y develar las posibilidades que hoy encontramos en los soportes tecnológicos. En esta oportunidad recogiendo y procesando las opiniones que los clientes dejan acerca de sus estadías en un hotel de 5 estrellas de la ciudad de Paraná en las páginas 2.0 de *Booking* y *TripAdvisor*, en el periodo de análisis desde 1 de octubre de 2017 al 30 de setiembre de 2018.

En general los resultados nos permitieron constatar que la interactividad entre clientes y hoteles presenta heterogeneidades, siendo mucho más acentuada la interacción en *TripAdvisor* que en

Booking. Esto puede ser factible debido a que dichas páginas en su generación o construcción apuntaban a objetivos diferentes, y aventuramos que con el tiempo estas estrategias comerciales han aumentado.

Aunque el período de tiempo del estudio fue acotado, durante el mismo pudimos constatar un crecimiento en la interactividad, pues esta modalidad con el tránsito del tiempo se selecciona más por la propia competencia que se genera al momento de seleccionar un hotel o indagar, pues actualmente en las decisiones de los posibles clientes cobran peso los comentarios.

Entre los ejes o problemáticas más relevantes planteadas por los usuarios o clientes en cada plataforma, cobra importancia el estado de las habitaciones, la vista del hotel, el servicio del personal y la calidad del desayuno.

En la descripción y presentación mencionamos las opiniones y las respuestas, pero no expresamos que hay extensos plazos en los tiempos empleados, entre la aparición de los comentarios y la respuesta a los mismos; en otros términos esto denota una lenta capacidad de interacción. En otros post´s se incurre con la falta total de respuestas por parte del hotel. Y en lo que se registran respuestas, éstas son cordiales y similares en sus formatos y escrituras.

Por último, observamos que el origen de los huéspedes que han interactuado con las mencionadas plataformas ha sido en mayor parte de nacionalidad argentina, seguida por nacionalidades de América Latina, y luego españoles y estadounidenses. En ambas páginas los comentarios positivos hacia el hotel y sus servicios tienen mayor relieve que los comentarios negativos.

Bibliografía citada

- ❖ Domas White, D. y E. Marsh, 2006. "Análisis de contenido: una metodología flexible". Disponible en: <https://www.ideals.illinois.edu/handle/2142/3670>
- ❖ Krippendorff, K., 1990. *Metodología de análisis de contenido. Teoría y práctica*, Paidós, Barcelona.
- ❖ Mayntz, R.; Holm, K. y P. Hubner, 1993. *Introducción a los métodos de sociología empírica*. Alianza Editorial S.A., Madrid.

- ❖ Marketing Digital, 2017. *El Blog de Content Marketing*. Disponible en: <https://coobis.com/es/cooblog/que-es-una-nube-de-palabras/>
- ❖ Mendoza, M., 2011. *Minería de datos en la Web*, disponible en: https://www.researchgate.net/publication/229068670_Mineria_de_datos_en_la_Web-
- ❖ Mingo, G., 2009. Ficha de cátedra Metodología Cualitativas: "Análisis de contenido cuantitativo". Facultad de Trabajo Social, UNER. Paraná.
- ❖ Moreiro González, J., 2002. "Aplicaciones al análisis automático del contenido provenientes de la teoría matemática de la información". *Anales de Documentación*, Universidad de Murcia. Disponible en: <https://revistas.um.es/analesdoc/article/view/2101>
- ❖ Llibre, M., 2015. *El golpe de internet a la hotelería con Booking, TripAdvisor y Airbnb*. Disponible en: <https://www.reportur.com/colombia/2015/08/01/el-golpe-de-internet-a-la-hoteleria-con-booking-tripadvisor-y-airbnb/>

Cita: Mingo, G. y M. Testaseca, 2021. "El uso tecnológico desde las páginas web 2.0 en el sector hotelero de Paraná (provincia de Entre Ríos)" (pp. 115-128), *Tiempo de Gestión* N° 30, FCG-UADER, Paraná.

