

# Imagen de un destino a partir de reseñas de turistas en buscadores de Internet

Image of a destination from tourists' reviews in Internet search engines

Estela Mariné-Roig<sup>1</sup>, Salvador Anton Clavé<sup>2</sup>

<sup>1</sup> Universidad de Lleida, España

<sup>2</sup> Universidad Rovira i Virgili, España

estela.marine@aegegn.udl.cat , salvador.anton@urv.cat

**RESUMEN.** En los últimos años, el crecimiento espectacular de los contenidos generados por los usuarios se ha reflejado en el campo del turismo en la proliferación de opiniones en línea sobre viajes (OTRs). El objetivo de este artículo es proponer un método para analizar la imagen percibida (y transmitida) por los viajeros, mediante los OTRs, tal como se presenta en los buscadores, haciendo hincapié en los componentes cognitivo, evaluativo y afectivo de la imagen del destino turístico situándola en el espacio y en el tiempo. Para ello, se realiza un análisis cuantitativo de algunos metadatos procesados por los buscadores, a partir de una muestra aleatoria de 150.000 OTRs alojados en TripAdvisor, sobre "cosas que hacer" en Cataluña, escritos en español por turistas entre 2012 y 2016. El análisis masivo de datos semiestructurados generados por usuarios da fiabilidad a los resultados obtenidos, que pueden ser de gran utilidad para optimizar la oferta turística.

**ABSTRACT.** In recent years, the spectacular growth of user-generated content has been reflected in the field of tourism with the proliferation of online travel reviews (OTRs). The aim of this paper is to propose a method to analyse the image perceived (and transmitted) by travellers, through OTRs, as presented in search engines, emphasizing the cognitive, evaluative and affective components of tourist destination image, and situating it in a certain space and time. To this end, a quantitative analysis of some metadata processed by search engines, from a random sample of 150,000 TripAdvisor OTRs on "things to do" in Catalonia, written in Spanish by tourists between 2012 and 2016 is conducted. The analysis is carried out on massive semi-structured data generated by users, which gives reliability to the obtained results, and that can be useful for tourism supply optimization.

**PALABRAS CLAVE:** Imagen de un destino, Reseñas de viaje, Buscadores, Datos masivos, Minería de datos, Metadatos HTML, Catalunya.

**KEYWORDS:** Destination image, Travel reviews, Search engines, Big data, Data mining, HTML metadata, Catalonia.

## 1. Introducción

El uso de Internet y de los medios de comunicación social se ha incrementado en los últimos años y sigue creciendo. En 2016, casi 4 de cada 5 europeos (79%), entre 16 y 74 años, usaron regularmente los servicios de Internet (Eurostat, 2017). Los medios sociales tienen grandes potencialidades para establecer relaciones con los usuarios, conocer sus imágenes y necesidades, permitir comentarios y participación / interacción, y comunicar con eficacia marcas de destino (Huertas & Marine-Roig, 2016, p. 292). En una encuesta realizada a más de 30.000 personas seleccionadas entre diferentes grupos sociales y demográficos de la Unión Europea (Eurobarometer, 2016), resultó que la segunda fuente de información más usada para planificar las vacaciones fue el boca a boca electrónico (eWoM: electronic Word-of-Mouth): sitios web que recogen y presentan comentarios, opiniones y valoraciones de viajeros (34%); ligeramente por encima de las fuentes primarias: experiencia personal (33%), y muy por encima de las fuentes de imagen turística inducida: sitios web gestionados por los proveedores de servicios o por los destinos turísticos (17%), y agencias de viajes y oficinas de turismo (13%). Además, una gran mayoría de encuestados reservó o contrató sus vacaciones a través de servicios en línea. En relación a los viajeros norteamericanos, el porcentaje de uso de Internet como fuente de información para planificar un viaje es superior al de los europeos y se sitúa en torno al 85% (Kim, 2015; Xiang et al., 2015). Los resultados anteriores son consistentes con los obtenidos por Llodrà-Riera, Martínez-Ruiz, Jiménez-Zarco e Izquierdo-Yusta (2015) que, en un estudio sobre visitantes de Mallorca, destacaron Internet como principal fuente de información y, dentro de ella, buscadores (p.e. Google), mapas (p.e. Google Maps) y portales con valoraciones de los viajeros (p.e. TripAdvisor). Similarmente, en una encuesta a 11.400 visitantes internacionales de Gran Bretaña (VisitBritain, 2017), resultó que los buscadores de Internet fueron la fuente en línea más influyente en la selección del destino turístico, seguido de los portales de comparación de precios y de los que recogen comentarios o valoraciones de los viajeros.

El eWoM se basa principalmente en contenidos generados por los usuarios (UGC: user-generated content) a los cuales se puede acceder en línea gratuitamente. El UGC es una información de primera mano, desinteresada y no solicitada y, por ello, se considera fiable por los demás usuarios. En el campo del turismo, el UGC se plasma principalmente en los diarios de viaje en línea, mediante los cuales la imagen percibida y transmitida, por el efecto eWoM, se convierte en imagen proyectada y contribuye a cerrar el círculo de construcción de la imagen del destino turístico (TDI: tourist destination image) desde una perspectiva holística (Marine-Roig, 2015). En los últimos años han proliferado enormemente los diarios de viaje conocidos como OTRs (online travel reviews) mediante los cuales los turistas relatan libremente sus experiencias en el destino que han visitado y dan su opinión y/o valoración sobre atracciones (monumentos, museos, parques, etc.) y servicios (hoteles, restaurantes, transportes, etc.) concretos. A título de ejemplo del crecimiento de los OTRs, se puede citar que, en Enero de 2015, Booking afirmaba tener 43 millones de OTRs de huéspedes reales verificados y TripAdvisor más de 200 millones de comentarios y opiniones y, en Septiembre de 2017, afirmaban haber superado los 130 y 535 millones de OTRs respectivamente.

Vista la importancia del eWoM marketing en la fase preparatoria de las vacaciones y, concretamente, de las fuentes de información en línea que recogen y presentan comentarios, opiniones y valoraciones de viajeros, así como de los motores de búsqueda para localizar y acceder a dichas fuentes de información, el objetivo de este artículo es proponer una metodología para el análisis masivo de OTRs sobre un destino turístico, con el fin de dilucidar la imagen percibida por sus visitantes, tal como se presenta y difunde a través de los buscadores de Internet. Para probar el método, se analizan determinados metadatos HTML (HyperText Markup Language) que procesan los navegadores y motores de búsqueda de Internet, a partir de una muestra aleatoria de 150.000 OTRs alojados en TripAdvisor, escritos en español por visitantes de Cataluña entre 2012 y 2016, y se obtienen resultados significativos en tres dimensiones de la TDI (cognitiva, evaluativa y afectiva) situándola en el espacio y en el tiempo.

## 2. Marco teórico

Una vez tratada en la introducción la idoneidad del UGC y, concretamente, de los OTRs para conocer las experiencias, opiniones, comentarios y valoraciones que conforman la imagen percibida por los viajeros, en



esta sección se estudian los fundamentos y componentes de la TDI, así como su proceso y presentación en los navegadores y buscadores de Internet.

## 2.1. La imagen de un destino y sus componentes

La TDI se puede definir como la suma de creencias, ideas e impresiones que una persona tiene sobre un destino (Crompton, 1979, p.18). Su estudio ha sido constante en la literatura científica sobre turismo (Chon, 1990; Li, Ali & Kim, 2015), porque se considera, entre otros motivos, que tiene un papel crucial en la toma de decisiones relacionadas con la adquisición individual de viajes (Chon, 1990).

De acuerdo con Baloglu y McCleary (1999), la construcción de la TDI depende de dos evaluaciones: la perceptiva / cognitiva, referida a las creencias o conocimiento sobre los atributos de un destino, y la afectiva, referida a los sentimientos o al apego hacia el mismo. La imagen percibida global está formada por la unión de ambas (Beerli & Martín, 2004; Marine-Roig, 2015; Tan & Wu, 2016). Un tercer componente se deriva directamente de los dos anteriores, el conativo / volitivo, que afecta al comportamiento del individuo a la hora de seleccionar un destino, en base a las imágenes percibidas durante la fase cognitiva y evaluadas durante la fase afectiva (Gartner, 1993). Esta dicotomía de imagen cognitiva-afectiva ha sido estudiada profusamente en el campo del turismo (Kim & Perdue, 2011). Ahora bien, las imágenes de un destino son percibidas subjetivamente por individuos en un entorno determinado y en un momento dado. Por ejemplo, la imagen percibida del paseo marítimo de un destino situado en la costa mediterránea no es la misma en verano que en invierno. En esta línea, como precedente de la TDI evaluativa, Pocock y Hudson (1978) contemplan la TDI designativa basada en el conocimiento individual del qué es y en qué lugar del entorno se sitúa ('whatness' y 'whereness' básicos de la imagen). Según estos autores, el aspecto designativo es informativo por naturaleza, concerniendo a la descripción y clasificación, y por sí mismo puede ser menos importante que el aspecto estimativo (el significado adjunto a, o evocado por, la forma física).

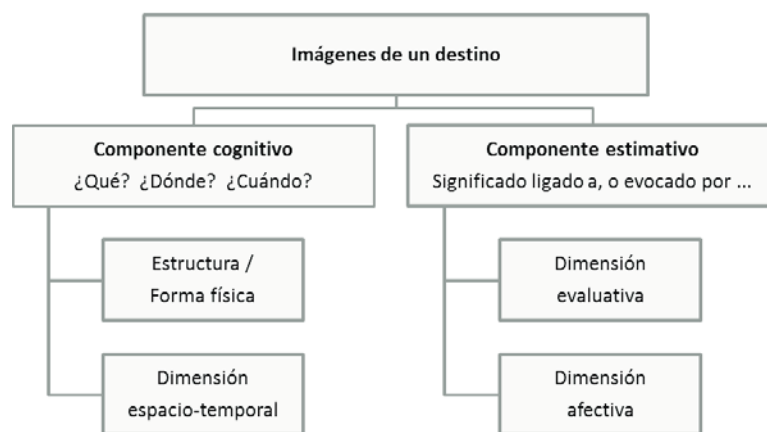


Figura 1. Componentes de la imagen de un destino. Fuente: Adaptación (Marine-Roig, 2017).

Para profundizar en el estudio de la TDI, en la Figura 1, se propone un esquema que sintetiza las dos dicotomías más comunes sobre los componentes de la imagen global: 'cognitive-affective' (Rapoport, 1977) y 'designative-appraisive' (Pocock & Hudson, 1978). La estructura / forma física incluye características como figura, tamaño, textura y diseño. Las características espaciales se refieren a la distancia, situación relativa y orientación. El aspecto espacial de la imagen designativa fue estudiado por Son (2005) mediante la obtención de croquis que, a su vez, son usados para medir mapas mentales, y por Marine-Roig y Anton Clavé (2016b) mediante coeficientes espaciales para elucidar la especialización de la imagen en destinos multiescalares. La dimensión temporal gira en torno a la fecha de percepción de la imagen. El componente estimativo incorpora tanto la evaluación como la preferencia, la primera incluye algunos estándares generales o externos y la segunda refleja un tipo más personal de valoración y afectación, que es la respuesta emocional relacionada con

el valor, el sentimiento y el significado vinculados a la imagen percibida (Pocock & Hudson, 1978).

## 2.2. Búsqueda de información turística en línea

Por su potencial para indexar y organizar enormes cantidades de información, los motores de búsqueda son herramientas poderosas para representar el mundo virtual y, por tanto, el dominio del turismo (Xiang, Wober & Fesenmaier, 2008, p. 146), y va creciendo la importancia del papel que juegan en los programas de marketing de las organizaciones turísticas en línea (Pan et al., 2010). Pan y Fesenmaier (2006) proponen un modelo conceptual de la planificación de un viaje a través de Internet basado en la interacción entre usuario y la parte de la Web relacionada con la industria y destinos turísticos. A partir de este marco, se puede plantear un modelo de búsqueda de información en Internet con tres componentes: el viajero, la interfaz y el espacio en línea (Xiang et al., 2008). La efectividad de una búsqueda depende de la situación, conocimientos y habilidades del usuario (viajero), de la cantidad y calidad de los sitios web relacionados (dominio turístico en línea), y de las funcionalidades de los exploradores web y motores de búsqueda (interfaz) usados para facilitar los resultados.

Los motores de búsqueda constan básicamente de dos partes: un analizador sintáctico que recorre atemporalmente la Web y recoge la información más significativa que sirve para construir una base de datos indexada por palabras o frases clave, y un componente en línea que atiende las consultas de los usuarios y les retorna los resultados correspondientes ordenados por idoneidad y visibilidad (Mariné-Roig, 2014). Estos resultados se suelen presentar en base a metadatos de la página web indexada, incluyendo el título con un hipervínculo y un breve resumen (Xiang et al., 2008). En la Figura 2 se puede observar la presentación de los resultados, sobre una misma consulta, por parte de los tres buscadores con más tráfico del mundo: Google, Yahoo y Baidu (Alexa.com, TopSites). Es de destacar que Yahoo no usa medios propios y presenta los resultados obtenidos por el motor de búsqueda Bing de Microsoft.

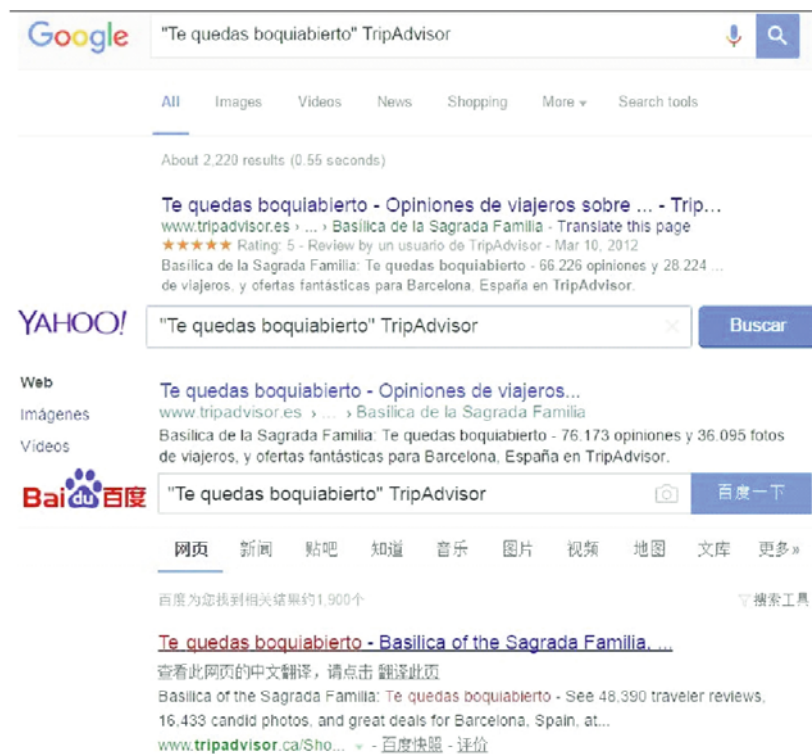


Figura 2. Presentación del mismo OTR en los 3 buscadores con más tráfico del mundo. Fuente: Resultado de la búsqueda el 30-05-2016 de un OTR en Google, Yahoo y Baidu.



Los metadatos son un conjunto de datos que describen o dan información acerca de otros datos. La etiqueta web <meta ... /> (W3Schools, 2017) contiene metadatos sobre un documento HTML. Dichos metadatos no se visualizan en la página web porque están destinados a facilitar información a los exploradores y a los motores de búsqueda de Internet. Los metadatos HTML consisten en parejas compuestas de nombre y contenido (Tabla 1). Los elementos más comunes son la descripción de la página web, la relación de palabras clave, y el nombre del autor del documento (W3Schools, 2017). A pesar de que la información contenida en la Tabla 1 y en la Figura 2 fue recogida el mismo día, se puede observar una discrepancia en la cantidad de opiniones y de fotos, debido a que la página web se actualiza diariamente, mientras que los tres motores de búsqueda, que capturaron los metadatos en fechas anteriores, no consideran necesario actualizarlos tan a menudo.

| Meta name          | Content                                                                                                                                                                                                                                                                     |
|--------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>title</i>       | Te quedas boquiabierto - Opiniones de viajeros sobre Basílica de la Sagrada Familia, Barcelona – TripAdvisor                                                                                                                                                                |
| <i>hyperlink</i>   | <a href="https://www.tripadvisor.es/ShowUserReviews-g187497-d190166-r125908839-Basilica_of_the_Sagrada_Familia-Barcelona_Catalonia.html">https://www.tripadvisor.es/ShowUserReviews-g187497-d190166-r125908839-Basilica_of_the_Sagrada_Familia-Barcelona_Catalonia.html</a> |
| <i>keywords</i>    | Basílica de la Sagrada Familia, Barcelona, España, Te quedas boquiabierto                                                                                                                                                                                                   |
| <i>description</i> | Basílica de la Sagrada Familia: Te quedas boquiabierto - 76.793 opiniones y 36.386 fotos de viajeros, y ofertas fantásticas para Barcelona, España en TripAdvisor                                                                                                           |

Tabla 1. Contenido (30-05-2016) de las meta-etiquetas HTML más destacadas (ver Figura 1). Fuente: OTR en TripAdvisor sobre la Sagrada Familia fechado el 10-03-2012.

### 3. Metodología

La metodología usada para el análisis masivo de los metadatos HTML es una adaptación del método para la recogida masiva, clasificación y depuración de blogs y reseñas de viaje descrito en Marine-Roig y Anton Clavé (2015) y detallado en Marine-Roig y Anton Clavé (2016a). Consta de las siguientes fases: Recolección de la información, minería de metadatos, y análisis cuantitativo.

#### 3.1. Caso de estudio

Cataluña es un destino turístico destacado a nivel europeo (Eurostat, 2017), cuya afluencia de visitantes se ha incrementado en los años considerados (2012-2016). La ocupación de los establecimientos hoteleros catalanes (OdTiMP, 2017) ha experimentado la siguiente evolución anual (en millones de viajeros y pernoctaciones): 2012 (16,2 y 48,4), 2013 (16,2 y 49,3), 2014 (16,8 y 50,2), 2015 (17,7 y 52,0) y 2016 (19,1 y 55,6). En 2016, igualmente se ha incrementado la ocupación en establecimientos extra hoteleros (campings, casas rurales, albergues y apartamentos turísticos). Aproximadamente el 75 % de todos estos viajeros procede del extranjero, habiéndose batido récords históricos en 2016 de llegada de turistas extranjeros de primer destino y gasto total (18,1 y 17.434,1 millones respectivamente).

Cataluña es un destino multiescalar (Marine-Roig & Anton Clavé, 2016b) que tiene dividido el territorio en nueve marcas turísticas (ACT, 2017): Barcelona (Barna), Costa Barcelona (cBarc), Costa Brava (cBrav), Costa Daurada (cDaur), Paisatges Barcelona (pBarc), Pirineus (Pyren), Terres de Lleida (tLlei), Terres de l'Ebre (tEbre), y Val d'Aran (vAran). Estas marcas están formadas por comarcas limítrofes con una oferta turística relativamente homogénea que especializa territorialmente la TDI, a pesar de que los destinos en Cataluña son gestionados y/o promocionados por distintos niveles administrativos (Datzira-Masip & Poluzzi, 2014). Los destinos de ámbito superior a la comarca (por ejemplo, provincias o regiones) se consideran no clasificados (unCla) en este caso de estudio.

#### 3.2. Recolección de la información

La principal fuente de relatos y opiniones en línea acerca de viajes son los sitios web que recogen blogs de viaje y OTRs, porque presentan la información de forma estructurada, lo cual permite automatizar su descarga, clasificación y análisis. En base a trabajos anteriores (Marine-Roig & Anton Clavé, 2016a) y a una actualización

de la búsqueda, se localizan 8 portales con información sobre Cataluña en la época considerada: GetJealous.com (GJ), MiNube.com (MN), MyTripJournal.com (MT), TravBuddy.com (TY), TravelBlog.org (TB), TravelJournals.net (TJ), TravellersPoint.com (TS), y TripAdvisor.com (TA). Para seleccionar la fuente más idónea se construye un ranking aplicando la fórmula ponderada (Mariné-Roig, 2014):

$$TBRH = 1*B(V) + 1*B(P) + 2*B(S)$$

donde 'B' corresponde al método de ordenación de Borda (de Borda, 1781), 'V' visibilidad de la web (cantidad y calidad de enlaces entrantes), 'P' su popularidad (visitas que recibe y tráfico en general) y 'S' el número de entradas relacionadas con el caso de estudio. TBRH consiste en una agregación ponderada de rankings: En base a tres listas completas (L) de webs ordenadas por 'V', 'P' y 'S', la función 'B' asigna una puntuación a cada candidato (c) consistente en el número de candidatos situados debajo de 'c' en 'L'. Esta puntuación se multiplica por el peso de cada ranking. Una vez sumadas las puntuaciones parciales, los candidatos se clasifican en orden descendente de puntuaciones totales.

Una vez aplicada dicha fórmula, resulta en primer lugar TA con mucha diferencia sobre las demás sitios web relacionados con los viajes. A título de ejemplo, se puede citar que, el 2017-10-10, el segundo clasificado (TB) disponía de 2.256 entradas de blog sobre Barcelona y TA de 2.405.267 opiniones de viajeros, o que MN disponía de 669 opiniones sobre la Basílica de la Sagrada Familia y TA de 121.860 opiniones y 72.273 fotografías. En principio, había dos portales que almacenaban gran cantidad de blogs y reseñas de viaje, TravedPod.com (TP) y VirtualTourist.com (VT), pero hace pocos años que fueron adquiridos por TripAdvisor Inc. y este año han sido cerrados definitivamente. Esta selección coincide con Baka (2016) que considera a TA la mayor fuente mundial de UGC en el dominio del turismo, así como numerosos autores (de Rosa, Bocci, & Dryjanska, 2017; Liu, Teichert, Rossi, Li, & Hu, 2017; Pantano, Priporas, & Stylos, 2017; Ríos, Ortega, & Matilla, 2016) que seleccionaron TA como fuente de información por sus múltiples ventajas. Descartando los centenares de miles de OTRs sobre hospedaje y restauración, TA (TripAdvisor, 2017) presenta unos 750.000 OTRs sobre "cosas que hacer" (Things to Do) en Cataluña de los cuales unos 250.000 están escritos en inglés y más de 150.000 en español entre 2012 y 2016. Los OTRs se pueden descargar mediante un programa copiador de webs como HTTrack Website Copier: Open source offline browser (httrack.com).

### 3.3. Minería de metadatos HTML

Observando la Figura 2 y la Tabla 1, ambas relativas al mismo OTR, se puede deducir que los tres buscadores obtienen la información presentada en el contenido delimitado por las meta-etiquetas HTML title y description. Además, asociado con el título hay un hiperenlace (hyperlink) que conduce a la página web que aloja el OTR. A su vez, title se compone de dos partes, el título propiamente dicho escrito por el viajero y una información asociada (atracción o servicio, destino y titular del sitio web) añadida por el administrador de la web (webmaster). El hiperenlace (Tabla 1) aporta muchos datos que facilitan el proceso de análisis de la información asociada: protocolo (https: Secure HyperText Transport Protocol), servidor (www.tripadvisor.es), finalidad de la página web (ShowUserReviews: mostrar OTRs de usuarios), código del destino (g187497: Barcelona), código de la atracción (d190166: Basílica de la Sagrada Familia), código del OTR (r125908839), nombre interno de la atracción (Basílica of the Sagrada Familia), nombre del destino (Barcelona), nombre de la región (Catalonia), y tipo de página web (html). Por otra parte, a diferencia de los otros dos buscadores, Google (sitio web con más tráfico del mundo) presenta una línea adicional con la valoración (en una escala de 1 a 5 estrellas) de la atracción o servicio, y la fecha del OTR (Mar 10, 2012).

Gracias a dicha estructura de las páginas web, se pueden extraer los metadatos descritos mediante expresiones (regex: Regular expression) simples de lenguaje regular (secuencias de caracteres que forman un patrón de búsqueda) a través de un programa que las admita y permita trabajar con grandes cantidades de datos, como UltraEdit (ultraedit.com) que admite ficheros con un peso superior a 4 GB (GigaBytes).

Para poder procesar los ficheros en un sistema operativo Windows, se convierten de UNIX/UTF-8 a



PC/ANSI y, de acuerdo con Marine-Roig y Anton Clavé (2016a), se substituyen todos los caracteres especiales (los superiores al ASCII 127: American Standard Code for Information Interchange de 7 bits) por el correspondiente carácter ISO 8859-15 (International Standards Organization) conocido también como alfabeto latino 9. Finalmente, se guardan los metadatos en formato CSV (comma separate values) para poder manipular unos ficheros de texto plano mediante una aplicación para hojas de cálculo de Microsoft Office o Apache OpenOffice por ejemplo.

### 3.4. Análisis cuantitativo de contenido

El análisis de contenido se puede definir brevemente como el análisis sistemático, objetivo y cuantitativo de las características de toda forma de comunicación registrada (Camprubí & Coromina, 2016; Neuendorf, 2017). Tiene como objetivo describir, con óptima objetividad, precisión y generalidad, lo que se dice sobre un tema determinado en un lugar y en un momento dados (Stepchenkova, 2012). Para ello se emplea una técnica replicable para comprimir palabras o frases clave en unas cuantas categorías que se pueden constituir a priori o se pueden deducir del propio texto. El análisis de contenido se basa generalmente en la cuenta de las palabras o frases, porque se asume que las palabras más frecuentes son las que generan mayor consenso, y permite a los investigadores tamizar grandes volúmenes de datos con relativa facilidad de una manera sistemática (Stepchenkova, 2012). En este caso, el análisis se centra en las dimensiones espacial y temporal de la TDI, así como en sus componentes cognitivo y estimativo (Figura 1).

- **Dimensión espacio-temporal.** Para poder estudiar la dimensión espacial de la imagen en un destino multiescalar es conveniente clasificar los OTRs por marcas turísticas territoriales. En la subsección 3.3 se ha visto que los OTRs están asociados a un destino, entonces es necesario preparar una lista de destinos de cada marca en el cual consten los nombres de la propia marca, de las comarcas, de los municipios y de otras entidades consideradas destinos como ‘La Pineda’ (zona urbanizada a caballo entre los municipios de Vila-seca y Salou) o ‘Baqueira-Beret’ (centro de vacaciones en el Valle de Arán). Comparando esta información con la extraída en la fase anterior, se pueden clasificar los OTRs en el espacio y en el tiempo (Tabla 2 y Figura 3).

|      | Barna | cBarc | cBrav | cDaur | pBarc | Pyren | tEbre | tLlei | vAran | unCla |
|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 2012 | 5395  | 336   | 821   | 760   | 139   | 170   | 10    | 44    | 14    | 5     |
| 2013 | 6721  | 528   | 1101  | 942   | 249   | 400   | 40    | 110   | 28    | 14    |
| 2014 | 11149 | 1253  | 1893  | 1610  | 419   | 755   | 104   | 162   | 107   | 58    |
| 2015 | 27343 | 3133  | 4989  | 4375  | 942   | 1654  | 577   | 509   | 276   | 91    |
| 2016 | 43679 | 6378  | 7844  | 5928  | 1680  | 2770  | 1081  | 859   | 412   | 143   |

Tabla 2. Los 150.000 OTRs sobre “cosas que hacer” en Cataluña por marca turística y año. Fuente: Muestra aleatoria de OTRs de TripAdvisor escritos en español entre 2012 y 2016.

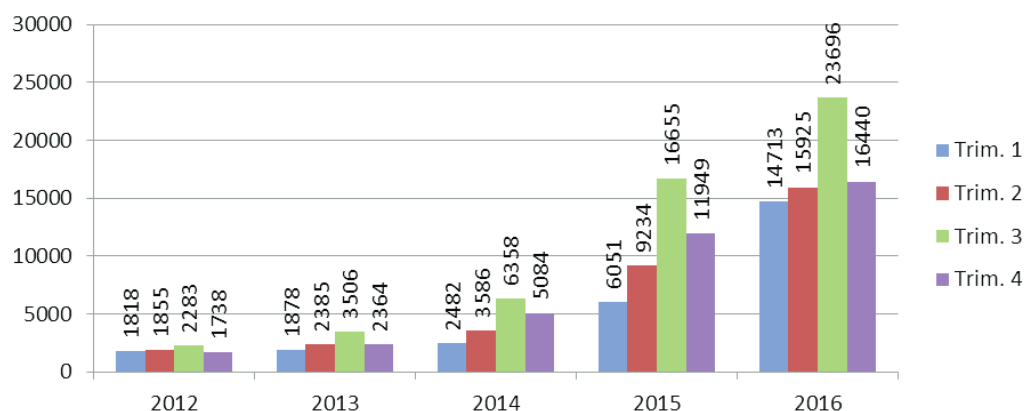


Figura 3. Los 150.000 OTRs sobre “cosas que hacer” en Cataluña por trimestres. Fuente: Muestra aleatoria de OTRs de TripAdvisor escritos en español entre 2012 y 2016.

- **Componente cognitivo.** Los metadatos obtenidos en la sección 3.3 permiten deducir que atributos del destino son más considerados por los autores de los OTRs, por ejemplo, qué atracciones se visitan con mayor frecuencia o qué servicios se usan más a menudo. También permiten agruparlos por tipos o categorías (patrimonio tangible, sol y playa, entorno urbano, ocio y esparcimiento, etc.) para poder comparar los resultados con los de otros destinos.

- **Componente estimativo (evaluativo y afectivo).** Por lo que respecta al componente estimativo, se dispone de dos fuentes de información (Figura 2): la valoración de los viajeros de una a cinco estrellas que sintetiza sus sentimientos y se puede contabilizar fácilmente, y sus propias palabras en el título del OTR que se pueden categorizar para proceder a su cuantificación. Para obtener una nota media del componente evaluativo basada en las estrellas o burbujas, se les ha dado una puntuación numérica entre cero y diez (Excelente: 10; Muy bueno: 7,5; Normal: 5; Malo: 2,5; y Pésimo: 0 puntos). Una categoría dicotómica simple para medir el componente afectivo puede ser la compuesta de buenas y malas impresiones y recomendaciones. Se trata de recopilar palabras (adjetivos calificativos positivos y negativos) o grupos de palabras que expresen dichos sentimientos. Por ejemplo, 'espectacular' y 'no te lo pierdas' expresan sentimientos positivos, mientras que 'decepcionante' y 'no vale la pena' expresan sentimientos negativos. Así como el componente cognitivo permite deducir, por ejemplo, las atracciones más visitadas, los componentes evaluativo y afectivo permite saber cuáles son las más apreciadas.

Para el análisis sintáctico de contenido se preparan dos listas, una con los grupos de dos o más palabras que tienen un significado propio como conjunto (p. e. 'Sagrada Familia' o 'No te lo pierdas'), y otra con las palabras no significativas en el caso de estudio como la mayoría de adverbios, conjunciones, determinantes, preposiciones y pronombres. En la Figura 4, se puede ver el pseudocódigo simplificado del algoritmo usado para el análisis cuantitativo del texto de los títulos de los OTRs.

```

Cargar compWord; // Palabra compuesta (lista de)
Cargar stopWord; // Palabra no significativa (lista de)
Cargar textWord; // Texto para analizar cuantitativamente
Crear result; // Tabla para almacenar la frecuencia y densidad
// de las palabras clave (keywords)

for each compWord do
{
  if exists compWord in textWord then
  {
    Contar apariciones del compWord en textWord;
    Añadir compWord a result con su frecuencia;
    Suprimir apariciones del compWord en textWord;
  }
}
for each word in textWord do
{
  if not exists word in stopWord then // pasa a ser un keyword
  {
    if exists keyword in result incrementar su frecuencia;
    else añadir keyword a result con frecuencia uno;
  }
}
Imprimir result;

```

Figura 4. Algoritmo usado para el análisis cuantitativo del contenido de los OTRs. Fuente: Elaboración propia.

## 4. Resultados

La clasificación de los OTRs (Tabla 2) por marcas turísticas territoriales y años permite comprobar las tendencias de la demanda en el espacio y en el tiempo, así como hacer estudios pormenorizados de la especialización de la imagen desde la perspectiva de los visitantes (Mariné-Roig & Anton Clavé, 2016b). A grandes rasgos, se pueden apreciar dos tendencias evidentes: el espectacular crecimiento de la cantidad de OTRs y la preponderancia de la marca Barcelona seguida de lejos por las tres marcas litorales (Costa Brava,



Costa Daurada y Costa Barcelona). Por otra parte, es evidente la estacionalidad (Figura 3) con el máximo número de OTRs durante el tercer trimestre de todos los años que coincide aproximadamente con las vacaciones de verano.

| Atracción o servicio   | Marca | OTRs  | Nota | ***** | **** | *** | **  | *   |
|------------------------|-------|-------|------|-------|------|-----|-----|-----|
| <i>Sagrada Familia</i> | Barna | 12073 | 9,26 | 9466  | 1916 | 482 | 119 | 90  |
| <i>Park Güell</i>      | Barna | 5053  | 8,21 | 2684  | 1488 | 610 | 170 | 101 |
| Gothic Quarter         | Barna | 4611  | 8,95 | 2920  | 1486 | 172 | 18  | 15  |
| <i>Casa Batlló</i>     | Barna | 4515  | 8,98 | 3151  | 1020 | 250 | 57  | 37  |
| PortAventura           | cDaur | 3861  | 7,41 | 1667  | 1221 | 448 | 210 | 315 |
| Camp Nou               | Barna | 3603  | 8,62 | 2308  | 824  | 324 | 67  | 80  |
| Sandemans New          | Barna | 3585  | 9,64 | 3168  | 353  | 40  | 13  | 11  |
| Magic Fountain         | Barna | 2565  | 8,83 | 1681  | 650  | 172 | 38  | 24  |
| Las Ramblas            | Barna | 2423  | 7,66 | 1025  | 806  | 392 | 122 | 78  |
| Santa Maria del Mar    | Barna | 2396  | 9,25 | 1756  | 568  | 64  | 5   | 3   |
| Palau de la Musica     | Barna | 2321  | 9,35 | 1836  | 402  | 56  | 17  | 10  |
| Passeig de Gracia      | Barna | 2114  | 8,78 | 1231  | 749  | 126 | 5   | 3   |
| Mercat de la Boqueria  | Barna | 1808  | 8,52 | 1060  | 517  | 166 | 38  | 27  |
| <i>La Pedrera</i>      | Barna | 1791  | 8,67 | 1081  | 528  | 139 | 27  | 16  |
| Aerobus Barcelona      | Barna | 1352  | 8,63 | 820   | 378  | 117 | 21  | 16  |
| Chicken Banana         | Barna | 1246  | 9,46 | 1018  | 205  | 12  | 5   | 6   |
| Parc de Montjuic       | Barna | 1107  | 8,58 | 572   | 451  | 78  | 4   | 2   |
| CosmoCaixa             | Barna | 971   | 8,85 | 602   | 302  | 59  | 5   | 3   |
| <i>Palau Güell</i>     | Barna | 852   | 8,85 | 538   | 259  | 39  | 9   | 7   |
| Barceloneta Beach      | Barna | 824   | 7,59 | 299   | 332  | 134 | 43  | 16  |
| Dali Theatre Museum    | cBrav | 811   | 8,49 | 479   | 225  | 70  | 23  | 14  |
| El Born                | Barna | 788   | 8,68 | 449   | 278  | 49  | 8   | 4   |
| Museu MNAC             | Barna | 767   | 8,92 | 497   | 225  | 32  | 11  | 2   |
| Parc de la Ciutadella  | Barna | 766   | 8,40 | 376   | 302  | 80  | 4   | 4   |
| Tibidabo               | Barna | 743   | 8,37 | 364   | 293  | 72  | 9   | 5   |
| Montserrat Monastery   | pBarc | 735   | 8,99 | 499   | 189  | 36  | 8   | 3   |
| Sant Pau Recinte       | Barna | 708   | 9,23 | 518   | 169  | 17  | 1   | 3   |
| Catedral Barcelona     | Barna | 688   | 8,32 | 330   | 288  | 49  | 8   | 13  |
| Placa de Catalunya     | Barna | 658   | 7,77 | 246   | 266  | 124 | 16  | 6   |
| Castell Tossa de Mar   | cBrav | 658   | 8,97 | 425   | 202  | 26  | 2   | 3   |

Nota: \*\*\*\*\*: valoraciones más positivas; \*: valoraciones más negativas

Tabla 3. Primeras 30 “cosas que hacer” en Cataluña (frecuencia y valoración). Fuente: Muestra de 150.000 OTRs escritos en español en TripAdvisor entre 2012 y 2016.

La Tabla 3 permite conocer las atracciones más visitadas y las mejor valoradas. En cuanto al componente cognitivo de la imagen, la Tabla 3 nos da indicios de los elementos y atributos (atracciones, lugares y servicios) principales que conforman la imagen ‘objetiva’ o ‘racional’ del destino, los más representativos entre los usuarios. Vemos que estos elementos están muy ligados al patrimonio tangible, especialmente de Barcelona. Sin lugar a dudas, destacan las obras maestras del arquitecto Antoni Gaudí (La Sagrada Familia, Parc Güell, Casa Batlló, Casa Milà - La Pedrera y Palau Güell) declaradas Patrimonio de la Humanidad (UNESCO, 2005). También vemos pero zonas de Barcelona como ‘Las Ramblas’ o ‘Barrio Gótico’ y elementos de otros tipos como ‘Camp Nou’ relacionado con el deporte o ‘Port Aventura’, un parque temático de ocio. En cuanto al componente evaluativo, las valoraciones de los usuarios nos indican en parte la satisfacción con su experiencia en estos lugares y atracciones (5\*: 89.037; 4\*: 42.061; 3\*: 12.790; 2\*: 3.013; y 1\*: 3.099). Los resultados muestran que en todas las 30 atracciones principales la puntuación mayoritaria es la más positiva (5\*) y que, en casi todos los casos, esta puntuación máxima es otorgada en más del 50% de las reseñas. Por su elevada puntuación, destacan un servicio de excursiones (tours) gratuitas (Sandemans New Europe) y un centro recreativo de juegos de escape (Chicken Banana). Las opiniones negativas son realmente minoritarias en todas las atracciones. Ello indica que el componente evaluativo de la imagen es en general muy positivo respecto a los elementos principales de la imagen del destino. Por lo que respecta al componente espacial de la imagen, la Tabla 3 nos muestra una gran concentración de los elementos principales del destino en su capital, Barcelona. Fuera del ámbito de Barcelona, sobresale el parque temático Port Aventura (Costa Daurada) y, en

menor medida, destacan el Monasterio de Montserrat (Paisajes Barcelona) y el Teatro-Museo Dalí (Costa Brava).

| Keyword         | Count | Percent | Keyword           | Count | Percent |
|-----------------|-------|---------|-------------------|-------|---------|
| 1 barcelona     | 6068  | 0,79    | 11 bonito         | 2930  | 0,38    |
| 2 visita        | 5921  | 0,78    | 12 buen           | 2752  | 0,36    |
| 3 excelente     | 4492  | 0,59    | 13 interesante    | 2736  | 0,36    |
| 4 lugar         | 4065  | 0,53    | 14 recomendable   | 2542  | 0,33    |
| 5 paseo         | 3592  | 0,47    | 15 gran           | 2521  | 0,33    |
| 6 impresionante | 3508  | 0,46    | 16 buena          | 2486  | 0,33    |
| 7 mejor         | 3461  | 0,45    | 17 vistas         | 2355  | 0,31    |
| 8 experiencia   | 3392  | 0,44    | 18 parque         | 2232  | 0,29    |
| 9 genial        | 3179  | 0,42    | 19 precioso       | 2225  | 0,29    |
| 10 espectacular | 3036  | 0,40    | 20 imprescindible | 2133  | 0,28    |

Tabla 4. Las 20 palabras clave más frecuentes en los títulos de los 150.000 OTRs. Fuente: Muestra aleatoria de OTRs de TripAdvisor escritos en español entre 2012 y 2016.

La Tabla 4 recoge la frecuencia de todas las palabras significativas (21.868 palabras distintas) que hay en los títulos de los 150.000 OTRs considerados (763.682 palabras en total). También indica su densidad en porcentaje del total de palabras (incluyendo las no significativas). Por una parte, destaca Barcelona porque es el destino más conocido y visitado de Cataluña con mucha diferencia sobre los demás; como se puede comprobar en la Tabla 2, más del 60% de OTRs están asociados con la marca Barcelona. Esto refuerza el hecho de que el componente espacial de la imagen está altamente concentrado geográficamente en la metrópolis. Por la otra, un gran número de palabras expresan sentimientos positivos que vienen a corroborar las valoraciones positivas recogidas en la Tabla 3 y nos muestran de nuevo un componente afectivo de la imagen altamente positivo expresado por los turistas entorno a las atracciones principales del destino, con palabras muy frecuentes como 'excelente', 'impresionante', 'genial', 'espectacular', 'bonito', 'interesante', 'recomendable', 'precioso' e 'imprescindible'.

## 5. Conclusiones y futuras investigaciones

En este estudio se ha reafirmado la validez de los OTRs como fuente de información para extraer los componentes cognitivo, evaluativo y afectivo de la TDI percibida (y transmitida) por los viajeros, situándola en el espacio y en tiempo, así como la importancia de los motores de búsqueda de Internet en la localización y presentación de los OTRs. El análisis masivo y sistemático de un gran volumen de información de primera mano, generada desinteresadamente por los usuarios, da fiabilidad a los resultados obtenidos y permite deducir la imagen global a partir de la suma de las imágenes percibidas individualmente. El análisis se ha centrado en los metadatos de la página web que procesan los buscadores y navegadores de Internet para hacer la presentación de los OTRs a requerimiento de los usuarios. La imagen resultante, percibida desde la perspectiva de la demanda, es relevante y puede ser de gran utilidad a las organizaciones gestoras del destino turístico para optimizar la oferta. Para el caso de Cataluña, se han podido identificar los principales elementos del componente cognitivo de la imagen, donde destaca el patrimonio tangible de Barcelona (especialmente del arquitecto Gaudí). En el caso de los componentes evaluativo y afectivo, el análisis ha permitido constatar el alto grado de satisfacción de los turistas con las principales atracciones y el uso masivo de adjetivos positivos. Finalmente, en referencia a la dimensión espacial se ha detectado una gran concentración de la imagen en la capital del destino, seguida a distancia por algunas atracciones situadas en zonas costeras. En cuanto a la dimensión temporal, destaca el crecimiento anual de la cantidad de OTRs (Tabla 2), superior al de la afluencia turística, y la estacionalidad registrándose máximos en el tercer trimestre (Figura 3).

El método propuesto, basado en el procesamiento de metadatos masivos estructurados, de libre acceso a través de Internet, permite profundizar en muchos aspectos de la imagen percibida y transmitida, sin necesidad de realizar o replicar costosas encuestas. Por ejemplo, se puede analizar la especialización de la TDI de Cataluña en el espacio y su evolución en el tiempo, tras los atentados de mediados de Agosto de 2017 en



Barcelona y Cambrils, o en épocas de incertidumbre política y jurídica. La principal limitación del método es que no todos los OTRs están indexados en los motores de búsqueda, pero los sitios web que los alojan, como TripAdvisor, gozan de gran popularidad y visibilidad y, por ello, aparecen sus contenidos en las primeras posiciones de las búsquedas relacionadas con los atributos, atracciones y servicios de los destinos turísticos.

## Agradecimientos

Este estudio es parte de un proyecto más amplio titulado “Efectos de la transformación de las pautas de movilidad global en la evolución de los destinos turísticos (MOVETUR)” financiado por el Ministerio de Economía y Competitividad español (CSO2014-51785-R).

### Cómo citar este artículo / How to cite this paper

Mariné-Roig, E.; Anton, S. (2017). Imagen de un destino a partir de reseñas de turistas en buscadores de Internet. *International Journal of Information Systems and Tourism (IJIST)*, 2(2), 41-52. (www.ijist-tourism.com)

## Referencias

- ACT. (2017). Dossier de prensa 2017: Turismo. Barcelona: Agencia Catalana de Turismo.
- Baka, V. (2016). The becoming of user-generated reviews: Looking at the past to understand the future of managing reputation in the travel sector. *Tourism Management*, 53, 148-162. <http://doi.org/10.1016/j.tourman.2015.09.004>
- Baloglu, S.; McCleary, K. W. (1999). A model of destination image formation. *Annals of Tourism Research*, 26(4), 868-897. doi: [http://doi.org/10.1016/S0160-7383\(99\)00030-4](http://doi.org/10.1016/S0160-7383(99)00030-4)
- Beerli, A.; Martín, J. D. (2004). Factors influencing destination image. *Annals of Tourism Research*, 31(3), 657-681. doi: <http://doi.org/10.1016/j.annals.2004.01.010>
- Camprubí, R.; Coromina, L. (2016). Content analysis in tourism research. *Tourism Management Perspectives*, 18, 134-140. doi: <http://doi.org/10.1016/j.tmp.2016.03.002>
- Chon, K.-S. (1990). The role of destination image in tourism : A review and discussion. *The Tourist Review*, 45(2), 2-9. doi: <http://doi.org/10.1108/eb058040>
- Crompton, J. L. (1979). An assessment of the image of Mexico as a vacation destination and the influence of geographical location upon that image. *Journal of Travel Research*, 17, 18-23. doi: <http://doi.org/10.1177/004728757901700404>
- Datzira-Masip, J.; Poluzzi, A. (2014). Brand architecture management: The case of four tourist destinations in Catalonia. *Journal of Destination Marketing and Management*, 3(1), 48-58. doi: <http://doi.org/10.1016/j.jdmm.2013.12.006>
- de Borda, J. C. (1781). Mémoire sur les élections au scrutin. In *Mémoire de l'Académie Royale* (pp. 657-665). Paris: Histoire de l'Académie des Sciences.
- de Rosa, A. S.; Bocci, E.; Dryjanska, L. (2017). Social representations of the European capitals and destination e-branding via multi-channel web communication. *Journal of Destination Marketing and Management*, (in press), 1-16. doi: <http://doi.org/10.1016/j.jdmm.2017.05.004>
- Eurobarometer (2016). Flash Eurobarometer 432: Preferences of Europeans towards tourism.
- Eurostat (2017). Tourism. In *Eurostat regional yearbook* (pp. 181-198). Luxembourg: Publications Office of the European Union.
- Gartner, W. C. (1993). Image formation process. *Journal of Travel & Tourism Marketing*, 2(2-3), 191-215. doi: [http://doi.org/10.1300/J073v02n02\\_12](http://doi.org/10.1300/J073v02n02_12)
- Huertas, A.; Marine-Roig, E. (2016). User reactions to destination brand contents in social media. *Information Technology & Tourism*, 15(4), 291-315. doi: <http://doi.org/10.1007/s40558-015-0045-9>
- Kim, D.; Perdue, R. R. (2011). The influence of image on destination attractiveness. *Journal of Travel & Tourism Marketing*, 28(3), 225-239. doi: <http://doi.org/10.1080/10548408.2011.562850>
- Kim, H. (2015). Comparing use of the Internet for trip planning between leisure and business travelers. *International Journal of Business, Humanities and Technology*, 5(4), 1-9.
- Li, J.; Ali, F.; Kim, W. G. (2015). Reexamination of the role of destination image in tourism: An updated literature review. *E-Review of Tourism Research*, 12(3-4), 191-209.
- Liu, Y.; Teichert, T.; Rossi, M.; Li, H.; Hu, F. (2017). Big data for big insights: Investigating language-specific drivers of hotel satisfaction with 412,784 user-generated reviews. *Tourism Management*, 59, 554-563. doi: <http://doi.org/10.1016/j.tourman.2016.08.012>
- Llodrà-Riera, I.; Martínez-Ruiz, M. P.; Jiménez-Zarco, A. I.; Izquierdo-Yusta, A. (2015). A multidimensional analysis of the information sources construct and its relevance for destination image formation. *Tourism Management*, 48, 319-328. doi: <http://doi.org/10.1016/j.tourman.2014.11.012>
- Mariné-Roig, E. (2014). A webometric analysis of travel blogs and review hosting: The case of Catalonia. *Journal of Travel & Tourism*
- Mariné-Roig, E.; Anton, S. (2017). Imagen de un destino a partir de reseñas de turistas en buscadores de Internet. *International Journal of Information Systems and Tourism (IJIST)*, 2(2), 41-52.

- Marketing, 31(3), 381-396. doi: <http://doi.org/10.1080/10548408.2013.877413>
- Mariné-Roig, E. (2015). Identity and authenticity in destination image construction. *Anatolia: An International Journal of Tourism and Hospitality Research*, 26(4), 574-587. doi: <http://doi.org/10.1080/13032917.2015.1040814>
- Mariné-Roig, E. (2017). Measuring destination image through travel reviews in search engines. *Sustainability*, 9(8), 1-18. doi: <http://doi.org/10.3390/su9081425>
- Mariné-Roig, E.; Anton Clavé, S. (2015). A method for analysing large-scale UGC data for tourism: Application to the case of Catalonia. In *Information and Communication Technologies in Tourism* (pp. 3-17). doi: [http://doi.org/10.1007/978-3-319-14343-9\\_1](http://doi.org/10.1007/978-3-319-14343-9_1)
- Mariné-Roig, E.; Anton Clavé, S. (2016a). A detailed method for destination image analysis using user-generated content. *Information Technology & Tourism*, 15(4), 341-364. doi: <http://doi.org/10.1007/s40558-015-0040-1>
- Mariné-Roig, E.; Anton Clavé, S. (2016b). Perceived image specialisation in multiscale tourism destinations. *Journal of Destination Marketing & Management*, 5(3), 202-213. doi: <http://doi.org/http://dx.doi.org/10.1016/j.jdmm.2015.12.007>
- Neuendorf, K. A. (2017). *The content analysis guidebook* (2n ed.). London, UK: SAGE Publications.
- OdTiMP (2017). Evolución anual de la oferta y la ocupación de los establecimientos hoteleros. (<http://observatoritrebball.gencat.cat>)
- Pan, B.; Fesenmaier, D. R. (2006). Online information search. Vacation planning process. *Annals of Tourism Research*, 33(3), 809-832. doi: <http://doi.org/10.1016/j.annals.2006.03.006>
- Pan, B.; Xiang, Z.; Tierney, H.; Fesenmaier, D. R.; Law, R. (2010). Assessing the dynamics of search results in Google. In *Information and Communication Technologies in Tourism 2010* (pp. 405-416). doi: [http://doi.org/10.1007/978-3-211-99407-8\\_34](http://doi.org/10.1007/978-3-211-99407-8_34)
- Pantano, E.; Priporas, C. V.; Stylos, N. (2017). "You will like it!" using open data to predict tourists' response to a tourist attraction. *Tourism Management*, 60, 430-438. doi: <http://doi.org/10.1016/j.tourman.2016.12.020>
- Pocock, D.; Hudson, R. (1978). *Images of the urban environment*. London, UK: Macmillan.
- Rapoport, A. (1977). *Human aspects of urban form*. Oxford, UK: Pergamon Press.
- Ríos, M. Á.; Ortega, F. J.; Matilla, M. (2016). La estancia perfecta en Hoteles de 4 y 5 estrellas de Sevilla a través del análisis de los comentarios en TripAdvisor. Determinación de los principales ítems. *International Journal of Information Systems and Tourism (IJIST)*, 1(1), 8-25.
- Son, A. (2005). The measurement of tourist destination image: applying a sketch map technique. *International Journal of Tourism Research*, 7(4-5), 279-294. doi: <http://doi.org/10.1002/jtr.532>
- Stepchenkova, S. (2012). Content analysis. In *Handbook of research methods in Tourism: Quantitative and qualitative approaches* (pp. 443-458). Cheltenham, UK: Edward Elgar Publishing.
- TripAdvisor (2017). Cosas que hacer en Cataluña. (<https://www.tripadvisor.es/Attractions-g187496-Activities-Catalonia.html>)
- UNESCO (2005). Works of Antoni Gaudi. (<http://whc.unesco.org/en/list/320>)
- VisitBritain (2017). Researching and Planning: Foresight, (150). London, UK. ([https://www.visitbritain.org/sites/default/files/vb-corporate/Documents-Library/documents/foresight\\_150\\_-\\_researching\\_and\\_planning.pdf](https://www.visitbritain.org/sites/default/files/vb-corporate/Documents-Library/documents/foresight_150_-_researching_and_planning.pdf))
- W3Schools (2017). HTML Tutorial. (<http://www.w3schools.com/html/>)
- Xiang, Z.; Wang, D.; O'Leary, J. T.; Fesenmaier, D. R. (2015). Adapting to the Internet: Trends in travelers' use of the Web for trip planning. *Journal of Travel Research*, 54(4), 511-527. doi: <http://doi.org/10.1177/0047287514522883>
- Xiang, Z.; Wober, K.; Fesenmaier, D. R. (2008). Representation of the online tourism domain in search engines. *Journal of Travel Research*, 47(2), 137-150. doi: <http://doi.org/10.1177/0047287508321193>

