



META-ANÁLISIS MASIVO DE OPINIONES EN LÍNEA SOBRE VIAJES

Estela Marine-Roig (estela.marine@aegern.udl.cat)
Salvador Anton Clavé

Universitat de Lleida
Universitat Rovira i Virgili

Resumen

En los últimos años han crecido enormemente en Internet los contenidos generados por los consumidores y usuarios (UGC) en paralelo con el auge de los medios de comunicación sociales. El UGC, como fuente de información desinteresada, goza de una gran credibilidad entre los demás usuarios. Este incremento del UGC se ha notado particularmente en el dominio del turismo a través de los blogs de viaje y las opiniones en línea acerca de viajes (OTRs) que reflejan la imagen percibida por los visitantes de un destino. A través del llamado boca a boca electrónico (eWOM), dicha imagen se difunde y contribuye a la construcción de la imagen global del destino turístico (TDI). Debido al gran volumen de información en línea disponible, los motores de búsqueda de Internet juegan un papel crucial en la localización y presentación de los relatos, opiniones, comentarios y valoraciones de los viajeros. El objetivo de este artículo es proponer un método para analizar la imagen percibida (y transmitida) tal como se presenta en los buscadores, haciendo hincapié en los componentes cognitivo, afectivo y espacial de la TDI. Para ello, se realiza un análisis cuantitativo de algunos metadatos procesados por los buscadores, a partir de una muestra aleatoria de 88.000 OTRs alojados en TripAdvisor, sobre “cosas que hacer” en Cataluña, escritos en español por turistas entre 2011 y 2015. El análisis masivo de datos semiestructurados generados por usuarios da fiabilidad a los resultados obtenidos, los cuales pueden ser de gran utilidad para optimizar la oferta turística.

Palabras clave: Imagen de un destino; datos masivos; minería de datos; metadatos; Catalunya; Catalonia

MASSIVE META-ANALYSIS OF ONLINE TRAVEL REVIEWS

Abstract

In recent years online consumer and user-generated contents (UGC) have grown enormously in parallel with the rise of social media. UGC, as a disinterested source of information, has great credibility among users. The growth of UGC has been especially remarkable in the field of tourism with travel blogs and online travel reviews (OTRs) that reflect visitors' perceived image of a destination. Through the so-called electronic word-of-mouth (eWOM), tourist image is spread out and this contributes to building the global tourist destination image (TDI). Due to the large volume of information available online, the Internet search engines play a crucial role in locating and presenting stories, opinions, reviews and ratings of travellers. The aim of this paper is to propose a method for the analysis of the perceived (and transmitted) image as presented in search engines, emphasizing the cognitive, affective and spatial components of the



TDI. To this end, a quantitative analysis of some metadata processed by search engines, from a random sample of 88,000 TripAdvisor OTRs on "things to do" in Catalonia, written in Spanish by tourists between 2011 and 2015 is conducted. The analysis is conducted on massive semi-structured data generated by users, which gives reliability to the obtained results, and that can be useful for tourism supply optimization.

Keywords: Destination image; big data; data mining; metadata; Catalunya; Catalonia

1. INTRODUCCIÓN

El uso de Internet se ha incrementado en los últimos años y sigue creciendo. Aproximadamente, el 75% de los europeos usa regularmente los servicios de Internet (Eurostat, 2015). 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 fuente de información más usada para planificar las vacaciones fue el boca a boca (WOM): recomendaciones de parientes, amigos y compañeros (51%); seguido del boca a boca electrónico (eWOM): sitios web que recogen y presentan comentarios, opiniones y valoraciones de viajeros (34%); muy por encima de las fuentes de imagen turística inducida: sitios web gestionados por los proveedores de servicios o por los destinos turísticos (17%). 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 muy superior al de los europeos y se sitúa en torno al 85%, seguido del WOM con un 42% (Kim, 2015; Xiang, Wang, O'Leary, & Fesenmaier, 2015). Estos resultados son consistentes con los obtenidos por Llodrà-Riera, Martínez-Ruiz, Jiménez-Zarco, & Izquierdo-Yusta (2015) en un estudio sobre visitantes de Mallorca. En los casos anteriores, no se ha tenido en cuenta la previa experiencia que ocupa también un lugar destacado en la planificación de los viajes.

El eWOM se basa principalmente en contenidos generados por los usuarios (UGC) 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 (Marine-Roig, 2010). En el campo del turismo, mediante los diarios de viaje, 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) desde una perspectiva global (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 del UGC, se puede citar que, en Enero de 2015, Booking afirmaba tener 43 millones de OTRs verificados y TripAdvisor más de 200 millones de comentarios y opiniones y, en Enero de 2016, afirmaban haber superado los 75 y 350 millones de OTRs, respectivamente.

En relación a los sitios web que resultan más útiles a los usuarios para recabar información sobre un destino, destacan en primera posición los buscadores de Internet (Llodrà-Riera et al., 2015), mientras que, para planificar un viaje, pasan a tercera posición en el caso de usuarios norteamericanos, detrás de las agencias de viaje en línea y de los portales de proveedores de servicios (Kim, 2015; Xiang et al., 2015).

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 que procesan los motores de búsqueda, a partir de una muestra aleatoria de 88.000 OTRs alojados en TripAdvisor, escritos en español por visitantes de Cataluña entre 2011 y 2015, y se obtienen resultados significativos en tres dimensiones de la TDI (cognitiva, afectiva y espacial).

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 motores de búsqueda 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 (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 & 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). Como precedente de la TDI evaluativa, Pocock & 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 evaluativo (el significado adjunto a, o evocado por, la forma física). El aspecto espacial de la imagen designativa fue estudiado por Son (2005) mediante la obtención de croquis que, a su



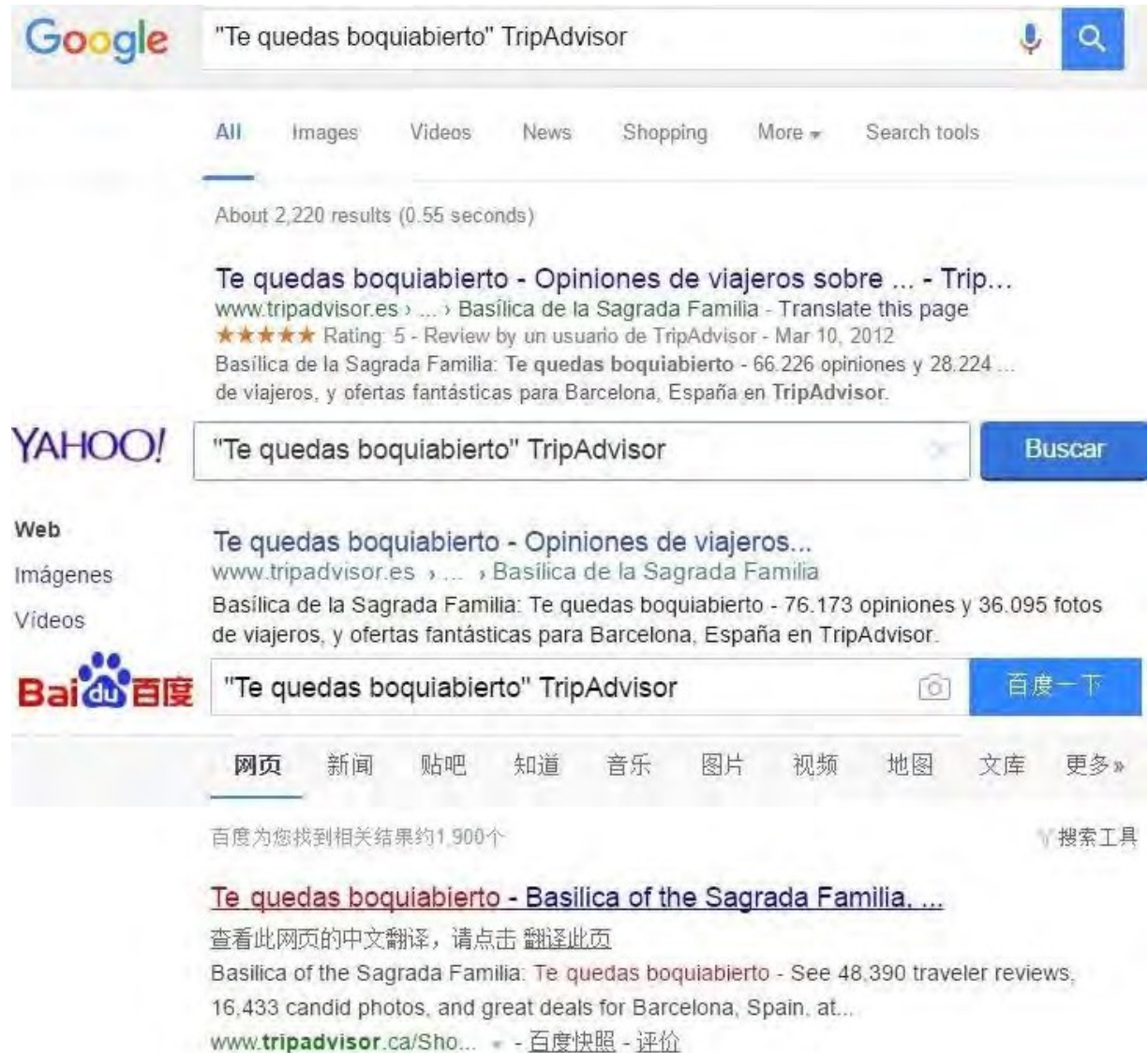
vez, son usados para medir mapas mentales, y por Marine-Roig & Anton Clavé (2016b) mediante coeficientes espaciales para elucidar la especialización de la imagen en destinos multiescalares.

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, Xiang, Tierney, Fesenmaier, & Law, 2010). Pan & 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 (Marine-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 1 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 1. 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, 2016) contiene metadatos sobre un documento HTML (*HyperText Markup Language*). 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, 2016). A pesar de que la información contenida en la Tabla 1 y en la Figura 1 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 a menudo.



Tabla 1. Contenido (30-05-2016) de las meta-etiquetas HTML más destacadas (ver Figura 1)

Meta name	Content
<i>title</i>	Te quedas boquiabierto - Opiniones de viajeros sobre Basílica de la Sagrada Familia, Barcelona - TripAdvisor
<i>hyperlink</i>	https://www.tripadvisor.es/ShowUserReviews-g187497-d190166-r125908839-Basilica_of_the_Sagrada_Familia-Barcelona_Catalonia.html
<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

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 es una adaptación del método para la recogida masiva, clasificación y depuración de blogs y reseñas de viaje descrito en Marine-Roig & Anton Clavé (2014), y detallado en Marine-Roig & 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, 2015), cuya afluencia de visitantes se ha mantenido e, incluso, incrementado en los años considerados (2011-2015). Según OdEIO (2016), la ocupación de los alojamientos turísticos catalanes (hoteles, campings, casas rurales, apartamentos turísticos y albergues) ha experimentado la siguiente evolución anual (en millones de viajeros y pernoctaciones): 2011 (20,1 y 69,0) y 2015 (22,2 y 75,3). De entre los destinos catalanes, destaca Barcelona que en 2015 ha alojado 8.303.649 viajeros (17.656.329 pernoctaciones). Aproximadamente el 75 % de todos estos viajeros procede del extranjero, habiéndose batido récords históricos en 2015 de llegada de turistas extranjeros de primer destino y gasto total (17,4 y 15.735,5 millones respectivamente).

Cataluña es un destino multiescalar (Marine-Roig & Anton Clavé, 2016b) que tiene dividido el territorio en nueve marcas turísticas: 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).

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 12 portales con información sobre Cataluña en la época considerada: GetJealous.com (GJ), MiNube.com (MN), MyTripJournal.com (MT), StaTravel.com (ST), TravBuddy.com (TY), TravelBlog.org (TB), TravelJournals.net (TJ), TravellersPoint.com (TS), TravelPod.com (TP), TripAdvisor.com (TA), Venere.com (VN), y VirtualTourist.com (VT). Para seleccionar la fuente más idónea se construye un ranking aplicando la fórmula ponderada (Marine-Roig, 2014):

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

donde \underline{B} corresponde al método de ordenación de Borda, \underline{V} visibilidad de la web (cantidad y calidad de enlaces entrantes), \underline{P} su popularidad (visitas que recibe y tráfico en general) y \underline{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 \underline{V} , \underline{P} y \underline{S} , la función \underline{B} asigna una puntuación a cada candidato (c) consistente en el número de candidatos situados debajo de \underline{c} en \underline{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 TripAdvisor con mucha diferencia sobre las demás webs. Esta selección coincide con Baka (2016) que considera a TripAdvisor la mayor fuente mundial de UGC en el dominio del turismo. Descartando los centenares de miles de OTRs sobre hospedaje y restauración, TripAdvisor presenta casi medio millón de OTRs sobre “cosas que hacer” (*Things to Do*) en Cataluña entre 2011 y 2015, de los cuales unos 200.000 están escritos en inglés y más de 88.000 en español. 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

Observando la Figura 1 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 hipervínculo (*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 hipervínculo (Tabla 1) aporta muchos datos que facilitan el proceso de análisis de la información asociada: protocolo (*https*: HTTP seguro), 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 (*Basilica 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 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*) 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 & Anton Clavé (2014), se substituyen todos los caracteres especiales (los superiores al ASCII 127) por el correspondiente carácter ISO 8859-15 (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.

3.4. Análisis cuantitativo

El análisis se centra en tres dimensiones de la TDI: espacial, cognitiva y afectiva.

- *Componente espacial.* Para poder estudiar el componente 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).

Tabla 2. Los 88.000 OTRs sobre "cosas que hacer" en Cataluña por marca turística y año

	Barna	cBarc	cBrav	cDaur	pBarc	Pyren	tEbre	tLlei	vAran
2011	864	35	96	160	17	29	0	5	5
2012	5.952	386	861	778	152	197	10	46	16
2013	7.477	600	1.105	972	263	452	30	113	29
2014	12.737	1.406	1.940	1.689	454	813	79	174	119
2015	30.792	3.534	5.071	4.462	1.008	1.766	424	570	312

Fuente: Muestra aleatoria de OTRs escritos en español en TripAdvisor entre 2011 y 2015



- **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.
- **Componente afectivo.** Por lo que respecta al componente afectivo, se dispone de dos fuentes de información: la valoración de los viajeros (Figura 1) de una a cinco estrellas (Excelente, Muy bueno, Normal, Malo y Pésimo) 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. Una categoría dicotómica simple 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, el componente afectivo permite saber cuáles son las mejor valoradas.

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 2, se puede ver el pseudocódigo simplificado del algoritmo usado para el análisis cuantitativo del texto de los títulos de los OTRs.

Figura 2. Algoritmo usado para el análisis cuantitativo del contenido 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;
```



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 (Marine-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).

Tabla 3. Primeras 25 “cosas que hacer” en Cataluña (frecuencia y valoración)

Atracción o servicio	OTRs	*****	****	***	**	*
1 Basilica of the Sagrada Familia	7.408	76 ,43	17 ,20	4 ,78	1 ,05	0 ,54
2 Park Guell	3.154	53 ,82	31 ,03	11 ,42	2 ,55	1 ,19
3 Gothic Quarter Barri Gotic	2.972	63 ,59	31 ,89	3 ,96	0 ,40	0 ,16
4 Casa Batllo	2.838	72 ,27	21 ,43	4 ,83	0 ,98	0 ,49
5 PortAventura Park	2.770	54 ,63	26 ,88	9 ,98	3 ,83	4 ,68
6 Camp Nou	2.252	61 ,97	26 ,78	8 ,48	1 ,84	0 ,93
7 The Basilica of Santa Maria del Mar	1.650	62 ,81	30 ,51	5 ,87	0 ,58	0 ,23
8 Magic Fountain Font Magica	1.623	64 ,14	24 ,68	7 ,74	2 ,00	1 ,44
9 Palau de la Musica Orfeo Catalana	1.577	75 ,67	20 ,10	3 ,45	0 ,47	0 ,32
10 Passeig de Gracia	1.498	54 ,26	37 ,87	7 ,23	0 ,46	0 ,19
11 Sandemans New Barcelona Tours	1.428	87 ,57	9 ,82	1 ,30	0 ,71	0 ,60
12 Las Ramblas	1.251	36 ,66	33 ,04	19 ,78	7 ,09	3 ,44
13 La Pedrera	1.124	56 ,26	30 ,80	9 ,88	2 ,37	0 ,68
14 Mercat de Sant Josep de la Boqueria	868	66 ,78	26 ,42	5 ,08	1 ,12	0 ,59
15 Parc de Montjuic	783	52 ,27	37 ,99	8 ,17	1 ,05	0 ,51
16 CosmoCaixa Barcelona	749	66 ,93	25 ,82	5 ,82	0 ,95	0 ,48
17 Palau Guell	717	61 ,51	28 ,95	7 ,77	1 ,17	0 ,60
18 El Born	598	63 ,27	31 ,73	4 ,07	0 ,69	0 ,23
19 Aerobus	592	72 ,70	21 ,70	4 ,00	1 ,01	0 ,58
20 Ciutat Vella	585	60 ,44	34 ,26	4 ,32	0 ,66	0 ,33
21 Catedral de Barcelona	581	55 ,83	35 ,13	7 ,73	0 ,76	0 ,55
22 Montserrat Monastery	576	73 ,25	21 ,29	4 ,11	0 ,98	0 ,38
23 Dali Theatre Museum	563	67 ,05	22 ,88	6 ,85	2 ,15	1 ,07
24 Museu MNAC	552	63 ,29	29 ,05	6 ,40	1 ,05	0 ,22
25 Tibidabo Mountain	531	57 ,60	31 ,34	8 ,84	1 ,68	0 ,55

Nota: *****: valoraciones más positivas en %; *: valoraciones más negativas en %

Fuente: Muestra de 88.000 OTRs escritos en español en TripAdvisor entre 2011 y 2015

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, La Pedrera y Palau Güell) declaradas Patrimonio de la Humanidad (UNESCO, 2005). También vemos pero zonas de Barcelona como Las Ramblas o Ciutat Vella y elementos de otros tipos como Camp Nou relacionado con el deporte o Port Aventura, un parque temático de ocio. En cuanto al componente afectivo, las valoraciones de los usuarios nos indican en parte la satisfacción con su experiencia en estos lugares y atracciones. Los resultados muestran que en todas las 25 atracciones principales la puntuación mayoritaria es la más positiva (****), y que en todas a excepción de Las Ramblas, esta puntuación máxima es otorgada en más del 50% de las reseñas. Las opiniones negativas son realmente minoritarias en todas las atracciones. Ello indica que el componente afectivo 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). Como ya se ha mencionado, entre las atracciones principales del destino también se encuentran zonas y barrios concretos de Barcelona.

Tabla 4. Las 20 palabras clave más frecuentes en los títulos de los 88.000 OTRs

Keyword	Count	Percent	Keyword	Count	Percent
1 muy	6.524	1,46	11 espectacular	1.700	0,38
2 barcelona	3.736	0,84	12 bonito	1.655	0,37
3 visita	3.155	0,71	13 interesante	1.612	0,36
4 excelente	2.645	0,59	14 buen	1.583	0,35
5 lugar	2.432	0,54	15 gran	1.482	0,33
6 mejor	2.211	0,49	16 parque	1.427	0,32
7 paseo	2.080	0,47	17 buena	1.401	0,31
8 impresionante	2.074	0,46	18 recomendable	1.327	0,30
9 experiencia	1.854	0,41	19 vistas	1.305	0,29
10 genial	1.706	0,38	20 precioso	1.246	0,28

Fuente: Muestra aleatoria de OTRs de TripAdvisor escritos en español entre 2011 y 2015



La Tabla 4 recoge la frecuencia de todas las palabras significativas (16.344 palabras distintas) que hay en los títulos de los 88.000 OTRs considerados (446.853 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 65% 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 Barcelona. 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 como excelente, impresionante, genial, espectacular, bonito, interesante y precioso.

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, afectivo y espacial de la TDI percibida (y transmitida) por los viajeros, 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. El análisis se ha centrado en los metadatos de la página web que procesan los buscadores 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 de Gaudí). En el caso del componente 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 al componente espacial se ha detectado la gran concentración de la imagen en la capital del destino.

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, es muy interesante conocer la especialización de la TDI en el espacio y su evolución en el tiempo; o bien explorar si se observa alguna diferencia de percepción entre turistas hispanohablantes y angloparlantes, dado que los primeros, en principio, tienen una relación más estrecha con el destino estudiado, o entre distintas nacionalidades de origen (turistas españoles, europeos y ultramarinos).



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).

6. REFERENCIAS BIBLIOGRÁFICAS

- 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. [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. <http://doi.org/10.1016/j.annals.2004.01.010>
- Chon, K.-S. (1990). The role of destination image in tourism: A review and discussion. *The Tourist Review*, 45(2), 2–9. <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. <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. <http://doi.org/10.1016/j.jdmm.2013.12.006>
- Eurobarometer. (2016). *Flash Eurobarometer 432: Preferences of Europeans towards tourism*. Brussels, Belgium.
- Eurostat. (2015). Tourism. In M. Kotzeva (Ed.), *Eurostat regional yearbook 2015* (pp. 189–206). Luxembourg: Publications Office of the European Union.
- Gartner, W. C. (1993). Image formation process. *Journal of Travel & Tourism Marketing*, 2(2-3), 191–215. http://doi.org/10.1300/J073v02n02_12
- Kim, H. (2015). Comparing use of the Internet for trip planning between leisure and business travelers, 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.
- 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. <http://doi.org/10.1016/j.tourman.2014.11.012>
- Marine-Roig, E. (2010). Los “Travel Blogs” como objetos de estudio de la imagen percibida de un destino. In A. J. Guevara Plaza, A. Aguayo Maldonado, & J. L.



- Caro Herrero (Eds.), *Turismo y Tecnologías de la Información y las Comunicaciones* (pp. 61–76). Málaga, Spain: Facultad de Turismo.
- Marine-Roig, E. (2014). A webometric analysis of travel blogs and review hosting: The case of Catalonia. *Journal of Travel & Tourism Marketing*, 31(3), 381–396. <http://doi.org/10.1080/10548408.2013.877413>
- Marine-Roig, E. (2015). Identity and authenticity in destination image construction. *Anatolia: An International Journal of Tourism and Hospitality Research*, 26(4), 574–587. <http://doi.org/10.1080/13032917.2015.1040814>
- Marine-Roig, E., & Anton Clavé, S. (2014). Recogida masiva, clasificación y depuración de blogs y reseñas de viaje: Caso de Cataluña. In A. Cerezo Medina, A. Guevara Plaza, & A. Aguayo Maldonado (Eds.), *Turismo y Tecnologías de la Información y las Comunicaciones* (pp. 83–96). Málaga, Spain: Facultad de Turismo.
- Marine-Roig, E., & Anton Clavé, S. (2016a). A detailed method for destination image analysis using user-generated content. *Information Technology & Tourism*, 15(4), 341–364. <http://doi.org/10.1007/s40558-015-0040-1>
- Marine-Roig, E., & Anton Clavé, S. (2016b). Perceived image specialisation in multiscalar tourism destinations. *Journal of Destination Marketing & Management*, 5(3), 202–213. <http://doi.org/http://dx.doi.org/10.1016/j.jdmm.2015.12.007>
- OdEiO. (2016). *2015 annual tourist balance sheet*. Retrieved from <http://observatoriempresaiocupacio.gencat.cat>
- Pan, B., & Fesenmaier, D. R. (2006). Online information search. Vacation planning process. *Annals of Tourism Research*, 33(3), 809–832. <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 U. Gretzel, R. Law, & M. Fuchs (Eds.), *Information and Communication Technologies in Tourism 2010* (pp. 405–416). Vienna, Austria: Springer. http://doi.org/10.1007/978-3-211-99407-8_34
- Pocock, D., & Hudson, R. (1978). *Images of the urban environment*. London, UK: Macmillan.
- Son, A. (2005). The measurement of tourist destination image: applying a sketch map technique. *International Journal of Tourism Research*, 7(4-5), 279–294. <http://doi.org/10.1002/jtr.532>
- Tan, W. & Wu, C. (2016). An investigation of the relationships among destination familiarity, destination image and future visit intention. *Journal of Destination Marketing & Management*, 5(3), 214–226. <http://dx.doi.org/10.1016/j.jdmm.2015.12.008>
- W3Schools. (2016). *HTML Tutorial*. Retrieved from <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. <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.
<http://doi.org/10.1177/0047287508321193>