

Article

Measuring Destination Image through Travel Reviews in Search Engines

Estela Marine-Roig 

Serra Hünter Fellow, University of Lleida, C/ Jaume II, 73, 25001 Lleida, Catalonia, Spain; estela.marine@aegern.udl.cat

Received: 16 July 2017; Accepted: 10 August 2017; Published: 12 August 2017

Abstract: In recent years, mobile phones and access points to free Wi-Fi services have been enhanced, which has made it easier for travellers to share their stories, pictures, and video clips online during a trip. At the same time, online travel review (OTR) websites have grown significantly, allowing users to post their travel experiences, opinions, comments, and ratings in a structured way. Moreover, Internet search engines play a crucial role in locating and presenting OTRs before and throughout a trip. This evolution of social media and information and communication technologies has upset the classic sources of information of the projected tourist destination image (TDI), allowing electronic word-of-mouth to occupy a prominent position. Hence, the aim of this paper is to propose a method based on big data technologies for analysing and measuring the perceived (and transmitted) TDI from OTRs as presented in search engines, emphasising the cognitive, spatial, temporal, evaluative, and affective TDI dimensions. To test this approach, a massive analysis of metadata processed by search engines was performed on 387,414 TripAdvisor OTRs on ‘Things to Do’ in Île de France, an outstanding smart tourist destination. The results obtained are consistent and allow for the extraction of insights and business intelligence.

Keywords: destination image; user-generated content; electronic word-of-mouth; search engines; big data; metadata mining; content analysis; business intelligence; Paris

1. Introduction

In the last decade, user-generated content (UGC) has grown dramatically, parallel to the rise of information and communication technologies (ICT) and social media. In the field of hospitality and tourism, said growth in the use of social media has been especially noticed in vacation planning. In a survey of 30,105 respondents selected from different social and demographic groups in the European Union [1], the primary source of information used to plan a holiday was stated to be word-of-mouth (WOM), recommendations of relatives, friends and colleagues (51%), followed by electronic word-of-mouth (eWOM), websites collecting and presenting comments, reviews, and ratings from travellers (34%). In relation to American travellers, the percentage of Internet use as a source of information for trip planning is higher than that in relation to European travellers and stands at around 85%, followed by WOM with 42% [2,3]. Nevertheless, search engines play a very important role in locating and presenting information about destinations [4,5].

eWOM communication is mainly based on UGC that can be accessed online for free. UGC is unsolicited and unbiased first-hand information. Therefore, it is deemed reliable by other users. In a survey [6], 54% of tourists responded to the statement ‘I trust reviews on social media from other tourists’ that they agree or strongly agree, and only 11% disagreed or strongly disagreed. In recent years, online travel reviews (OTRs) have proliferated greatly. In OTRs, tourists freely recount their experiences at the destination they visited and give their opinion and/or evaluation of specific attractions (monuments, museums, parks, etc.) and services (hotels, restaurants, transport, etc.). As an

example of the growth of UGC, in January 2015, Booking claimed to have 43 million verified reviews from real guests, and TripAdvisor received over 200 million comments and opinions from travellers, and each received 123 and 500 million OTRs, respectively, in July 2017.

According to Xiang et al. [7], primary research has traditionally been done by communication-based studies such as surveys and in-depth interviews, designed to compile data directly from users and consumers. Today, due to the aforementioned potential and exponential growth in the use of social media in travelling, the tourism and hospitality industry appears to be an ideal field for social media analytics. For instance, big data has been used by Miah et al. [8] for tourist behaviour analysis; Jabreel et al. [9] for semantic comparison of emotional values; Kirilenko et al. [10] to conduct sentiment analysis of public attitudes; and by Liu et al. [11] to analyse the satisfaction of guests in the hospitality field.

In the field of tourism, through trip diaries, a perceived (and transmitted) image becomes a projected image by the eWOM effect and contributes to closing the circle of tourist destination image (TDI) formation from a holistic perspective [12]. Therefore, the main objective of this article is to build a theoretical and methodological framework for the mass analysis and measurement of the TDI projected by OTRs through search engines. This framework applies to the case of a prominent smart tourist destination (STD), as indicated in the measurement-approach section.

2. Theoretical Framework

The study of TDI has been a constant in the scientific literature on tourism [13,14] because images are of decisive importance in transposing the representation of an area inside the minds of potential tourists, giving them a pre-perception of the destination [15], and are considered to play a crucial role in an individual's travel purchase decision-making [14,16]. Reynolds [17] emphasised that product and brand images are created by consumers and that an image is the mental construct developed by the consumer on the basis of a few selected impressions from the flood of total impressions (p. 69). In a simple way, the TDI can be defined as the sum of beliefs, ideas, and impressions that a person has of a destination [18] (p. 18). However, after reviewing 45 valid definitions of the TDI, Lai et al. [19] (p. 1074) propose a very elaborate definition: TDI is 'a voluntary, multisensory, primarily picture-like, qualia-arousing, conscious, and quasi-perceptual mental (i.e., private, nonspatial, and intentional) experience held by tourists about a destination. This experience overlaps and/or parallels the other mental experiences of tourists, including their sensation, perception, mental representation, cognitive map, consciousness, memory, and attitude of the destination.' To facilitate its analysis, TDI is divided into three distinct but hierarchically interrelated components: cognitive, affective, and conative [20,21].

2.1. TDI Components

According to Rapoport [21], the individual-environment interaction involves three areas: knowing something, feeling something about it, and then doing something about it. These areas correspond to three stages: cognitive, affective, and conative. In this vein, Baloglu et al. [22] state that the construction of the TDI depends on two evaluations; perceptual/cognitive, referring to beliefs or knowledge about the attributes of a destination, and affective, referring to feelings or attachment to the same. The overall perceived image is formed by the union of both [12,23]. A third component is directly derived from the previous two, the conative, which affects the individual's behaviour when selecting a destination, based on the images received during the cognitive phase and evaluated during the affective phase [20]. This dichotomy of a cognitive-affective image has been extensively studied in the field of tourism [16].

Subsequent to Rapoport's [21] classification of the image areas, Pocock et al. [24] proposed a parallel but more detailed alternative schema (Figure 1): (1) The designative image, concerning description and classification, which is informative in nature and based on an individual's knowledge of what is and where its environment is set (the basic whatness and whereness of the image); (2) the appraisive image, meaning attached to, or evoked by, the physical form, which is based on the appraisal or assessment; and (3) the prescriptive image, which relates to predictions and inferences of both the designative and appraisive natures.

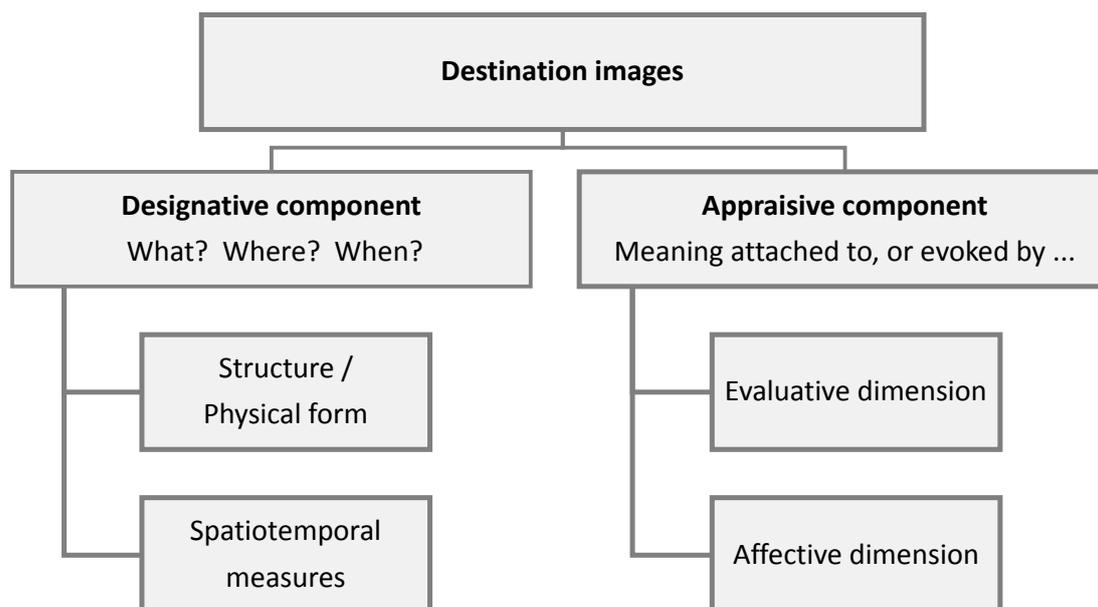


Figure 1. Tourist destination image (TDI) components. Source: author, derived from Pocock et al. [24].

Figure 1 shows two basic components of the TDI; the designative or informational image, based on the categorisation of cognitive elements of the environment, and the appraisive image, concerning the appraisal or assessment of these elements. The structure and physical form include features such as texture, shape, size, and layout. The spatial characteristics refer to distance, relative location, and directional relationships [24]. The spatial aspect of the designative image was studied by Son [25], who obtained sketches, which were in turn used to measure mental maps, and by Marine-Roig et al. [26] through spatial coefficients for elucidating image specialization in multiscale destinations. The appraisive component incorporates both evaluation and preference, the former including some general or external standards and the latter reflecting a more personal type of appraisal and affection, which is the emotional response concerned with the value, feeling, and meaning attached to the perceived image [24] (p. 30). In other words, the designative component is related to the structure of the environment and the appraisive to its sense or meaning. In this vein, Lynch [27] stated that each person has a visual image of the city and that the image gives the city structure, meaning, and identity.

In addition to placing the TDI in space, it must be taken into account that the image is perceived by an individual at a given time. The image may vary over the years or may change from season to season; for instance, the perceived image of Japan during the time that the cherry-trees are in blossom or the image of the Mediterranean coast in summer and in winter. Therefore, the spatiotemporal dimension of the TDI must be considered.

2.2. TDI Information Sources

TDI information sources are classified as primary and secondary [28]. Primary TDI sources are derived from information about a destination that visitors acquire from their own experiences, while secondary sources are formed from information received through other people or organisations. In turn, secondary sources are divided into induced, autonomous, and organic sources [20]. Induced sources are characterised by the fact that agents depend on destination organisations, that is, they are an interested party in the process of selecting the trip. These sources can be subdivided into overt-induced and covert-induced based on the degree of knowledge that the recipient has about their origin. Autonomous TDI formation agents are independent of the destination such as reportages, newspaper articles, and documentaries. Organic sources (solicited and unsolicited) emanate from friends, colleagues, and relatives (WOM advertising). The credibility of the sources is inversely proportional

to the degree of control that the destination has over the source. The more control the destination has over the source, the less credible it appears to the traveller. Thus the most valued sources are autonomous and organic.

It has been almost 25 years since Gartner [20] published its classification of TDI information sources, and travellers have changed their habits to attain information to select a destination. For instance, in the macro-survey [1] seen in the introduction, websites that collect and present OTRs were ranked second, behind WOM but above primary sources and far above service providers, tourist offices, travel agencies, etc. According to a survey of 11,400 international tourists [29], the top three online influences on destination choice were search engines, price comparison sites, and traveller review sites. In a multidimensional analysis on information sources for the formation of TDI [5], the Internet was ranked first, and, within the web platforms, search engines (e.g., Google), maps (e.g., Google Maps), and webpages with assessments by users (e.g., TripAdvisor) were frequently utilised.

From previous surveys, OTRs have been demonstrated to be a significant source of information in all cases, and search engines have occupied a preponderant place in the search for information about a destination. However, search engines are not a source of information themselves. Instead, websites that collect OTRs obtain first-hand information about the destination, which can be located and presented to users by search engines. Then Gartner's [20] framework is valid by adding the content generated online by travellers (travel blogs and OTRs) within the unsolicited-organic information sources. Otherwise, OTRs are spread primarily through social media (eWOM) and search engines.

2.3. TDI Dimensions and Paratextual Elements of OTRs

The term 'paratext' was introduced by Genette [30] to define a set of productions (an author's name, a title, a preface, illustrations, etc.) accompanying the text of a literary work, which can be discussed whether they belong in the traditional sense to the text or not, but in any case surround and extend it. This author divides them into 'peritext' and 'epitext' according to the distance of the elements from the text itself. The production of the paratext is directly, but not exclusively, the responsibility of the publisher or the publishing house. Marine-Roig [31] proposes a framework to adapt the theory of Genette to the case of OTRs on attractions or services and classifies as peritext their titles, ratings, language, subjects, type, dates, and geographic locations, followed by the author's profile and, as epitext, the related reviews and comments of other users and contextual advertising. In this case, the paratext is generated by the author (UGC), by the webmaster (WGC: webhost- or webmaster-generated content), or by both together.

The title is the most important peritextual element, fulfils the function of summarizing and previewing the experience reported in the OTR [32], and consists of two parts, the title in a strict sense (UGC) and the complementary information (WGC). The UGC part of the title, for its great content of adjectives and recommendations, is useful in deducing the affective dimension, and the WGC part for the designative component. The rating fits with the evaluative dimension. The subject (attraction or service) and type (sights, landmarks, etc.) contribute to the analysis of the cognitive dimension. Finally, the date and location delimit the spatiotemporal dimension.

2.4. OTRs on Search Engines

For their potential to index and organise vast amounts of information, search engines are powerful tools that represent the virtual world and therefore the domain of tourism [33] (p. 146), and their role is becoming increasingly important in the marketing programs of online tourist organisations [34]. Therefore, search engines have great potential to capture the projected TDI and facilitate travel planning [35].

Pan et al. [36] propose a conceptual model of planning a trip through the Internet based on the interaction between users and the portion of the Web related to the industry and tourist destinations. From this framework, an online search information model emerges with three components: the traveller, the interface, and the online space [33]. The effectiveness of a search

depends on the situation, knowledge and skills of the user (traveller), the quantity and quality of related websites (tourist domain online), and the functionalities of the web browsers and search engines (interface) used to facilitate the results.

Search engines basically consist of two parts; a parser that timelessly runs the Web and collects or updates the most significant information used to build a database indexed by key words or phrases and an online component receiving users' queries and returning corresponding results sorted by relevance and visibility [37]. These results are often presented based on the metadata of the indexed webpage, including the title, with a link and a brief summary [33].

An analysis by Xiang et al. [38] showed that social media carries substantial weight in search results related to travel planning, with OTRs representing a significant amount of social media for travel purposes [29]. In order to show that data from an OTR appears in search engines (Figure 2), we have chosen an OTR from TripAdvisor, the largest user-generated online review site in the tourism domain [7,39], and the three search engines with the most traffic, Google, Baidu, and Yahoo (Alexa.com, TopSites). It is noteworthy that Yahoo does not use its own means and presents the results obtained by Live.com through Microsoft's Bing search engine.

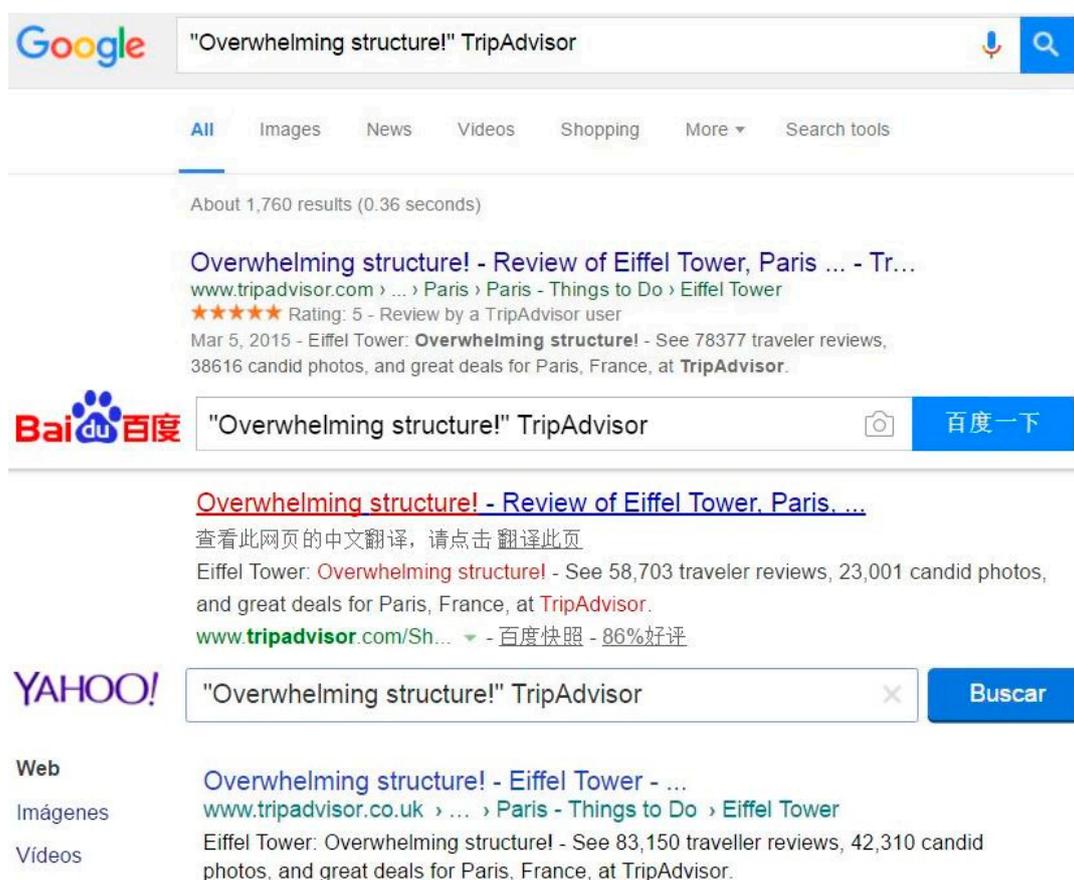


Figure 2. View of the same online travel review (OTR) in the three search engines with the most traffic.

Metadata is a set of data describing or giving information about other data. The web label <meta ... /> [40] contains metadata about an HTML (HyperText Mark-up Language) page. These metadata are not displayed on the website because they are intended to provide information to browsers and search engines on the Internet. HTML metadata consist of pairs of tags composed of a name and content (Table 1). The most common elements are the description of the webpage, the list of keywords, and the name of the author of the document [40]. Although the information contained in Table 1 and in Figure 2 was collected on the same day, a discrepancy can be seen in the number of views and

photos because the webhost is updated daily, while the three search engines that captured metadata on previous dates see no need to update so often.

Table 1. Contents of the most important HyperText Mark-up Language (HTML) meta-tags (see Figure 2).

Meta Name	Content
<i>title</i>	Overwhelming structure!—Review of Eiffel Tower, Paris, France—TripAdvisor
<i>hyperlink</i>	https://www.tripadvisor.com/ShowUserReviews-g187147-d188151-r257879021-Eiffel_Tower-Paris_Ile_de_France.html
<i>keywords</i>	Eiffel Tower, Paris, France, Overwhelming structure!
<i>description</i>	Eiffel Tower: Overwhelming structure! —See 86,418 traveler reviews, 44,570 candid photos, and great deals for Paris, France, at TripAdvisor

Source: TripAdvisor OTR on the Eiffel Tower dated 5 March 2015.

In the example used in Figure 2, we can see the query made in three search engines, which is the title of an OTR (composed of two words and an exclamation mark) and the domain name where it is hosted. The title matches a key phrase that is in the meta-tag keywords (Table 1), by which search engines indexed the OTR webpage. The three search engines return the title, description, and web address of the OTR (Figure 2 and Table 1). Google, the global website with the most Internet traffic (Alexa.com, TopSites), also returns the score, author, and date of the OTR by a script.

2.5. Measurement Approach

In short, the TDI is made up of the physical environment and its meaning and values, as perceived by individuals at a given time; thus image-building is individualistic and subjective [41]. Therefore, the more opinions that are analysed, the more precise the results will be on the perceived TDI as a whole. The websites that host OTRs have a lot of free opinions from very different users, and big data technologies facilitate their processing [11,42,43].

TDI has a considerable influence on users when planning trips or holidays. A significant diffusion of TDI is performed through eWOM communication, which occurs in online information sources that collect and present traveller comments, reviews, and ratings, and through search engines that locate these sources and present a summary of the data. Therefore, the aim of this paper is to propose a methodology for a massive analysis of OTRs on a tourist destination in order to elucidate and measure the projected image, consisting of the image perceived by visitors as presented by the web host and spread through Internet search engines. To test the method, we used certain HTML metadata processing search engines to obtain a sample of 387,414 Things-to-Do OTRs housed on TripAdvisor, written in English by tourists from more than 150 countries, who were visiting Île de France between 2007 and 2016, and we obtained significant results in five dimensions of the TDI (cognitive, spatial, evaluative, affective, and temporal). Furthermore, to explore to what extent the OTR paratextual productions are significant in building the TDI, another sample of 123,726 OTRs on the four most popular landmarks is analysed and discussed.

3. Materials and Methods

The approach is founded on the theoretical framework proposed by Marine-Roig [31] on the relationship between an OTR and the paratextual productions around it, which are both generated by the traveller (UGC) as well as the by the webmaster (WGC). The entity-relationship diagram constructed by this author shows the closeness between the writing body of an OTR and its surrounding peritextual elements to the extent that the reviewers' experiences, opinions, or assessments are meaningless; for example, if they are not placed in time and space. With regard to implementation, the method follows the batch-processing paradigm, that is, the big data are first stored and then

analysed [43], but it is not necessary to work on a distributed system because approximately one million HTML files (250 GB: GigaBytes) can be processed on a single workstation. The proposed method consists of the following phases: data collection, HTML metadata mining, and quantitative analysis.

3.1. Case Study

According to the World Tourism Organization (UNWTO), Europe was the most frequently visited region in the world in 2015, and Île de France was the most touristic continental region of the European Union, with 77.7 million overnight stays [44]. Île de France is an outstanding STD and has a peculiar geographical distribution of departments, comprising a metropolis surrounded by an inner ring, and this, in turn, is surrounded by an outer ring (Figure 3), which allows for the study of the spatial dimension of the TDI. Moreover, France is a non-English-speaking country, so problems caused by special characters (i.e., characters above ASCII 127: 7-bit American Standard Code for Information Interchange) that are not part of the English alphabet should be attended to.

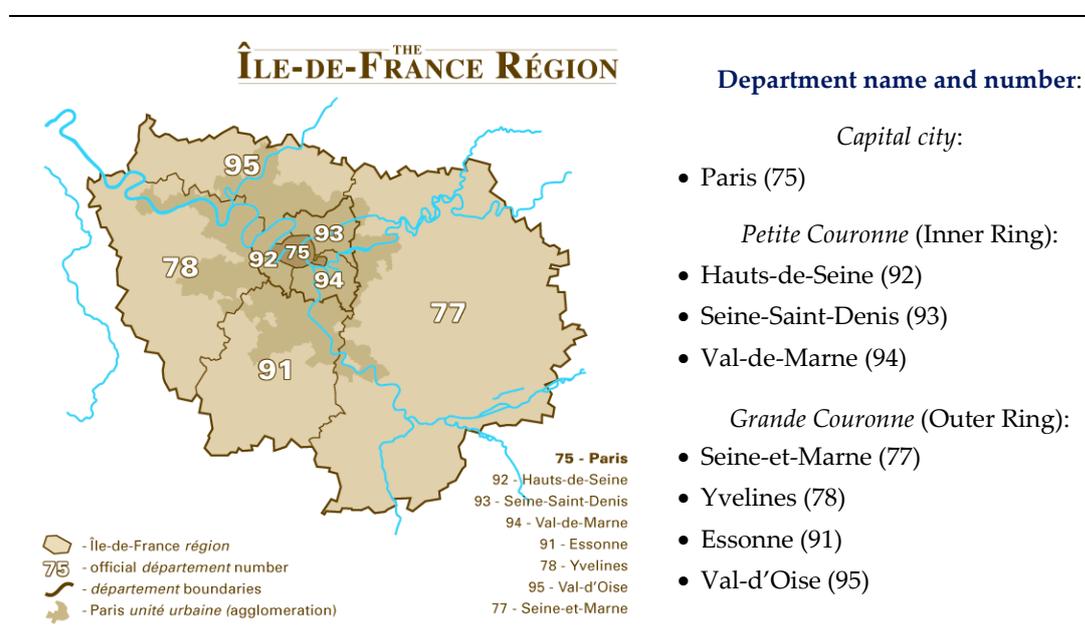


Figure 3. Map of Île-de-France's departments. Source: author, derived from J. M. Schomburg (WikiMedia Commons).

3.2. Data Collection

The main online sources of travel-related stories and opinions are websites that host travel blogs and OTRs because they present the information in a structured way, allowing a person to automate the download, classification, and analysis. On the basis of past work and an update of the search, 12 portals are located with abundant information about Île de France during the studied period (2007–2016). To choose the most suitable source, a weighted formula of aggregation of rankings [37] based on Borda's [45] positional method (B) is applied with the webometric variables visibility (V), popularity (P), and size (S):

$$\text{TBRH} = 1 \times \text{B(V)} + 1 \times \text{B(P)} + 2 \times \text{B(S)}$$

Once this formula is applied, TripAdvisor comes first and outscores the other webhosts by far. This selection matches Baka (2016), who considers TripAdvisor the world's largest source of UGC in the domain of tourism, and other authors [11,46], explaining the advantages of collecting a set of open data in TripAdvisor because of the huge amount of user-generated reviews that it hosts. Moreover, Yoo et al. [47] note that TripAdvisor's reputation management system helps to determine

the helpfulness of reviews and/or reviewers (which enables viewing profiles, other reviews, votes, and ratings) and motivates users to contribute reliable reviews (through intrinsic and extrinsic motivations).

Dismissing reviews about hotels and restaurants for their high degree of specialization, in January 2017, 890,682 Things-to-Do OTRs on Île de France in several languages were downloaded [48] by means of a web copier, Offline Explorer Enterprise (OEE). OEE delivers high-level downloading technology and industrial-strength capabilities, downloads up to 100 million URLs per project, and archives websites automatically on a regular basis (MetaProducts.com). In this case study, the most representative language of TripAdvisor is English, due to the greater volume of OTRs and the variety of countries (more than 150) of origin among foreign reviewers. These countries are (in order from the most to fewest number of OTRs): United States, United Kingdom, Australia, Canada, India, Ireland, New Zealand, Singapore, Germany, South Africa, Netherlands, Israel, etc. Then, a sample of 387,414 reviews written in English between 2007 and 2016 was selected (Table 2 and Figure 4). To check that the sample collected all the Things-to-Do OTRs written in English, the hyperlink codes (*) of the webpages of each OTR (ShowUserReview*.html) were crossed with those of the webpages of each attraction or service (Attraction_Review*.html).

Table 2. Sample of 387,414 TripAdvisor OTRs on Île de France per district and year.

	75	77	78	91	92	93	94	95
2007	316	68	13	0	3	1	1	1
2008	705	129	19	0	2	1	0	2
2009	1511	177	61	0	2	4	1	0
2010	2538	200	85	0	3	8	1	2
2011	10,123	795	363	0	17	28	8	6
2012	46,260	2846	1054	2	96	88	32	25
2013	49,205	3094	1361	8	86	148	41	26
2014	58,811	3459	1991	19	137	200	55	59
2015	93,707	5253	3046	24	249	239	68	116
2016	89,492	5105	2965	24	278	327	80	144

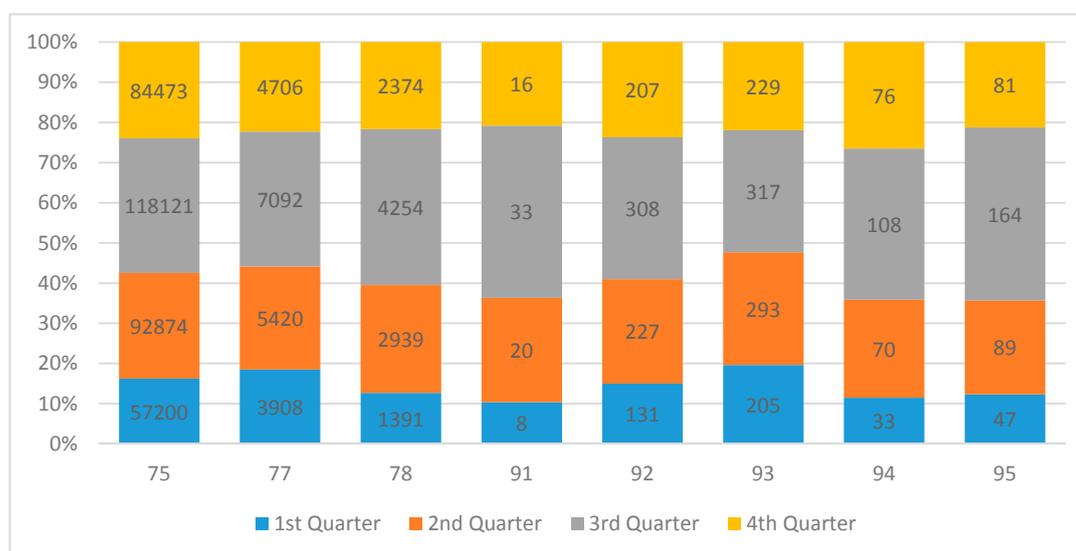


Figure 4. Sample of 387,414 TripAdvisor OTRs (2007–2016) on Île de France per district and trimester.

3.3. HTML Metadata Mining

In examining Figure 2 and Table 1, both on the same OTR, one can deduce that the three search engines retrieve the information presented in the title and description content delimited by the HTML

meta-tags. In addition, associated with the title, there is a hyperlink leading to the webpage hosting the OTR. In turn, the title consists of two parts; the title in strict sense written by the traveller (UGC) and the associated information (attraction or service, destination and domain) added by the webmaster (WGC). The hyperlink (Table 1) provides a lot of data that facilitates the process of the associated information analysis, including the protocol ([https:secureHTTP](https://secureHTTP)), server (www.tripadvisor.com), purpose of the webpage (show user's reviews), destination code (g187147: Paris), attraction code (d188151: Eiffel Tower), OTR code (r257879021), internal name of attraction (Eiffel_Tower), subregion name (Paris), region name (Ile_de_France), and type of webpage (html). On the other hand, unlike the other two search engines, Google presents an additional line with the rating (Rating: 5) of the attraction or service, author (TripAdvisor user), and OTR date (5 March 2015).

Thanks to the structure of webpages, we can extract the metadata described by simple expressions (regex: Regular expressions) of regular language (sequences of characters forming a search pattern) through a programme such as UltraEdit (ultraedit.com), which supports files larger than 4 GB, that admits regex and allows users to work with large amounts of data.

Special characters can be encoded in different ways in HTML pages. For example, Sacré Cœur (Sacred Heart) has two special characters: lowercase *e* acute and lowercase *œ* ligature. This ligature can be represented with at least four codes: Friendly (ö), numerical (œ), hexadecimal (œ), and UTF-8 (Ã°). These encodings baffle the parser and should be unified; one solution is to replace the special characters with the corresponding ISO 8859-15 (Latin alphabet 9) character. Finally, metadata in CSV format (comma separate values) was stored to handle files in plain text using a spreadsheet application.

3.4. Content Analysis

Content analysis can be defined as a systematic, replicable technique for compressing key words or key phrases into a few content categories. It allows researchers to sift through large volumes of data with relative ease in a systematic way [49]. The quantitative content analysis used in this research consists of three phases: parser configuration, frequency analysis, and categorisation.

3.4.1. Parser Settings

To classify and count the words in a text, the parser needs to know which characters are word separators, which are composite words, and which words are not considered keywords.

Word delimiters are generally regarded as word-separator characters blanks, commas, semicolons, etc.; however, to achieve greater precision, in this case all characters that are not letters in the English and French languages have been considered.

Composite words are groups of words that have different meanings together or separately such as 'pick pockets' or compound nouns like 'Notre Dame'.

A black list is a list of non-significant words for the case study and includes most adverbs, conjunctions, determiners, prepositions, and pronouns.

3.4.2. Frequency Analysis

Content analysis is usually based on a word-frequency count because, despite its flaws, it is assumed that the words mentioned most frequently reflect the greatest concerns [49]. Figure 5 shows the pseudocode algorithm used for frequency analysis. The algorithm is case insensitive and has two counters; one for total words in the text including stop words and another for unique keywords. To optimize the execution time, the stop words are stored in an ordered list and the results in a set (without repetitions) on a binary tree in order to implement the searches with a logarithmic (dichotomous) asymptotic cost.

```

Load compWord; // composite word (list of)
Load stopWord; // non-significant word (list of)
Load textWord; // text to analyse quantitatively
New result; // table to store frequency of keywords

for each compWord do
{
  if exists compWord in textWord then
  {
    Count occurrences of compWord in textWord;
    Add compWord to result with its frequency;
    Delete occurrences of compWord in textWord;
  }
}
for each word in textWord do
{
  if not exists word in stopWord then // becomes a keyword
  {
    if exists keyword in result increase its frequency;
    else add keyword to result with frequency 1;
  }
}
Print result;

```

Figure 5. Simplified algorithm used for frequency analysis. Source: Author.

3.4.3. Categorisation

There are two models of categorization; categories established a priori and categories analysed and deducted from the text itself (emergent coding) [49,50]. To analyse the cognitive, affective, and spatial components of the TDI, three categories of keywords were created; attractions, feelings, and locations.

Attractions: by means of a preliminary analysis, this category has been extracted from the WGC title (Table 1), that is, from the part of the OTR title generated by the webmaster, in order to be able to process the attractions with the same name that TripAdvisor uses. For example, although ‘Tour Eiffel’ is the original name of a landmark, the study considered the English version, ‘Eiffel Tower’, used by TripAdvisor.

Feelings: a dichotomous category has been constructed a priori divided into good feelings and bad feelings. Both are formed by American and English adjectives, interjections, and recommendations. For example, ‘beautiful’, ‘amazing’, ‘nice’, ‘wonderful’, ‘wow!’, ‘must see’, and ‘don’t miss’ represent good feelings, while ‘poor’, ‘disappointing’, ‘overcrowded’, ‘yuck’, ‘not great’, ‘not worth it’, and ‘beware of pickpockets’ are representative of bad feelings. On the other hand, the reviewers give a global rating (between 1 and 5) to the attraction or service: Excellent (5) and Very good (4) qualifications have been considered positive, Poor (2) and Terrible (1) negative, and Average (3) neutral.

Locations: the destinations are classified by areas (see Figure 3 and Table 2). For example, Versailles belongs to the department of Yvelines (78), and Marne-la-Vallée (where Disneyland Paris is located) is astride three departments, Seine-Saint-Denis, Val-de-Marne, and Seine-et-Marne, but it is identified here as Seine-et-Marne (77) due to that being the most touristic department of the region, after Paris (75).

4. Results and Discussion

Preliminary results of the spatiotemporal distribution of the 387,414 OTRs are obtained. In Table 2, a considerable growth of the quantity of OTRs is observed, with a great concentration in district 75

(Paris), which has more than 91% of the OTRs of the whole region. For example, three districts that border Paris (92, 93, and 94), which together form the Inner Ring (see Figure 3), when added up only represent 0.57% of the OTRs that Île de France had between 2007 and 2016. However, in 2016, a decrease in the OTRs of the main districts (75, 77, and 78) of Île de France was observed, which may be due to the impact of the awful 2015 attacks in Paris. Figure 4 shows that, in all districts, the third quarter is the most touristic by the number of OTRs, followed by the second, fourth, and first.

Table 3 shows the 387,414 analysed OTR titles (2,369,507 words). In the UGC keywords column, good feelings are predominant. Other columns indicate the number of occurrences of every keyword and the percentage that it represents of the total of words (including stop words).

Table 3. Top 20 OTR title keywords (user-generated content (UGC)).

Rank	Keyword	Count	Density	Rank	Keyword	Count	Density		
		Total: 2,369,507	Unique: 25,810	%			Total: 2,369,507	Unique: 25,810	%
1	paris	43,668	1.84292	11	see	10,518	0.44389		
2	great	30,925	1.30512	12	must see	9117	0.38476		
3	beautiful	22,817	0.96294	13	experience	9003	0.37995		
4	tour	17,867	0.75404	14	view	8786	0.37079		
5	amazing	15,052	0.63524	15	very	8754	0.36944		
6	museum	13,421	0.56640	16	way	8721	0.36805		
7	best	12,672	0.53479	17	nice	8720	0.36801		
8	place	12,301	0.51914	18	good	8025	0.33868		
9	visit	11,848	0.50002	19	day	7980	0.33678		
10	worth	10,745	0.45347	20	wonderful	7658	0.32319		

Source: Sample of 387,414 TripAdvisor OTRs (2007–2016) written in English.

4.1. Designative Image

Whatness: Table 4 shows the reviewed attractions as a result of filtering WGC titles by categories of attractions. The most visited attractions, judging by the number of OTRs, are Tour Eiffel, Musée du Louvre, Musée d’Orsay, and Cathédrale Notre Dame.

Table 4. Top 20 Île de France reviewed attractions per district, frequency, and ratings.

Attraction	District	Count	Excellent	Very good	Average	Poor	Terrible
Eiffel Tower	75	44,438	30,106	10,510	2990	538	294
Musee du Louvre	75	31,865	21,084	7657	2350	516	258
Musee d’Orsay	75	24,139	19,203	4093	627	117	99
Notre Dame Cathedral	75	23,284	15,742	5924	1404	163	51
Arc de Triomphe	75	13,595	8166	4233	1065	99	32
Disneyland Park	77	11,571	5102	3019	1832	911	707
Luxembourg Gardens	75	10,548	7182	2867	438	54	7
River Seine	75	10,027	6435	2886	577	83	46
Basilica du Sacre-Coeur	75	9094	5492	2595	614	211	182
Sainte-Chapelle	75	8486	6374	1487	463	111	51
Chateau de Versailles	78	7644	4051	2016	870	405	302
Musee de l’Orangerie	75	7366	5372	1525	385	59	25
Champs-Elysees	75	5124	2351	1531	937	215	90
Fat Tire Tours Paris	75	5124	4377	475	154	65	53
Palais Garnier	75	5027	3925	901	142	31	28
Musee Rodin	75	5022	3331	1294	323	58	16
Moulin Rouge	75	4870	2208	1229	703	356	374
Walt Disney Studios	77	4715	2284	1484	651	182	114
Montmartre	75	4223	2566	1120	327	130	80
Le Marais	75	4190	2897	1096	161	28	8

Source: Sample of 387,414 TripAdvisor OTRs (2007–2016) written in English.

Whereness: According to Marine-Roig et al. [26] (p. 204), destination image is territorially specialized; TDI specialisation refers to the degree to which certain places are communicated and perceived through certain imagery, activities, attributes, feelings, or identity components that

distinguish them from others as tourist destinations. Coinciding with the results of Table 2, which shows a large concentration of OTRs in the metropolis, in Table 4, the most visited attractions are in Paris (75), except for Disneyland Park and Walt Disney Studios, which are in the 77th district, and Château de Versailles, which is in the 78th.

4.2. Appraisive Image

With the data provided by the search engines (Figure 2 and Table 1), there are two methods used to analyse feelings toward the destination attributes; user ratings and the expressions of the reviewers in the OTR titles.

Evaluative dimension: Google shows the rating given by each user for each attraction or service, which allows for the calculation of the average of the 387,414 reviews (Table 4): Excellent (252,509: 65.18%), Very good (91,633: 23.65%), Average (27,928: 7.21%), Poor (8404: 2.17%), and Terrible (6940: 1.79%). This assessment can be grouped into three cases; Positive scores (88.83%), Neutral scores (7.21%), and Negative scores (3.96%).

Affective dimension: Filtering the UGC keywords column of Table 3 with the feelings category, 215,031 keywords are related to Good feelings (9.07% of total words including stop words) and 11,257 to Bad feelings (0.48%).

The results obtained with both methods are similar but do not match exactly because the Positive scores are more than 22 times higher than the Negative in the case of the evaluative dimension, whereas the Good feelings are less than 19 times higher than the Bad in the case of the affective dimension. This inconsistency shows that both dimensions (evaluative and affective) are useful to measure the appraisive component of an image.

4.3. Zoom on the Four Main Attractions

To explore and gain insight into the image of attractions, we extracted 123,726 OTRs on the four principal attractions (Eiffel Tower, Musee du Louvre, Musee d'Orsay, and Notre Dame cathedral). An analysis of frequencies was done (Table 5), and these frequencies were superficially compared. In general, the titles of the four attractions contain many positive adjectives like 'amazing', 'beautiful', and 'great'.

Table 5. Top 25 OTR title keywords (UGC).

Eiffel Tower		Musee Louvre		Musee Orsay		Notre Dame	
Unique: 5342	Total: 259,871	Unique: 4946	Total: 192,729	Unique: 3703	Total: 140,012	Unique: 3684	Total: 119,189
4583	paris	2996	museum	5212	museum	3539	beautiful
3035	eiffel tower	2390	louvre	2745	art	1652	cathedral
2731	amazing	1925	amazing	2197	paris	1448	paris
2594	view	1728	art	2162	great	1292	amazing
2180	night	1501	see	1737	impressionist/s	1221	must see
2132	great	1405	visit	1499	best	975	notre dame
2054	must see	1381	must see	1426	beautiful	912	visit
1964	beautiful	1344	paris	1016	must see	795	architecture
1748	views	1320	great	999	wonderful	793	worth
1462	visit	1154	day	988	visit	756	church
1454	worth	983	time	935	amazing	717	great
1391	top	962	mona lisa	916	collection	670	stunning
1333	icon/ic	919	world	909	favorite	629	place
1134	experience	915	huge	814	louvre	508	very
1122	tower	913	best	803	building	473	impressive
962	see	911	much	688	worth	472	history
938	line/s	910	place	554	place	465	breathhtaking
892	queue/s	866	worth	497	musee d'orsay	424	gothic
794	best	844	beautiful	487	fantastic	403	nice

Table 5. Cont.

Eiffel Tower		Musee Louvre		Musee Orsay		Notre Dame	
Unique: 5342	Total: 259,871	Unique: 4946	Total: 192,729	Unique: 3703	Total: 140,012	Unique: 3684	Total: 119,189
763	long	618	very	447	better	396	building
757	time	616	big	446	see	393	free
721	day	557	overwhelming	443	world	364	magnificent
703	awesome	551	crowded	442	excellent	350	see
699	ticket/s	531	experience	421	love	330	experience
697	must do	503	wonderful	398	impressionism	298	wonderful

Source: Sample of 123,726 TripAdvisor OTR UGC titles on four main attractions.

The titles of the Eiffel Tower frequently contain singular keywords: ‘wait/s (621)’, ‘ticket/s (699)’, ‘in advance (343)’, ‘line/s (938)’, ‘queue/s (892)’, etc., which relate mainly to the problems of queuing and waiting to acquire tickets or access to the tower and reviewers recommending buying tickets in advance. Also, they often contain the keyword ‘icon/ic’ because the Eiffel Tower is the symbol of Paris and of the whole of France.

The titles of the Musee du Louvre have many occurrences of the keyword ‘mona lisa’ because La Gioconda (known as Mona Lisa) is one of the most famous works of art in the world; keywords such as ‘crowd/s (390)’, ‘crowded (551)’, ‘overcrowded (44)’, etc. also appear to complain about the crowds that occur in the museum (see Figure 6, and the keywords ‘time (983)’ and ‘need/s (490)’ also appear because many reviewers recommend taking more time to visit the museum.



Figure 6. Crowds in the Louvre (Mona Lisa). Author: Max Fercondini (Wikimedia Commons).

The titles of the Musee d’Orsay demonstrate the use of the keywords ‘impressionism’ and ‘impressionist/s’ since the museum houses the largest collection of Impressionist masterpieces, and

the titles of Notre Dame have the keywords ‘cathedral’, ‘church’, ‘architecture’, and ‘gothic’ because the cathedral is widely considered to be one of the finest examples of French Gothic architecture.

Figure 7 shows the feelings of visitors to the four studied attractions as a result of filtering Table 5 with the feelings category. The feeling bars in the graph represent the percentage of occurrences of keywords on feelings in relation to the total words (including stop words) of the OTR titles for each attraction. Additionally, score bars represent the ratio of scores for every 10 OTRs. For example, ‘Tour Eiffel Score+ 9.14’ means that 91.4% of OTRs on the Eiffel Tower have an Excellent or Very Good rating.

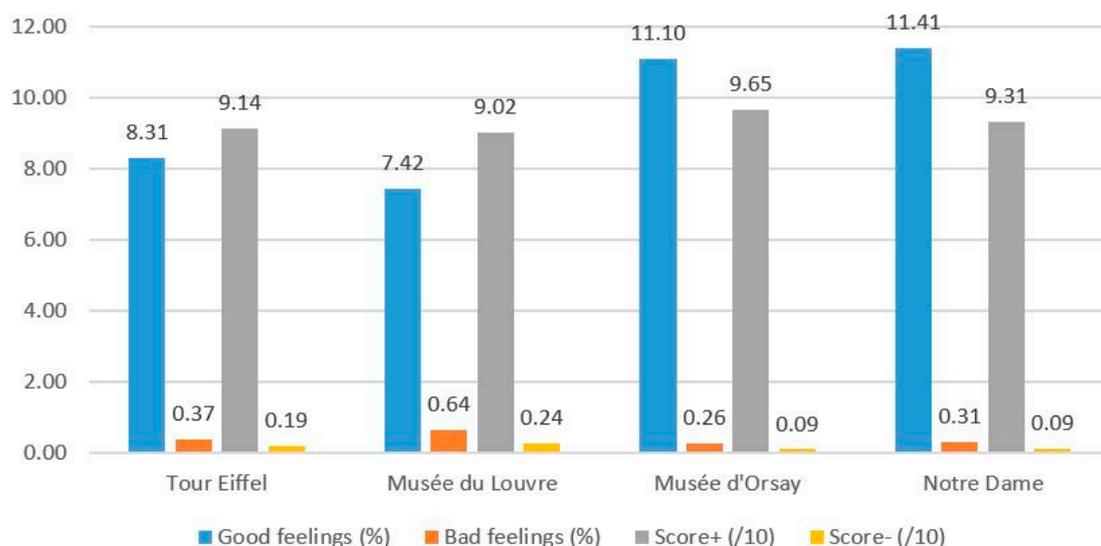


Figure 7. Appraisive image of four Île de France attractions. Source: Sample of 123,726 OTR UGC titles.

Two attractions (Eiffel Tower: 8.31%; Musée du Louvre: 7.42%) demonstrate a percentage of good feelings below the average (9.07%) of the 387,414 OTRs, and the other two attractions (Musée d’Orsay: 11.10%; Notre Dame: 11.41%) demonstrate a percentages of good feelings above the average. These results are consistent with the ratings in Table 4, where the first two (Eiffel Tower: +91.40%, −1.87%; Musée du Louvre: +90.20%, −2.43) show lower grades than the second two (Musée d’Orsay: +96.51%, −0.89%; Notre Dame: +93.05, −0.92%). The above results may be due to the problems identified in Table 5 on queues and waiting at the Eiffel Tower and overcrowding (see Figure 6) and insufficient time to visit the Musée du Louvre. However, there is an inconsistency in the evaluative and affective dimensions of the appraisive image between the two best rated attractions (Musée d’Orsay and Notre Dame), as can be seen by the order of the key figures (Figure 7). As with the inconsistency seen above, in these cases it is also demonstrated that the two dimensions are necessary to measure the appraisive image component.

5. Concluding Remarks

The proposed method allows the perceived image by travellers as transmitted by OTR webhosts and displayed in search engines to be analyzed and measured. This perceived and transmitted image becomes a projected image and contributes to closing the circle of the TDI.

The case study is meaningful because it includes the search engines with more traffic and, especially, Google (the website with the most traffic in the world); TripAdvisor, the travel-related website with the greatest volume of content generated by users; and Île de France, the most touristic continental region of the European Union. The number of analysed OTRs (387,414 in general and 123,726 in particular) represents the totality of the OTRs that meet the requirements (reviews on Things to Do in Île de France written in English between 2007 and 2016), and, therefore, the results allow for the derivation of reliable insights and business intelligence.

The method is reliable for several reasons: the quantitative content analysis of stored big data has little likelihood of error; the source of information (UGC data) is trustworthy according what the majority of researchers support; and the analysis of the appraisive component of the image has demonstrated the usefulness of its two dimensions (evaluative and affective) to measure the TDI. Furthermore, the available information is large and relatively easy to obtain because it is possible to access freely on trip-related websites hosting travel blogs and OTRs.

The main information from a TripAdvisor OTR that search engines show comes from the content of the HTML meta-tag <title>. The title is divided into two parts: a summary of the opinion of the tourist on an attraction or service (UGC) and the name and location of the attraction or service (WGC). The analysis of the first part is very useful for inferring the affective dimension of the TDI and the second part for the cognitive and spatial dimensions. Evaluative dimension analysis has revealed a high degree of satisfaction among tourists with the main attractions, and affective dimension analysis has shown a massive use of positive adjectives. The analysis of cognitive and spatial components found a high concentration of attractions in the metropolis, constituting more than 90% of the entire region.

5.1. Scientific Implications

While the dichotomy of the cognitive-affective image [20,21] has been commonly studied in the field of tourism, the parallel dichotomy of the designative-appraisive image [24] has received very little attention, and we believe it is more adequate to measure the perceived TDI from UGC. From the research background, through a significant case study (the most touristic region of the continent most frequently visited in the world) and a method of quantitative content analysis, it was demonstrated that the information (UGC and WGC) of the OTRs submitted by search engines contributes to the construction of TDI c, especially in five dimensions: cognitive, spatial, temporal, evaluative, and affective (Figure 1). In other words, both designative aspects (whatness and whereness) of the TDI as feelings or attachment to a destination's attributes are analysed and measured. The methodology employed in this study outlines a process of gathering, mining, and analysing massive tourism-related user-generated content (hundreds of thousands of visitors from more than 150 countries), which collectively constitutes the perceived image of the destination as a whole.

5.2. Managerial Implications

On the one hand, the information was analysed to find out the spatiotemporal distribution of OTRs and to know what and where the most visited attractions and the best rated are. These metrics allow destination management organisations (DMOs) to compare various attractions or services from the same or other destinations, in addition to places, territorial brands, or whole regions. The temporal dimension also allows for an analysis of the evolution of the tourist destination over time and the change (or permanence) of images perceived during different years or seasons. On the other hand, managers can acquire business intelligence (BI), with, for example, the problem detected in Paris concerning queues for purchasing tickets and accessing the Eiffel Tower or the crowds and the need to allow more time on visits to the Louvre Museum. Consequently, the occurrence of 'queues' and 'crowds' among the most-frequent words in OTRs of the two main landmarks is detrimental to the image of Paris, which aims to have sustainable tourism [51]. Focusing on the attribute-based TDI, this spontaneous demand-side information can be useful for DMOs to optimise products or services in the tourism supply chain and can contribute to the improved allocation of destination resources [52]. Moreover, the proposed framework allows one to gain insights and is very cost-effective in relation to the realization or replication of expensive surveys or interviews to determine the preferences and opinions of visitors.

5.3. Limitations

The main problem is that not all OTRs are indexed in search engines, but websites like TripAdvisor appear in the top positions of the results returned. Associated hyperlinks allow access to webhosts

that have a hierarchical structure of geographic classification of destinations and subclassifications like hotels, restaurants, and things to do at each destination. Moreover, the prescriptive [24] or conative [21] component of the analysis of the image is beyond the scope of this study.

5.4. Future Work

The writing body of the OTRs, together with the paratextual elements seen in this work, allows an in-depth analysis beyond the most frequent words. By using regular language expressions (regex), one can construct search patterns to extract phrases or groups of words related to the image, identity, authenticity, sustainability, or smartness of tourist destinations, and so on.

Regarding the implementation of the method, the algorithm in Figure 5 generates the frequency tables in near real-time, but the problem is the huge amount of noise in webpages since only an average of 4% of the internal content of downloaded pages is useful for this case study using the proposed method based on the batch-processing paradigm. HTML pages generally have a consistent structure, which allows researchers to take advantage of them. In this respect, additionally, other big data technologies could be used within the streaming-processing paradigm [43]; that is, they could filter data as they arrive and only store the useful page sections. The algorithm is complex, but the improvement is substantial because it would reduce the volume of stored data by about 25 times.

Acknowledgments: This work was supported by the Spanish Ministry of Economy and Competitiveness [Grant id.: MOVETUR CSO2014-51785-R]. Special thanks to Esteve Mariné for his expert advice in the field of Information Technology. The author states that she has not received funds to cover the costs of publishing in open access.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Eurobarometer. *Flash Eurobarometer 432: Preferences of Europeans towards Tourism*; European Commission: Brussels, Belgium, 2016.
2. Kim, H. Comparing use of the Internet for trip planning between leisure and business travelers. *Int. J. Bus. Humanit. Technol.* **2015**, *5*, 1–9.
3. Xiang, Z.; Wang, D.; O’Leary, J.T.; Fesenmaier, D.R. Adapting to the Internet: Trends in travelers’ use of the Web for trip planning. *J. Travel Res.* **2015**, *54*, 511–527. [CrossRef]
4. Camprubí, R.; Coromina, L. The influence of information sources in the image formation. *PASOS: Revista de Turismo y Patrimonio Cult.* **2016**, *14*, 781–796.
5. Llodrà-Riera, I.; Martínez-Ruiz, M.P.; Jiménez-Zarco, A.I.; Izquierdo-Yusta, A. A multidimensional analysis of the information sources construct and its relevance for destination image formation. *Tour. Manag.* **2015**, *48*, 319–328. [CrossRef]
6. VisitBritain. *Technology and Social Media: Foresight—Issue 152*; British Tourist Authority Board: London, UK, 2017.
7. Xiang, Z.; Du, Q.; Ma, Y.; Fan, W. A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tour. Manag.* **2017**, *58*, 51–65. [CrossRef]
8. Miah, S.J.; Vu, H.Q.; Gammack, J.; McGrath, M. A big data analytics method for tourist behaviour analysis. *Inf. Manag.* **2016**, in press.
9. Jabreel, M.; Moreno, A.; Huertas, A. Semantic comparison of the emotional values communicated by destinations and tourists on social media. *J. Destin. Mark. Manag.* **2016**, in press.
10. Kirilenko, A.P.; Stepchenkova, S.O. Sochi 2014 Olympics on Twitter: Perspectives of hosts and guests. *Tour. Manag.* **2017**, *63*, 54–65. [CrossRef]
11. Liu, Y.; Teichert, T.; Rossi, M.; Li, H.; Hu, F. Big data for big insights: Investigating language-specific drivers of hotel satisfaction with 412,784 user-generated reviews. *Tour. Manag.* **2017**, *59*, 554–563. [CrossRef]
12. Marine-Roig, E. Identity and authenticity in destination image construction. *Anatolia Int. J. Tour. Hosp. Res.* **2015**, *26*, 574–587. [CrossRef]
13. Li, J.; Ali, F.; Kim, W.G. Reexamination of the role of destination image in tourism: An updated literature review. *E-Rev. Tour. Res.* **2015**, *12*, 191–209.

14. Chon, K.-S. The role of destination image in tourism: A review and discussion. *Tour. Rev.* **1990**, *45*, 2–9. [[CrossRef](#)]
15. Fakeye, P.C.; Crompton, J.L. Image differences between prospective, first-time, and repeat visitors to the Lower Rio Grande Valley. *J. Travel Res.* **1991**, *30*, 10–16. [[CrossRef](#)]
16. Kim, D.; Perdue, R.R. The influence of image on destination attractiveness. *J. Travel Tour. Mark.* **2011**, *28*, 225–239. [[CrossRef](#)]
17. Reynolds, W.H. The role of the consumer in image building. *Calif. Manag. Rev.* **1965**, *7*, 69–76. [[CrossRef](#)]
18. Crompton, J.L. An assessment of the image of Mexico as a vacation destination and the influence of geographical location upon that image. *J. Travel Res.* **1979**, *17*, 18–23. [[CrossRef](#)]
19. Lai, K.; Li, X. Tourism destination image: Conceptual problems and definitional solutions. *J. Travel Res.* **2016**, *55*, 1065–1080. [[CrossRef](#)]
20. Gartner, W.C. Image formation process. *J. Travel Tour. Mark.* **1993**, *2*, 191–215. [[CrossRef](#)]
21. Rapoport, A. *Human Aspects of Urban Form*; Pergamon Press: Oxford, UK, 1977.
22. Baloglu, S.; McCleary, K.W. A model of destination image formation. *Ann. Tour. Res.* **1999**, *26*, 868–897. [[CrossRef](#)]
23. Beerli, A.; Martín, J.D. Factors influencing destination image. *Ann. Tour. Res.* **2004**, *31*, 657–681. [[CrossRef](#)]
24. Pocock, D.; Hudson, R. *Images of the Urban Environment*; Macmillan: London, UK, 1978.
25. Son, A. The measurement of tourist destination image: Applying a sketch map technique. *Int. J. Tour. Res.* **2005**, *7*, 279–294. [[CrossRef](#)]
26. Marine-Roig, E.; Anton Clavé, S. Perceived image specialisation in multiscalar tourism destinations. *J. Destination Mark. Manag.* **2016**, *5*, 202–213. [[CrossRef](#)]
27. Lynch, K. *The Image of the City*; The MIT Press: Cambridge, MA, USA, 1960.
28. Phelps, A. Holiday destination image—The problem of assessment: An example developed in Menorca. *Tour. Manag.* **1986**, *7*, 168–180. [[CrossRef](#)]
29. VisitBritain. *Researching and Planning: Foresight—Issue 150*; British Tourist Authority board: London, UK, 2017.
30. Genette, G. *Paratexts: Thresholds of Interpretation*; Cambridge University Press: New York, NY, USA, 1997.
31. Marine-Roig, E. Online travel reviews: A massive paratextual analysis. In *Analytics in Smart Tourism Design: Concepts and Methods*; Xiang, Z., Fesenmaier, D.R., Eds.; Springer: Heidelberg, Germany, 2017; pp. 179–202.
32. De Ascaniis, S.; Gretzel, U. Communicative functions of Online Travel Review titles: A pragmatic and linguistic investigation of destination and attraction OTR titles. *Stud. Commun. Sci.* **2013**, *13*, 156–165. [[CrossRef](#)]
33. Xiang, Z.; Wober, K.; Fesenmaier, D.R. Representation of the online tourism domain in search engines. *J. Travel Res.* **2008**, *47*, 137–150. [[CrossRef](#)]
34. Pan, B.; Xiang, Z.; Tierney, H.; Fesenmaier, D.R.; Law, R. Assessing the dynamics of search results in Google. In *Information and Communication Technologies in Tourism 2010*; Gretzel, U., Law, R., Fuchs, M., Eds.; Springer: Vienna, Austria, 2010; pp. 405–416.
35. Fesenmaier, D.R.; Xiang, Z.; Pan, B.; Law, R. A framework of search engine use for travel planning. *J. Travel Res.* **2011**, *50*, 587–601. [[CrossRef](#)]
36. Pan, B.; Fesenmaier, D.R. Online information search. Vacation planning process. *Ann. Tour. Res.* **2006**, *33*, 809–832. [[CrossRef](#)]
37. Marine-Roig, E. A webometric analysis of travel blogs and review hosting: The case of Catalonia. *J. Travel Tour. Mark.* **2014**, *31*, 381–396. [[CrossRef](#)]
38. Xiang, Z.; Gretzel, U. Role of social media in online travel information search. *Tour. Manag.* **2010**, *31*, 179–188. [[CrossRef](#)]
39. Baka, V. The becoming of user-generated reviews: Looking at the past to understand the future of managing reputation in the travel sector. *Tour. Manag.* **2016**, *53*, 148–162. [[CrossRef](#)]
40. W3Schools. HTML Tutorial. Available online: <http://www.w3schools.com/html/> (accessed on 7 July 2017).
41. Suthasupa, S. The portrayal of a city's image by young people. *Procedia-Soc. Behav. Sci.* **2013**, *38*, 284–292. [[CrossRef](#)]
42. Ylijoki, O.; Porras, J. Conceptualizing big data: Analysis of case studies. *Intell. Syst. Account. Financ. Manag.* **2016**, *23*, 295–310. [[CrossRef](#)]

43. Hu, H.; Wen, Y.; Chua, T.-S.; Li, X. Toward scalable systems for big data analytics: A technology tutorial. *IEEE Access* **2014**, *2*, 652–687.
44. Eurostat. Tourism. In *Eurostat Regional Yearbook 2016*; Kotzeva, M., Ed.; Publications Office of the European Union: Luxembourg, 2016; pp. 177–192.
45. De Borda, J.C. Mémoire sur les élections au scrutin. In *Mémoire de l'Académie Royale; Histoire de l'Académie des Sciences*: Paris, France, 1781; pp. 657–665.
46. Pantano, E.; Priporas, C.V.; Stylos, N. 'You will like it!' using open data to predict tourists' response to a tourist attraction. *Tour. Manag.* **2017**, *60*, 430–438.
47. Yoo, K.-H.; Sigala, M.; Gretzel, U. Exploring TripAdvisor. In *Open Tourism: Open Innovation, Crowdsourcing and Co-Creation Challenging the Tourism Industry*; Gger, R., Gula, I., Walcher, D., Eds.; Springer: Berlin, Germany, 2016; pp. 239–255.
48. TripAdvisor. Things to Do in Ile-de-France. Available online: https://www.tripadvisor.com/Attractions-g187144-Activities-Ile_de_France.html (accessed on 1 January 2017).
49. Stemler, S. An overview of content analysis. *Pract. Assess. Res. Eval.* **2001**, *7*, 137–146.
50. Stepchenkova, S. Content analysis. In *Handbook of Research Methods in Tourism: Quantitative and Qualitative Approaches*; Dwyer, L., Gill, A., Seetaram, N., Eds.; Edward Elgar Publishing: Cheltenham, UK, 2012; pp. 443–458.
51. ParisInfo. Sustainable Tourism in Paris. Available online: <https://en.parisinfo.com/discovering-paris/sustainable-tourism-in-paris> (accessed on 7 July 2017).
52. Jiang, Y.; Ramkissoon, H.; Mavondo, F.T.; Feng, S. Authenticity: The link between destination image and place attachment. *J. Hosp. Mark. Manag.* **2017**, *26*, 105–124. [[CrossRef](#)]



© 2017 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).