



Assessing Online Sustainability Communication of Italian Cultural Destinations – A Web Content Mining Approach

Valentina Marchi¹, Valentina Apicerni^{2(✉)}, and Alessandra Marasco²

¹ CNR-IBE, Via G. Caproni 8, 50145 Florence, Italy
valentina.marchi@ibe.cnr.it

² CNR-IRISS, Via G. Sanfelice 8, 80135 Naples, Italy
{v.apicerni, a.marasco}@iriss.cnr.it

Abstract. Online sustainability communication attracts a considerable attention in tourism research. This study focuses on sustainability communication in official destination websites for informing and motivating visitors to adopt sustainable practices and behaviors. To advance previous research in this area, it adopts a web content mining approach to assess the characteristics of online contents of a sample of 20 Italian cultural destinations. The main dimensions and typologies of sustainability-oriented practices in tourism are used as basis to develop a text classifier for the automated content analysis. A total of 2.975 web pages from official city websites and official tourism promotion websites of the destinations is analyzed through this approach to investigate the online contents relating to the environmental, economic, socio-cultural and general dimensions sustainability as well as their emotional appeal. The analysis reveals that about 15.8% of total online texts contains information to promote sustainability-oriented behaviors at the destination. It indicates that the communication is more specifically focused on environmental/economic/socio-cultural practices than generically referring to sustainable/responsible tourism. However, in line with previous research, it highlights that destinations do not sufficiently balance these pillars in their online communication. Further, the websites' texts scarcely leverage the persuasive potential of affective appealing messages. The implications of this automated approach for improving the design of online sustainability-related contents are also discussed.

Keywords: Online sustainability communication · Sustainable behaviors · Cultural destinations · Automated web content mining · Emotions

1 Introduction

One of the greatest challenges for sustainable tourism development is to encourage tourists to adopt sustainable behaviors in order to minimize the negative impacts of tourism on host communities, cultural heritage and the environment while providing its positive contribution to local economies [1–3]. In this direction, destination marketing and communications tools can be used to influence tourists' decision-making and

purchasing behaviors for more sustainable choices and practices [4, 5]. As the first touchpoint with the destination [6], tourism destination websites represent an important source for sustainability communication to travelers [7]. Official destination websites primarily serve the goal of providing useful information to potential visitors and promoting their products and services [8] and can influence tourists' behaviors by making them aware of destination's sustainable policies and products and stimulating pro-sustainable choices, actions and practices at the destination [9–11].

Although the topic of sustainability communication has been receiving an increasing attention in the last decade [7], there is still limited research on sustainability communication in official destination websites and further research is needed to advance the design of persuasive messages to engage visitors in sustainability behaviors [7, 11]. To contribute to fill this gap, this study aims to understand how to design effective sustainability messages for destinations online communication in order to promote visitors' sustainable behaviors. With respect to research based on manual approaches, it adopts an automated approach for analyzing the communication provided by official destination websites to inform about and motivate sustainable choices and practices by visitors. The final aim is to develop a reliable tool to perform automated analysis of online texts that could facilitate systematic, large-scale and comparable analyses of online communications. Previous research introduced automated approaches and techniques from data mining domain as effective methods for analyzing and evaluating destination websites [12, 13]. Gill et al. [14] performed an automated content analysis of sustainability communication in corporate websites, considering economic, environmental, and social indicators, in conjunction with semantic analysis in order to investigate level and type of sustainability reporting across firms. Whereas web automated analysis may not be appropriate to evaluate the audience impact of sustainability communication [15], it presents several benefits by allowing to analyze thousands of online texts in a speedy and accurate way and to reduce subjective interpretation in coding [14].

This study presents a web content mining analysis based on natural language processing (NLP) with a total set of 2.975 web pages by the official websites of the top 20 competitive Italian cultural destinations. The main dimensions and typologies of sustainability-oriented practices in tourism were used as basis to develop a text classifier for the automated analysis of sustainability-related contents in 39 official websites. Through emotional and sentiment analysis, the study also investigated the affective appeal of online sustainability-related contents as one of the main persuasive characteristics of communication identified by previous research. The paper presents the findings from the application of this automated approach and discusses its potential for measuring the effectiveness of official sustainability communication.

2 Related Work

Sustainability communication “sets out to make consumers aware of the availability of sustainable travel products, to inform consumers how these offerings meet their needs and comply with sustainability criteria, and ultimately to stimulate pro-sustainable purchases” [7, p. 10]. According to the recent literature review by Tölkes [7], tourism

organizations' websites are the most researched channels of sustainability communication. Smith and Font [15, 16] investigated the responsible tourism communication messages provided by volunteer tour operators. Through online content analysis, they analyze and score the organizations' webpages across 19 responsibility criteria (including donations, local conservation, respect heritage, respect wild-life, etc.). Overall, there is a lack of responsible marketing practices which represents a crucial feature of sustainability communication. Based on the analysis of Malaysian hotels, Joseph et al. [17] found that websites are not fully utilized to create awareness about sustainable development and mostly report economic information, followed by social and environment sustainability information. In the same line, Santos et al. [18] show that only a minority of hotels in Portugal explicitly practice sustainability communication, but they do it in a quite superficial way and using a rational appeal. With reference small and medium accommodation firms in the Azores, Tiago et al. [19] concluded that digital communication could be substantially improved and is linked to website sophistication.

Previous research also highlights that tourism organizations do not fully utilize their websites to motivate customers to behave more sustainably. In particular, Villarino and Font [20, p. 326] denoted a "sustainability marketing myopia", because communication tends to be mostly focused on products rather than on customers' needs. They identified four dimensions of persuasiveness in sustainability communication: type of action (including theme and beneficiary), structure (explicit versus implicit, active versus passive and denotative versus connotative), content (appeal versus logic, social norms and level of experience) and authority. Based on the analysis of sustainability contents in accommodations' websites, they highlighted the importance of communication specificity and the persuasive effects of emotional appealing and experiential messages to engage tourists in sustainable behaviors [20].

Literature in this area also investigated sustainability communication in official destinations' websites, with particular emphasis on its role and effectiveness in encouraging sustainable behaviors by prospective travelers. Pennington-Gray and Thapa [21] analyzed the role of websites for promoting culturally responsible behaviors based on a manual content analysis; they found that only a small number of destinations provided information related to cultural responsibility, rules or guideline. Based on the five pillars identified by the Sustainable Tourism for Development Guidebook by UNWTO, Garbelli et al. [22] performed a content analysis of online resources relating to Victoria Falls World Heritage Site; results showed there is room to improve the online communication to educate prospective travelers to behave in a sustainable and responsible manner during their visit to this UNESCO site. D'Angella and De Carlo [10] investigated the relationship between the orientation to sustainability in destinations' online communications and their strategic positioning. They proposed and applied Green D-web score and highlighted the importance of communication oriented to sustainability for destination competitiveness and for attracting new segments of environmentally sensitive tourists. A qualitative approach is adopted by Mura and Sharif [6] to perform a benchmarking of official websites of destinations in Southeast Asia. They stressed the important educative role of official websites for sustainability issues and recommended more content in this regard.

Finally, Ghanem and Elgammal [11] developed an online sustainability communication checklist for a web content analysis of the top 50 competitive destinations.

Their results indicated a lack of appropriate online approach to informing, motivating and engaging stakeholders in sustainability practices along with an unbalanced communication in relation to the three sustainability pillars. Based on their analysis, they call for more research to advance the design of effective online sustainability messages at destination level.

3 Research Methodology

This study adopted a web content mining approach to examine the characteristics of online contents provided by Italian destinations to inform and engage prospective visitors in sustainable practices. An automated approach is deemed to be useful to advance previous studies on sustainability that relied on manual web content analysis as well as to contribute to the development of reliable instruments for assessing the effectiveness of sustainability communication at destination level. In particular, the automated approach is used in this study to analyze contents communicated in official websites in relation to characteristics that the literature suggested may influence their persuasiveness. These include the communication of specific contents on topics relating to all the (environmental, economic, socio-cultural) sustainability dimensions [7, 11, 20] and the emotional appeal of contents [7, 18, 20].

3.1 Research Design

In the first stage of the research, literature was reviewed to identify a typology of sustainability dimensions and practices, which formed the basis for the development of a text classifier to be used for the analysis. The review also considered manuals and guidelines by international tourism organizations, including the *Sustainable visitor tips* by the World Monument Fund and the *Responsible traveler tips* by UNWTO. Further, the sample of destinations' websites was identified. In detail, the analysis focused on the official websites of the top 20 Italian cultural destinations, which collectively accounted for about 80% of international arrivals (almost 47 million) and 75% of nights spent in accommodation (more than 121 million) in 2018 [23]. The destinations in this sample include Italian cities classified by Italian Institute of Statistics [24] as cultural destinations that have been presenting positive growth rates of tourist arrivals in 2018. This sample was deemed appropriate to explore online sustainability communication, also in consideration of the relevance of online communication for sustainability in relation to cultural heritage sites and destinations [22]. In the second stage, based on the keywords identified in the first stage a text classifier was created for the automated analysis of online sustainability-related contents. Finally, data collection and analysis were performed as detailed below.

3.2 Data Collection and Analysis

The study used two sources of online sustainability content provided by the 20 Italian destinations in our sample: i) official city websites (20); ii) official tourism promotion websites (19). In the first step, the researchers selected 39 web sites through a manual

preliminary screening. Contents were then collected by using a web scraping procedure, which is a technique to extract data without the need of a user interaction [25]. It retrieved all texts and subpages directly accessible from web pages [26]. In order to focus the analysis on the main visible and accessible sustainable contents, the study excluded those not linked to the homepage. The web scraping procedure was performed using R software [27] on the 39 websites in Italian language during the period February–April 2020. In a first stage, 4,058 web pages were extracted. A pre-processing operation was carried out with the aim of deleting duplicated results and additional noise. The final database was composed of 2,975 web pages.

In the second step, the study adopted a content analysis technique to detect words related to sustainability practices and behavior in the online communications. For the purpose of the study, a dictionary to detect messages targeting visitors for promoting sustainable/responsible tourism was built. A semi-automatic dictionary was created following the proposed scheme of Deng et al. [28] in order to combine the researcher's knowledge with the assistance of text analysis software (Fig. 1). Computational text analysis was performed using R, a statistical opensource software, which allows replicability. The corpus creation, which consists in the set of documents on which the dictionary is developed [28], was based on different textual contents identified: i) guidelines by international tourism organizations and scientific articles analyzed through a manual analysis; ii) web sites of the 20 Italian cultural destination manually analyzed in the preliminary phase; iii) 2,975 web pages of the 20 Italian cultural destination extracted using web scraping procedure and processed through an automated text analysis. Afterwards, a pre-processing operation was carried out with the aim of removing stop word, unnecessary information and reducing all words to lowercase. Moreover, the cut-off criteria were applied, based on term frequency that shows how certain concepts occur in the text. Later, the study focused on the identification of categories and categorizing entries. In line with previous research on sustainability and sustainability communication [11, 20] the researchers included the four following categories recognized as the main dimensions for sustainable tourism: economic, socio-cultural, environmental and general.

Keywords identified consist in word lists, without term overlaps between the four sustainability dimensions, composed by unigram, bigram and trigram. Researchers manually reviewed the words and phrases emerging from each category. Synonym words were manually detected and included in the dictionary. This step generated new words. For example in the economic category were included also terms like 'Km0' to indicate farmers markets. The validation method of the dictionary considered the Keyword-in-context (KWIC) Indexing. KWIC is an automatic system that allows to search a particular keyword in the text and analyze its local meaning in relation to a number of words immediately preceding and following it [29]. It was performed using the R package called 'Quanteda' for managing and analyzing text [30]. The final sustainable dictionary was composed of 124 sustainable keywords.

In the third step, a preliminary sentiment analysis was carried out to understand the characteristics of contents published online to promote sustainable practices by the Italian destinations. The study adopted the OpenNER Sentiment Lexicon developed in six languages [31], including Italian, thus allowing an applicability of the analysis even on a sample in different languages. The research applied the Italian Sentiment Lexicon,

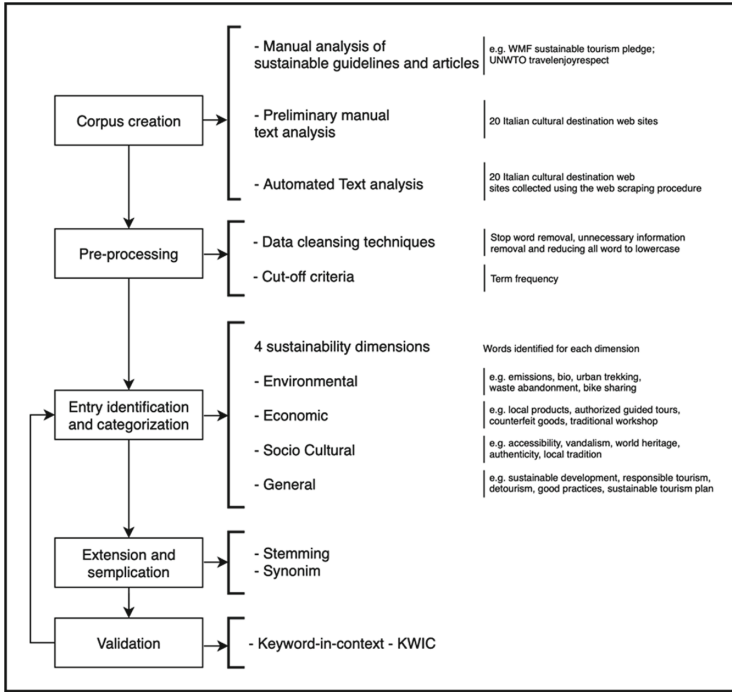


Fig. 1. The semi-automatic dictionary building process (S-DBP) implemented to build a dictionary on sustainable tourism following the design scheme proposed by Deng et al. [28].

which allows to analyze the sentiment polarity classification (positive, negative and neutral).

Further, the study developed an ad-hoc emotional dictionary with the aim to explore more in-depth the affective appeal of sustainability-related contents. This dictionary includes emotional and experiential keywords identified through a preliminary manual text analysis and an automated text analysis. The words in this dictionary are also included in the NRC Emotion Lexicon [32], which consists of a list of English words and their association with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy and disgust) and two sentiments (positive and negative). This lexicon, also known as EmoLex, has been previously used to detect the emotions conveyed by destinations' online texts [33].

4 Results

Table 1 presents the results relating to the frequency count of sustainability-related words, which is calculated as the total number of words for each of the four dimensions in the total texts. It also reports the number of web pages containing sustainability-related words and the percentage of web pages on the total pages containing words related to the four sustainability dimensions. Overall, the analysis revealed that 15.8%

of the total online texts contains information to promote sustainability-oriented behaviors at the destination. It further shows that environmental sustainability has the highest frequency of words, followed by the economic dimension. These results align with previous studies reporting that online information is mostly concerned with environmental sustainability, overlooking the other two dimensions [11, 14, 19, 20]. Further, they indicate that the communication is more specifically focused on environmental/economic/socio-cultural practices (14.59% of the total web pages) than generically referring to sustainable/responsible tourism (1.21% of the total web pages), which can be considered a relevant characteristic of credible and thus persuasive communication according to Villarino and Font [20].

Table 1. Frequency of sustainability-related words across dimensions

| Dimensions | Word frequency | Text per topics | % Text per topics |
|----------------|----------------|-----------------|-------------------|
| Environmental | 342 | 193 | 6.49 |
| Economic | 329 | 191 | 6.42 |
| Socio-Cultural | 159 | 124 | 4.17 |
| General | 116 | 68 | 2.29 |

Table 2. Frequency of sustainability-related words by city

| Cities | Env. | Eco. | Soc. | Gen. | TOT | MEAN | SD |
|---------|-------|-------|-------|-------|-------|-------|--------|
| Assisi | 0.020 | 0.016 | 0.007 | 0.010 | 0.053 | 0.013 | 0.0059 |
| Bari | 0 | 0.025 | 0 | 0.031 | 0.056 | 0.014 | 0.0164 |
| Bologna | 0.011 | 0.025 | 0.039 | 0.003 | 0.078 | 0.020 | 0.0159 |
| Como | 0.013 | 0.022 | 0.002 | 0.009 | 0.046 | 0.011 | 0.0083 |
| Firenze | 0.099 | 0.021 | 0.007 | 0.021 | 0.148 | 0.037 | 0.0419 |
| Genova | 0.030 | 0.092 | 0.020 | 0.008 | 0.150 | 0.038 | 0.0374 |
| Milano | 0.037 | 0.026 | 0.016 | 0.006 | 0.085 | 0.021 | 0.0133 |
| Napoli | 0.011 | 0.009 | 0.017 | 0.017 | 0.054 | 0.013 | 0.0041 |
| Padova | 0.104 | 0.044 | 0.009 | 0.044 | 0.201 | 0.050 | 0.0394 |
| Palermo | 0.005 | 0.005 | 0.024 | 0.005 | 0.039 | 0.010 | 0.0095 |
| Parma | 0.018 | 0.022 | 0.004 | 0.018 | 0.062 | 0.016 | 0.0079 |
| Perugia | 0.009 | 0.028 | 0.003 | 0.004 | 0.044 | 0.011 | 0.0116 |
| Pisa | 0.011 | 0.006 | 0.003 | 0.005 | 0.025 | 0.006 | 0.0034 |
| Ravenna | 0.021 | 0.038 | 0.013 | 0.002 | 0.074 | 0.018 | 0.0152 |
| Roma | 0.022 | 0.033 | 0.012 | 0.007 | 0.074 | 0.019 | 0.0115 |
| Siena | 0.121 | 0.023 | 0.052 | 0.006 | 0.202 | 0.050 | 0.0507 |
| Torino | 0.036 | 0.011 | 0.013 | 0.006 | 0.066 | 0.017 | 0.0133 |
| Trieste | 0.074 | 0.115 | 0 | 0 | 0.189 | 0.047 | 0.0571 |
| Venezia | 0.036 | 0.022 | 0.015 | 0.032 | 0.105 | 0.027 | 0.0095 |
| Verona | 0.034 | 0.026 | 0.010 | 0.007 | 0.077 | 0.019 | 0.0129 |

The number of frequencies in each text are divided by the total number of words of each text to avoid that longer texts are over-represented.

This is confirmed by the results relating to the frequency count by city (Table 2), which also highlights a small group of cities with a higher frequency of sustainability-related words (Firenze, Genova, Padova, Siena, Trieste and Venezia). Among these, it is interesting to note that Genova and Trieste present high percentages of economic sustainability-related words, as they often refer to the promotion of local products food, markets, as well as traditional workshops.

However, the analysis by destination shows that communication is mostly unbalanced across the dimensions, confirming evidence provided by Ghanem and Elgammal [11] about the lack of an effective approach balancing all sustainability pillars to better inform, motivate, and engage stakeholders in sustainable tourism.

Table 3 provides a detailed analysis of the contents that are most frequently communicated in relation to the four dimensions of sustainability. Sustainability-related words concerning the environment are mainly related to green practices and areas, while those relating to the economic pillar concern the specific products/providers for encouraging to shop locally. The promotion of excursions and authorized guided tours to discover destinations emerge in the communication related to the environmental and economic dimensions, respectively. In relation to the socio-cultural dimension, online texts most frequently refer to the authenticity of the resources at the destination and their preservation. This dimension also includes information about accessibility of hospitality services and attractions along with recommendations/codes of conduct that are aimed to prevent damaging behaviors by visitors (e.g., vandalism, graffiti). The last dimension includes words relating to general sustainability practices (e.g., Detourism, good practices, sustainable development) and awareness campaigns promoting sustainability.

Table 3. Top 5 sustainability-related words across dimensions

| Environment | % | Economic | % | Socio-cultural | % | General | % |
|-----------------|-------|-------------------------|-------|-------------------|-------|---------------------|-------|
| Excursions | 28.07 | Authorized guided tours | 45.90 | Authenticity | 14.46 | Sustainability | 21.55 |
| Green areas | 14.32 | Local products | 14.89 | Preservation | 13.84 | Good practices | 14.65 |
| Green practices | 12.86 | Local food | 13.07 | Cultural Heritage | 11.94 | Sust. Develop. | 11.21 |
| Cycle paths | 9.94 | Traditional workshops | 11.85 | Accessibility | 8.17 | Awareness campaigns | 10.34 |
| Bike sharing | 7.31 | Local markets | 4.56 | Damage | 5.03 | Detourism | 8.62 |

Table 4 and 5 present the results of the sentiment analysis, indicating the occurrences of positive, negative and neutral words in sustainability-related online texts across sustainability dimensions and by city, respectively. Overall, the results show that destinations do not take much advantage of an emotionally appealing language, in line with previous research [20]. Looking more in detail at the results of the sentiment analysis by destination (Table 5), some cases emerge presenting a higher percentage of

positive sentiment along with a lower percentage of negative sentiment, notably Firenze, Perugia and Trieste.

Table 4. Sentiment analysis of sustainability-related contents across dimensions

| Dimensions | Positive | % | Negative | % | Neutral | % |
|----------------|----------|------|----------|------|---------|------|
| Environmental | 14845 | 1.89 | 3163 | 0.40 | 29909 | 3.81 |
| Economic | 16971 | 1.85 | 3787 | 0.41 | 34284 | 3.74 |
| Socio-Cultural | 17938 | 1.85 | 3978 | 0.41 | 36015 | 3.71 |
| General | 18815 | 1.86 | 4200 | 0.42 | 37545 | 3.71 |

Table 5. Sentiment analysis of sustainability-related contents by destination

| Cities | Positive (%) | Negative (%) | Neutral (%) |
|---------|--------------|--------------|-------------|
| Trieste | 2.23 | 0.22 | 3.23 |
| Firenze | 2.17 | 0.49 | 4.16 |
| Perugia | 2.16 | 0.34 | 3.23 |
| Bari | 1.76 | 0.50 | 3.99 |
| Verona | 1.51 | 0.35 | 4.52 |
| Pisa | 1.32 | 0.22 | 2.79 |

Table shows the three cities with the highest percentage of positive sentiment and the three cities with the lowest negative score

Finally, the research developed an emotional dictionary, which was tested on web pages coded as *no sustainable* and on those pages coded as *sustainable*, based on the sustainable dictionary previously created. On the pages regarding sustainability there is a greater use of emotional words. The result is statistically significant with a confidence level of 99% (p-value = 0.006943).

Table 6 reports the top emotional words, based on their frequency count in relation to the emotional dictionary. Based on EmoLex [32], they provide indications of the range of positive emotions associated the texts, including surprise (“unique”), joy (“beautiful”, “happy”), trust (“responsible”, “authentic”), anticipation (“passionate”).

The analysis on the use of emotional words does not reveal substantial differences between the four dimensions of sustainability. Table 7 shows the top five emotional-related words that are mostly communicated, and highlighted that, although most of the emotional words are common between dimensions, yet their frequency is different. Measuring emotions through specific words has limits, as highlighted by Zhang and Fesenmaier [33, p. 87], because emotions are also expressed through some important experiential connotations that may not be immediately grasped with dictionary.

Table 6. Top 10 emotional words

| Words | % |
|---------------|-------|
| Unique | 19,08 |
| Beautiful | 18,10 |
| Discovery | 15,38 |
| Extraordinary | 8,07 |
| Respect | 7,63 |
| Passionate | 3,05 |
| Pleasant | 2,94 |
| Difficult | 2,73 |
| Responsible | 2,62 |
| Authentic | 2,62 |

Table 7. Top 5 emotional-related words across dimensions

| Dimensions | Top 5 emotional words |
|----------------|--|
| Environment | Unique (20.9%), Discovery (15.2%), Beautiful (12.9%), Responsible (8.1%), Respect (7.5%) |
| Economic | Discovery (19.2%), Beautiful (16.6%), Unique (14.1%), Respect (7.6%), Extraordinary (7.1%) |
| Socio-Cultural | Unique (18.3%), Responsible (15.9%), Discovery (15.6%), Beautiful (7.6%), Respect (6.7%) |
| General | Unique (20.3%), Discovery (18.2%), Responsible (12.3%), Participation (8.0%), Extraordinary (8.0%) |

5 Conclusions and Future Research

This study aims to contribute to research on online sustainability communication through a first attempt in the direction of developing a reliable tool for the automated analysis of online texts that could advance research based on manual approaches. When validated, such an instrument could facilitate systematic, large-scale and comparable analyses of online communications to understand how effectively destinations communicate to create awareness on and encourage sustainable behaviors through their websites as well as provide insights for the design of messages to improve their persuasiveness. Future research is needed to overcome the limits of this study, especially related to the size and the composition of the sample that includes only few Italian urban destinations. In this study, 124 sustainability related words have been identified; however, further testing on online texts relating to other destinations is necessary to assess the validity of the dictionaries and consequently refine them. Related to this issue is the challenge of multilinguality. The present study analyzed only online contents written in Italian language. Although “focusing on the largest languages is not a sustainable strategy in an increasingly multilingual world” [34, p. 430], future research should address the need to analyze online contents in other languages. Moreover, it would be interesting to test the dictionaries on tourism sustainability promotion contents and campaigns posted by DMOs on social networks. Finally, text-based sentiment analysis

should take into account some important experiential connotations that are not immediately captured through the dictionary. Future research should be oriented to create an ad-hoc emotional dictionary for sustainability.

References

1. Budeanu A (2007) Sustainable tourist behaviour—a discussion of opportunities for change. *Int J Consum Stud* 31(5):499–508
2. Buonincontri P, Marasco A, Ramkissoon H (2017) Visitors' experience, place attachment and sustainable behaviour at cultural heritage sites: a conceptual framework. *Sustainability* 9(7):1–19
3. UNEP, UNWTO (2005) *Making Tourism More Sustainable – A Guide for Policy Makers*. UNEP, Paris, WTO, Madrid
4. Font X, McCabe S (2017) Sustainability and marketing in tourism: Its contexts, paradoxes, approaches, challenges and potential. *J Sustain Tour* 25(7):869–883
5. Hanna P, Font X, Scarles C, Weeden C, Harrison C (2018) Tourist destination marketing: From sustainability myopia to memorable experiences. *J Destin Market Manage* 9:36–43
6. Mura P, Sharif SP (2015) Exploring rural tourism and sustainability in Southeast Asia through the lenses of official tourism websites. *Worldwide Hospital Tour Themes* 7(5):440–452
7. Tölkes C (2018) The role of sustainability communication in the attitude–behaviour gap of sustainable tourism. *Tour Hospital Res* 20(1):117–128
8. Teichmann K, Zins AH (2009) Information elements on DMO-websites: Alternative approaches for measuring perceived utility. In: O'Connor P, Höpken W, Gretzel U (eds) *Information and communication technologies in tourism*. Springer, Vienna, pp 209–219
9. Ali A, Frew AJ (2014) Technology innovation and applications in sustainable destination development. *Inf Tech Tour* 14(4):265–290
10. d'Angella F, De Carlo M (2016) Orientation to sustainability and strategic positioning of destinations: an analysis of international tourism websites. *Curr Issues Tour* 19(7):624–633
11. Ghanem M, Elgammal E (2017) Communicating sustainability through a destination's website: a checklist to inform, motivate, and engage stakeholders. *J Travel Tour Market* 34(6):793–805
12. Költringer C, Dickinger A (2015) Analyzing destination branding and image from online sources: a web content mining approach. *J Bus Res* 68(9):1836–1843
13. Law R, Qi S, Buhalis D (2010) Progress in tourism management: a review of website evaluation in tourism research. *Tour Manag* 31(3):297–313
14. Gill D, Dickinson S, Scharl A (2008) Communicating sustainability: a web content analysis of North American, European and Asian firms. *J Commun Manage* 12(3):243–262
15. Smith VL, Font X (2014) Volunteer tourism, greenwashing and understanding responsible marketing using market signalling theory. *J Sustain Tour* 22(6):942–963
16. Smith VL, Font X (2015) Marketing and communication of responsibility in volunteer tourism. *Worldwide Hospital Tour Themes* 7(2):159–180
17. Joseph C, Lin VCS, Nichol EO, Jussem PM (2014) Sustainability disclosure on Malaysian hotel websites. In: *Proceedings conference: international conference on governance (ICG 2014)*, vol 29, Universiti Utara, Malaysia
18. Santos MC, Veiga C, Águas P, Santos JAC (2019) Sustainability communication in hospitality in peripheral tourist destinations. *Worldwide Hospital Tour Themes* 11(6):660–676

19. Tiago F, Gil A, Stemberger S, Borges-Tiago T (2020) Digital sustainability communication in tourism. *J Innovat Knowl* (article in press)
20. Villarino J, Font X (2015) Sustainability marketing myopia: The lack of persuasiveness in sustainability communication. *J Vacation Market* 21(4):326–335
21. Pennington-Gray L, Thapa B (2004) DMOs and culturally responsible behaviors: an exploratory analysis. *Tourism* 52(2):183–194
22. Garbelli M, Adukaite A, Cantoni L (2015) Communicating tourism sustainability online the case of victoria falls world heritage site. *E-Rev Tour Res* 6:1–5
23. Apicerni V, Marasco A (2019) Dalla competitività delle città d'arte italiane alla sfida dell'innovazione smart per il turismo culturale In: *Rapporto sul Turismo Italiano - XXIII Edizione 2018/2019*. Rogosi Editore, Napoli, pp 341–360
24. ISTAT (2015) *Classificazione dei comuni per circoscrizione e tipo di località turistica*
25. Sozzi A (2017) Measuring sustainability reporting using web scraping and natural language processing, conference paper. NTTS
26. Bianchi G, Bruni R, Scalfati F (2018) Identifying e-commerce in enterprises by means of text mining and classification algorithms. *Math Prob Eng* 2018:1–8
27. Krotov V, Tennyson M (2018) Research note: scraping financial data from the web using the R language. *J Emerg Tech Acc* 15(1):169–181
28. Deng Q, Hine MJ, Ji S, Sur S (2019) Inside the black box of dictionary building for text analytics: a design science approach. *J Int Tech Inf Manage* 27(3):119–159
29. Luhn HP (1960) Key word-in-context index for technical literature (kwic index). *Am Doc* 11(4):288–295
30. Benoit K, Watanabe K, Wang H, Nulty P, Obeng A, Müller S, Matsuo A (2018) Quanteda: an R package for the quantitative analysis of textual data. *J Open Source Softw* 3(30):774
31. Maks I, Izquierdo R, Frontini F, Agerri R, Azpeitia A, Vossen P (2014) Generating polarity lexicons with WordNet propagation in five languages. In: *Proceedings of the 9th international conference on language resources and evaluation (LREC 14)*. European Language Resources Association (ELRA), Reykjavik, pp 1155–1161
32. Mohammad SM, Turney PD (2013) *NRC Emotion Lexicon*, NRC Technical report. <http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm>
33. Zhang W, Fesenmaier DR (2018) Assessing emotions in online stories: comparing self-report and text-based approaches. *Inf Tech Tour* 20(1–4):83–95
34. Karlgren J, Sahlgren M, Olsson F, Espinoza F, Hamfors O (2012) Usefulness of sentiment analysis. In: *Baeza-Yates R et al (eds) European conference on information retrieval*. Springer, Heidelberg, pp 426–435

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

