Journal of King Saud University – Computer and Information Sciences 34 (2022) 10125–10144

HOSTED BY

ELSEVIER

Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

Natural language processing applied to tourism research: A systematic review and future research directions

Miguel Á. Álvarez-Carmona [a], Ramón Aranda [b], Ansel Y. Rodríguez-Gonzalez [c], Daniel Fajardo-Delgado [d], María Guadalupe Sánchez [d], Humberto Pérez-Espinosa [c], Juan Martínez-Miranda [c], Rafael Guerrero-Rodríguez [e], Lázaro Bustio-Martínez [f], Ángel Díaz-Pacheco [g,*]

[a] *Centro de Investigacion en Matemáticas, Sede Monterrey, Km. 10 Autopista al Aeropuerto, Parque de Investigación e Innovación, Apodaca, Nuevo León, 66628 México, Investigador por México*
[b] *Centro de Investigación en Matemáticas, Sede Merida, Carretera Sierra Papacal Chuburna Puerto Km 5, Papacal, Yucatán, 97302 México, Investigador por México*
[c] *Unidad de Transferencia Tecnológica Tepic, Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE-UT3), Andador 10 Tepic, Nayarit 63173, México*
[d] *Departamento de Sistemas y Computación, Instituto Tecnológico de Ciudad Guzmán, Av. Tecnológico 100, Mpio. de Zapotlán el Grande, Jalisco 49000, México*
[e] *Departamento de Gestión y Dirección de Empresas, División de Ciencias Económico-Administrativas, Universidad de Guanajuato, Fraccionamiento 1, Col. El Establo S/N. Guanajuato, Guanajuato 36250, México*
[f] *Universidad Iberoamericana Ciudad de México, Prolongación Paseo de Reforma 880, CDMX, 01219, México*
[g] *Departamento de Ingeniería Electrónica, División de Ingenierías, Campus Irapuato-Salamanca, Universidad de Guanajuato, Carretera Salamanca-Valle de Santiago km 3.5, Comunidad de Palo Blanco, 36885, Salamanca, Guanajuato, México*

A R T I C L E I N F O

Article history:
Received 1 July 2022
Revised 16 September 2022
Accepted 12 October 2022
Available online 15 October 2022

Keywords:
Tourism
Natural language processing
Systematic review

A B S T R A C T

The social networks and the rapid development of new technologies have led to considerable changes in the tourism industry. Artificial intelligence, in particular natural language processing (NLP), presupposes a significant advantage in obtaining information on the mass content generated by online users concerning tourism services and products. This work presents a systematic review of the use of NLP in the tourism industry and research. We used the well-known PRISMA methodology, and 227 relevant studies over the last decade have been reviewed. Our analysis identified the main methodologies, tools, data sources, and other relevant features in the field. One of the principal contributions of this study is a taxonomy for using NLP in tourism. In addition, metadata were examined using a threefold approach: (i) general statistics, (ii) abstract text analysis, and (iii) keyword networks. Automatic analyses have identified six major topics in applying NLP to tourism issues and have shown that China, the United States, Thailand, and Spain share similar tourism issues or approaches.

Contents

* Corresponding author.
E-mail address: isc.diazpacheco@gmail.com (Á. Díaz-Pacheco).

Peer review under responsibility of King Saud University.

ELSEVIER Production and hosting by Elsevier

https://doi.org/10.1016/j.jksuci.2022.10.010

1. Introduction

Based on the World Economic Forum report, in 2018, the travel & tourism industry generated 10.4% of the world GDP and supported over 319 million jobs (Uppink and Soshkin, 2022; Calderón and Blanco, 2017; Alvarez-Carmona et al., 2022c) stated that 75% of international travelers plan their trips by digital means, and a big part of their decisions rely on other travelers' shared online information. To increase the revisit intentions of this industry's customers, it is essential to meet their expectations. Big Data analytics has been playing a crucial role in coping with this need (Marr, 2016; Gupta et al., 2017; Nadkarni et al., 2019). These massive quantities of Online User-Generated Content (UGC) can be an asset for companies and governments to define tourism-related strategies (Li et al., 2019; Xiang et al., 2015).

It is essential to use algorithms from the Artificial Intelligence field to handle UGC, specifically from the Natural Language Processing area (NLP). According to Nadkarni et al. (2011), NLP began in 1950 because of the interest in intersecting artificial intelligence and linguistics. NLP covers many artificial intelligence methods to analyze and represent naturally occurring text at one or more linguistic examination levels. This sub-field of artificial intelligence aims to achieve human-like processing capabilities of the language for diverse scopes (Liddy, 2001; Cai et al., 2016; Chowdhury, 2003).

Some recent NLP applications have been used to improve strategies in tourism like sentiment analysis, chatbots, and hotel recommendations, among others (Putri et al., 2019; García-Pablos et al., 2016; Prameswari et al., 2017a; Bulchand-Gidumal, 2020). Over the past decade, many research articles on tourism solutions have been published. One strategy to summarize the information of the research articles is utilizing a Systematic Review (SR). SR is a powerful method to structure knowledge. Thus, we identified several reviews that explored a wide range of tourism-scope challenges (Loyola-González et al., 2020; Rodríguez-Ruiz et al., 2019). For example, Weed (2006) reviewed 80 articles on the sports-tourism matter to identify the state-of-the-art in the field. The methodology employed for review was developed by the author and is precisely explained to allow replicability.

Anderson et al. (2015) investigated the link between tourism and the introduction of non-native species in marine environments, revising 32 articles. The results showed that non-native species are statistically higher in touristic sites. They employed the Collaborative Environmental Evidence methodology. Gomezelj (2016) examined the state of research on innovation in tourism revising 152 papers. They performed a bibliometric analysis to show the theoretical foundations of the studies. With clusters of the co-citation in the documents, they identified some trends that characterize the field. They do not mention the review methodology employed. Li et al. (2018) reviewed 144 articles on using the Big Data paradigm in tourism research. They wanted to understand the full-scale types of big data, data characteristics, analytic techniques, and touristic issues addressed in this field. The methodology followed for the review was not specified, but the implemented procedure is well explained for replicability. Marasco et al. (2018) studied the collaborative innovation in tourism with a review of 79 articles. They considered the studies' location, the perspective of analysis, the employed methodology, the level of research, and the specific themes addressed. The reviewing methodology's name was not specified, but the process was thoroughly explained. Reyes-Menendez et al. (2019) scrutinized 17 articles to understand the phenomena of fake text reviews on digital platforms and how it has been addressed in tourism. They employed the PRISMA methodology and concluded that further research on alternative approaches' impact is necessary to detect fake online reviews for the tourism business. Wattanacharoensil and La-ornual (2019) investigated cognitive bias in the tourist's decision-making. Thirty-seven articles were reviewed. The authors designed the employed methodology, which is detailed in the study. Chi et al. (2020) reviewed 63 studies to comprehend how Artificial Intelligence technology has been incorporated into the service sector. The methodology proposed for systematic reviews was employed by Tranfield et al. (2003) for systematic reviews. They concluded that AI research is still in its childhood, particularly in services. Samara et al. (2020) evaluated the benefits and challenges of incorporating Big Data and Artificial Intelligence in the hospitality and tourism industry. They analyzed 102 articles. The systematic review methodology proposed by Siddaway (2014) was employed in that work. They discovered that the overall travel experience could become richer at every level pre-, post-, and during the trip.

Based on the above, studies presented by Li et al. (2018), Reyes-Menendez et al. (2019), and Samara et al. (2020) explored technological approaches in hospitality and tourism, such as the use of the Big Data paradigm (applications and challenges) and the fake reviews phenomena in the User Generated Content. However, our research aimed to assess state-of-the-art NLP applications to address tourism-related issues. Due to the IT perspective presented in this work, we wanted to analyze technical aspects, areas of opportunity, and challenges. We believe our work can provide practical information for researchers who do not have experience in the field and may struggle to search for data sources, identify the problems that NLP can solve in tourism, or determine the method for the situation. In addition, our analysis can benefit field experts in search of new research niches.

In order to know how NLP is used in the tourism industry, the edge techniques, and where data is obtained, this work presents a systematic review of 227 studies from the last decade. Our research efforts are intended to assist two significant groups. The first is for professionals and young researchers looking for a starting point in this field. In this sense, we summarize the available data sources, popular approaches, and pre-processing methods for text. The second group is for researchers in the sector looking for new directions and areas of opportunity. As a result, we have proposed novel approaches and challenges to tourism and NLP research.

This study is organized as follows: In Section 2, our methodology to conduct the systematic review is detailed to allow replicability. Section 3 presents a general background of the NLP, a summary of the outstanding work reviewed, and the analyses conducted on meta-data derived from this study. Section 4 presents the discussion and answers the research questions. Section 5 shows the opportunity areas and challenges of the NLP in tourism, and Section 6 is for the conclusions and consideration for future research.

2. Material and methods

For this systematic review, we rely on the *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA) (Moher et al., 2009; Liberati et al., 2009). This methodology comprises elements for meta-analysis that focus on reporting research on randomized trials and can also structure systematic reviews. In this sense, to comprehend the current status of NLP research in hospitality and tourism and to provide crucial information for beginners in the area, we propose the following research questions:

1. In which tourism issues are the NLP techniques applied?
2. Which state-of-the-art NLP algorithms/methods are applied for tourism issues?
3. What information sources (data) are required to apply NLP techniques for tourism issues?

Using the research questions as our guide for this study, we selected only the documents that could answer at least one of them. Additionally, we discard those with a smaller length than three pages. We employed this **minor exclusion criterion** to reject studies like workshop briefs, congress posters, and short papers because they present ongoing research. Regarding the **information sources**, all documents were searched in online databases. Only research in English was considered, and the databases were queried for documents indexed from January first, 2010, until December thirty-first, 2020. The search was conducted on the following academic databases: ScienceDirect, Wiley Online Library, Sage Journals, the Web of Science, Emerald, and Scopus. The **search query** considered the use of several keywords aside from "Tourism" and "Natural Language Processing" as "travel", "visit", "hospitality", "hotel", and "destination" for the tourism field. Regarding NLP, other keywords employed were: "text mining", "sentiment analysis", "chatbot", "UGC" (User Generated Content), and "consumer review". Keywords concatenation was set with "OR" operator, and operator "AND" was employed for the intersection of the keywords of the considered areas. The search retrieved 500 records. ScienceDirect yielded 124 studies, 22 from Wiley Online Library, 70 from Sage Journals, 124 from the Web of Science, 45 from Emerald, and 168 from Scopus. Fifty-two duplicated records were removed to have 448 documents stored for eligibility. The exclusion criteria rejected 98 records. One hundred twelve

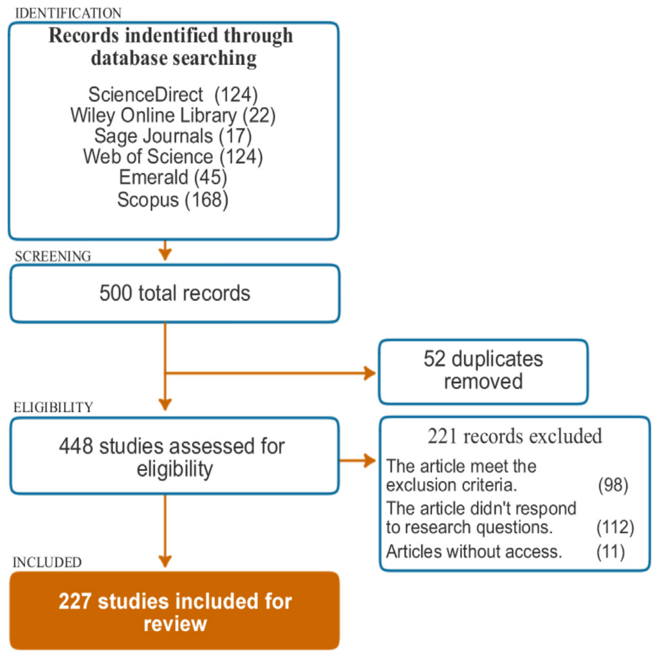

Fig. 1. Flowchart of the review process.

records were discarded because they did not respond to the research questions; 11 were not assessed because they were not written in English (Chinese). A total of 227 studies were included in the review process. For a more straightforward interpretation, the flowchart of the process is presented in Fig. 1.

For the **study selection** stage, the first two reviewers assessed the suitability of the documents based on the minor exclusion criterion (detailed above). Then, all the studies not discarded were separately assessed by two different reviewers. Inclusion was based on a full-text review. Disagreements were settled by consensus. The data extracted or **data items** from each study is as follows: Keywords, publication year, country of the first author, tourism issues addressed, database from which the article was obtained, countries considered in the study, language or languages of source data, obtained results, source of the database analyzed in the article, number of records in the database, data preprocessing techniques, feature selection algorithms, data representations, techniques for the proposed solution, and evaluation metrics.

2.1. Descriptive Analysis

Based on the data collected from the studies, we have built up many graphs to understand the different perspectives and issues addressed. Research protocols, performance measures, and approaches were too different among studies to allow comparisons and other analyzes; therefore, descriptive statistics answer the research questions in this study.

As a first step, we wanted to understand the evolution of adopting NLP approaches in tourism. Fig. 2 shows the number of studies per year. Considering the increase in our digital participation[1] over the last decade, it is not surprising that NLP has gained popularity in tourism research (with a significant increase from 2014). To plan their trips, many travelers use apps, read other travelers' opinions online, get suggestions from recommendation systems, and write about their experiences (Guerreiro et al., 2019; Xiang et al., 2015).

[1] The term digital participation refers to the active involvement in digital society through the use of modern information and communication technology (Seifert and Rössel, 2019).

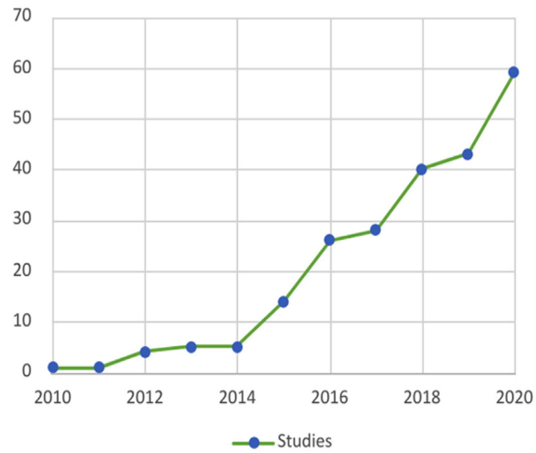

Fig. 2. Contributions per year.

NLP plays a crucial role in how we experience our vacations, providing practical information for travelers and business intelligence for the tourism industry.

Another important aspect is the language examined with NLP. Studies in this field need to gather text samples to build models. We ranked the top 10 languages of the different corpus implemented in the literature (see Fig. 3). While many studies focus on domestic tourism and China produces more studies, English is the most widely used language. One cause of this could be the status of English as a lingua franca (Zahedpisheh et al., 2017; Akopyants et al., 2017; Jocuns, 2018), and the availability of tools and frameworks to analyze this language. However, it is highly remarkable that the second most crucial language was Chinese.

As for the technical aspects, we summarized four important elements of the NLP process used in the literature: preprocessing techniques, representation methods, machine learning algorithms, and performance metrics. The top three most used preprocessing methods in the literature are *Remove stop words*, *Tokenization*, and *Stemming/Lemmatization* (see Fig. 4).

According to Hardeniya et al. (2016), the stages mentioned above of text preprocessing are as follows. The removal of *stop words* refers to deleting all the non-informative words such as determiners or prepositions from the text. *Tokenization* means splitting the raw text string into meaningful units (words) that can be used in further analysis. *Stemming* and *Lemmatization* attempt to transform all the grammatical/inflected forms of a word into its root. The main difference between both approaches is that Stemming uses a set of rather simplistic rules to chop a word to reduce the variations, while lemmatization uses context and part of speech techniques to determine the inflected forms that need transformation. Preprocessing, in general, removes non-informative words like articles, punctuation, numbers, emoticons, and hashtags, among others (Pinarbasi and Taskiran, 2020). For the representation methods (see Fig. 5), the top three are *bag of n-grams*, *bag of words*, and *topic-based representation*. Representation methods aim to represent words with vectors or other numerical characterizations. The principal machine learning algorithms employed are *decision tree*, *support vector machines*, and *Bayesian networks* (see Fig. 6). IBM defined machine learning as a branch of artificial intelligence (AI) and computer science that focuses on using data and algorithms to imitate the way humans learn, gradually improving its accuracy[2]. Finally, Fig. 7 shows the most used performance metrics employed in the studies: *Accuracy*, *Recall*, and *Precision* were the top three. Performance metrics offer a quantitative evaluation of the model's fitness and depend on the task carried out.

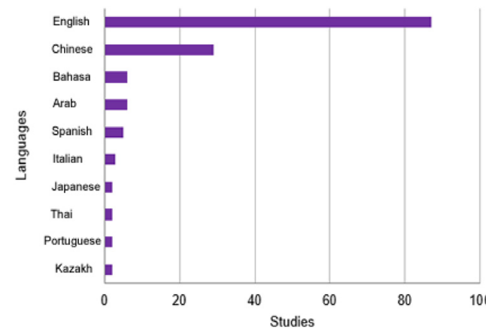

Fig. 3. Bar chart of papers and which language analyzed with NLP.

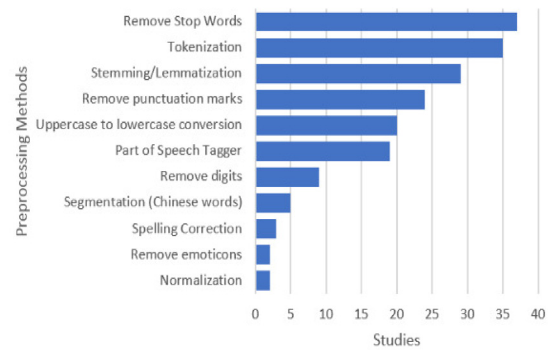

Fig. 4. Preprocessing techniques used in the revised literature.

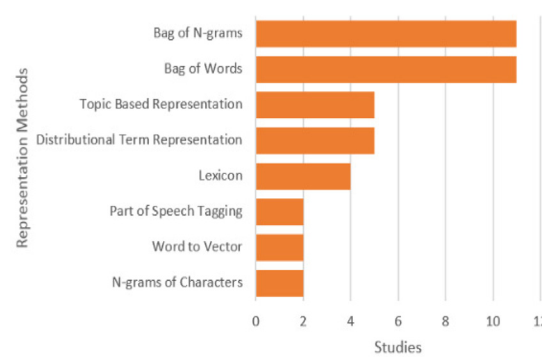

Fig. 5. Representation methods used in the revised literature.

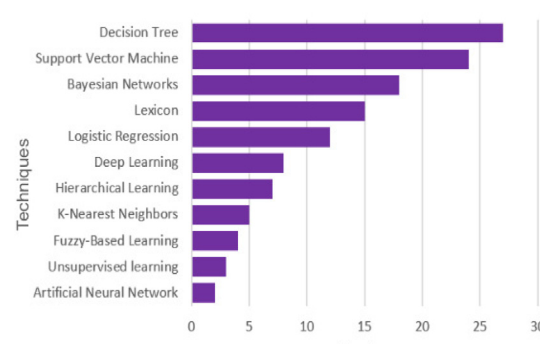

Fig. 6. Learning algorithms used in the revised literature.

[2] https://www.ibm.com/cloud/learn/machine-learning

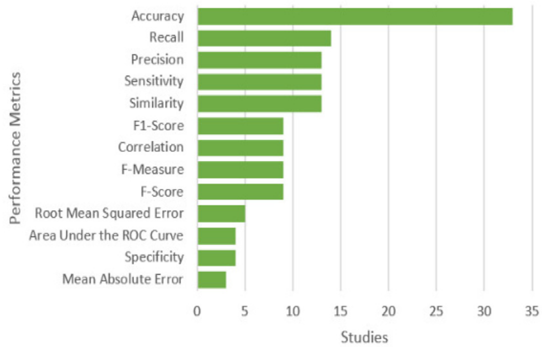

Fig. 7. Performance metrics used in the revised literature.

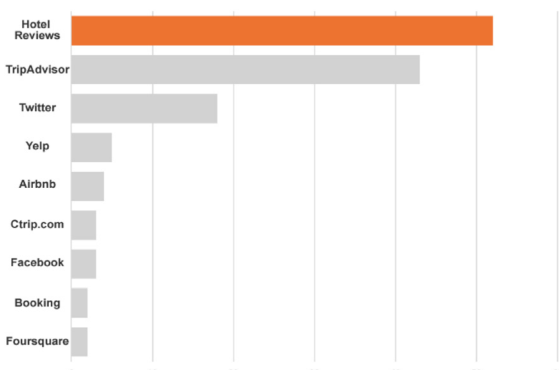

Fig. 8. Main sources of data on the analyzed studies.

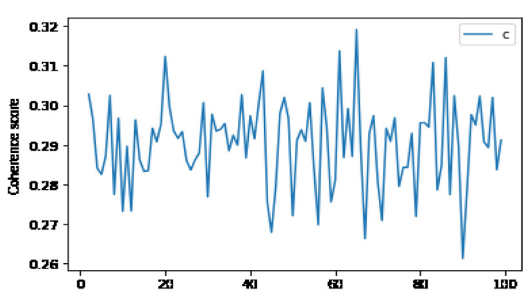

Fig. 9. Coherence score with a different number of topics.

In order to provide a better understanding, Santhappan and Chokkalingam (2020) define *accuracy* as the ratio of the total number of correct predictions concerning the total number of predictions made. On the other hand, *recall* is the proportion between true positives (considering the basic scenario of two classes) to the total of positive cases. Finally, *precision* is the ratio between the true positives and the sum of the total positives and the false positives. An analysis of the remaining performance metrics is outside this work's scope. For further information, please consult (Sokolova and Lapalme, 2009; Botchkarev, 2019).

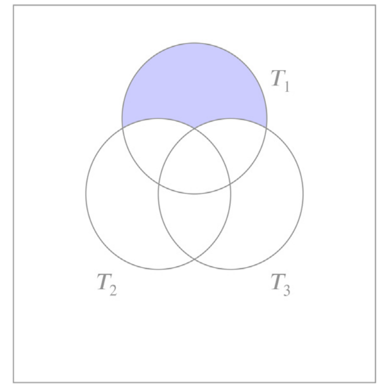

Fig. 10. $T_1 - (T_2 \cup T_3)$ which represents the unique elements of T_1.

Fig. 8 depicts the information extracted from the studies about the principal data sources. It is important to emphasize that some studies employed diverse data sources for their experiments; data sources with too few mentions are not shown. In the graphic, *Hotel Reviews* refers to information extracted from a hotel's website or surveys on the guests. As mentioned in the work of Ert and Fleischer (2016), online travel booking platforms have been one of the most important tourist marketing channels because of the development of information technologies. Although many studies have relied on surveys of specific hotels, the preeminence of online platforms has been growing during the past decade. In the study of Tsao et al. (2015), they found that about 80% of travelers read reviews of hotels before they venture into a trip, and 53% claim that they will not book a hotel that has no reviews. Considering the above, we thought that the popularity of platforms like TripAdvisor, Yelp, or Twitter, will become predominant as the technology develops.

Up to this point, our intention behind these questions was to provide some insights for newcomers to NLP research into tourism. We aim to discuss a deeper perspective of the technical aspects in the latter subsections.

2.2. Textual abstracts analysis

Generally, an article abstract is concise and describes the objectives and scope of an investigation. For this reason, it is possible to determine thematic directions within an article through its abstract. For this, we propose applying Latent Semantic Analysis (LSA), which is an algorithm for automatic detection of hidden topics (Landauer et al., 2013), overall the articles' abstracts. This analysis was performed using the Python programming language and the libraries NLTK and GENSIM. The parameters and processes are described below. The number of topics is obtained empirically by measuring the degree of coherence of each topic. The coherence score was calculated in the range $[2, 100]$ topics. In Fig. 9, it can be seen in which number of topics the most significant coherence was obtained. The global maximum is 68 topics; however, a very close value and many fewer topics are 22, a local maximum. For this reason, these 22 topics will be analyzed.

Table 1
First two topics obtained with LSA over the abstracts corpus.

Topic 1:	review	sentiment	hotel	use	analysis
Topic 2:	sentiment	review	hotel	use	analysis

Table 2
Resulting topics that maximize the unique elements of each set.

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6
data	use	help	service	analysis	destination
tourist	aspect	travel	travel	data	travel
text	features	recommend	evaluation	text	analysis
tourism	user	rate	approach	inform	system
bangkok	tweet	context	comment	opinion	help
sentiment	number	data	information	help	base
object	online	evaluation	analysis	twitter	social
extract	transport	custom	rate	method	dialogue
custom	data	consum	research	sentiment	media
visit	base	text	object	hotel	arabic
Sentiment analysis	Travel analysis	Recommendation systems	Destination evaluation	Sentiment analysis for hotels	Destination analysis

Table 1 shows the top 5 words of the first two topics. It is possible to observe that the topics are identical since only the order of the words changes. To avoid topics with huge intersections, we propose an index to calculate each topic's degree of contribution. For this, we will see each topic as a set. That is, there are 22 topics in the collection, and it is represented as follows:

$$T = \{T_1, T_2, \ldots, T_k, \ldots, T_{22}\}. \quad (1)$$

The degree of contribution of a topic k is computed as:

$$Cont(T_k) = \frac{|T_k - \cup T^{-k}|}{|T_k|}, \quad (2)$$

where $T^{-k} = T - \{T_k\}$. The idea behind of Eq. 2 is to compute the different unique words of each topic. To illustrate it, Fig. 10 shows the example with three sets. The blue area represents the unique elements for the T_1 set. If all elements in T_k are unique, then, $Cont(T_k) = 1$, if T_k does not have any unique element, then $Cont(T_k) = 0$.

When computing the $Cont(T_k)$ for all k in $[1, 22]$, ten topics with a value of 0 were obtained. These sets were eliminated since their contribution is null. By recalculating $Cont(T_k)$ for the remaining 12 topics, this time, the topics with values below 0.1 were eliminated. The process was repeated until the minimum value of $Cont(T_k)$ was 0.3.This is expressed in the algorithm 1.

Algorithm 1 Set selection algorithm proposed

Require: T
Ensure: $T\prime$
$threshold \leftarrow 0$
while $threshold \leqslant 0.3$ **do**
$del \leftarrow Empty\ set$
for $i = 1$ **to** $len(T)$ **do**
$index = Count(T_i)$
if $index \leqslant threshold$ **then**
$del.add(T_i)$
end if
end for
$T \leftarrow T - del$
$threshold \leftarrow threshold + 0.1$
end while
$T\prime \leftarrow T$
return $T\prime$

Finally, As a result of the proposed algorithm, we found six topics to maximize the value of $Cont(T_k)$ for all k. These topics can be seen in the Table 2. The last row in this table represents each topic's name, tagged by humans, derived from each topic's words.

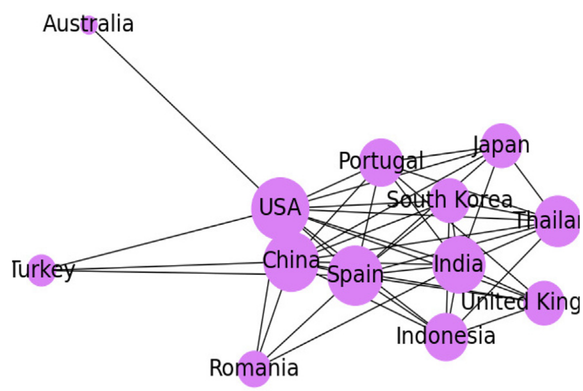

Fig. 11. Countries connected by 5 keywords.

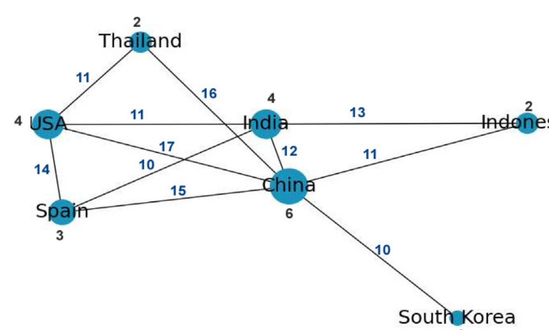

Fig. 12. Countries connected by 10 keywords. The degree is shown above or below the node and the weight is over the edge.

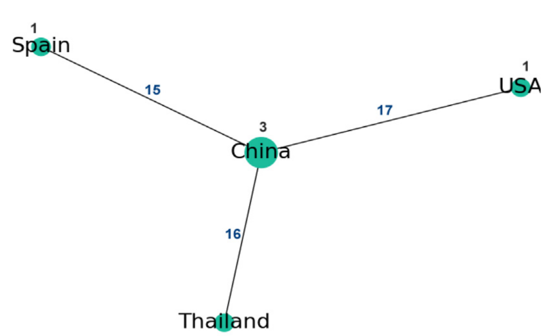

Fig. 13. Countries connected by 15 keywords. The degree is shown above or below the node and the weight is over the edge.

It can be seen that the main topics automatically extracted that are addressed in the abstracts of all the articles analyzed in this systematic review are sentiment analysis, Travel, Recommendation systems, Destination branding, Sentiment analysis for hotels, and Destination recommendation.

2.3. Keywords analysis

Keywords in the scientific literature are words selected by the authors to identify relevant information about their researches. Most of the keywords depict methods and main topics or issues. Lu et al. (2020) show high-frequency coincidence between author–selected keywords and similar research keywords. We extract the research keywords by country to determine which countries work have a relationship in their research. With the help of the Python programming language and the NetworkX library, we generated graphs to show the networks present in the studies of the different countries in the review. Each node is formed with the keywords of each study (separated by country), and the links between two nodes are established based on the number of keywords they have in common. By varying this parameter (the number of keywords), we can generate different networks to understand the underlying relations among the studies of the countries. Figs. 11–13 presents the countries' network graphs by using different connectivity criteria. Information on weights and node degrees is presented in the graphs themselves for Figs. 12 and 13. Since Fig. 11 depicts a more complex graphic representation, this information is presented as an adjacency matrix (Table 3). Fig. 11 shows a network of countries with almost five keywords in common. The countries with the highest node degrees (ND) are the United States of America with 12, China, and Spain with 11. Thus, we can show that those countries have more variability in their main topics. When we change the connectivity criterion to have almost ten keywords in common (see Fig. 12), the countries with the highest ND are China (6), the United States (4), and India (4). In Fig. 13, we set the connectivity criteria to 15. Thus, we can see that China is the central node in the USA, Spain, and Thailand network. Although Thailand is not one of the countries with more variability in the keywords, it has many similar topics to China. The tourism issue from keywords that the USA, Spain, and Thailand work in common with China are customer dissatisfaction/satisfaction, hospitality, chatbot, Kansei engineering, review helpfulness, sentiment analysis, and service development.

The algorithm/method in common with keywords are artificial neural networks, big data analysis, clustering, data mining, topic model, visual analytics, lexicon, recommendation system, and text mining to analyze data from hotel reviews, ewom, social media, Twitter, TripAdvisor, as well as google trends and web mining in general.

3. Results

One of the most relevant findings of this work is that thanks to a comprehensive analysis conducted by our NLP experts, we have classified 227 studies according to the problems discussed. We found five major categories: *sentiment analysis*, *destination branding*, *question–answering*, *NLP for assisting in tourism*, and *miscellaneous*. Regarding the category "NLP for assisting in tourism", it comprises two sub-categories: recommendation systems and NLP for assisting in travels. The "miscellaneous" category is composed of studies that solve specific problems that cannot be grouped into the categories mentioned earlier. Fig. 14 shows a graphic representation of the taxonomy, and Fig. 16 presents a chart with the volume of research efforts in each family. To understand where our taxonomy comes from, Fig. 15 presents its genealogy and inherited traits from its ancestors. Our taxonomy is based on the one presented by Kennedy-Eden and Gretzel (2012) for mobile applications in tourism and the taxonomy on NLP capabilities presented by Roukos and Soffer (2020). Table 10 provides information on the manuscripts analyzed in this work, classified by taxonomic families to assist newcomers to NLP applications in tourism. We

Fig. 14. Taxonomy of the analyzed studies.

Table 3
Adjacency Matrix of the weighted undirected graph in Fig. 11. The degree of each node is shown in the last column of the table. The names of the countries are abbreviated according to the standard *ISO-3166–1 ALPHA-3* of three characters.

	AUS	USA	CHN	ROU	PRT	TUR	ESP	JPN	KOR	IND	IDN	THA	GBR	Node Degree
AUS	-	6												1
USA	6	-	17	6	5	5	14	7	6	11	7	11	8	12
CHN		17	-	7	6	6	15	9	10	12	11	16	7	11
ROU		6	7	-			5			6				4
PRT		5	6		-		7	6		6		5	7	7
TUR		5	6			-	5							3
ESP		14	15	5	7	5	-	7	8	10	5	6	9	11
JPN		7	9		6		7	-		7		6		6
KOR		6	10				8		-	6	5	6		6
IND		11	12	6	6		10	7	6	-	13	7	8	10
IDN		7	11				5		5	13	-	9	7	7
THA		11	16		5		6	6	6		9	-		7
GBR		8	7		7		9			8	7		-	6

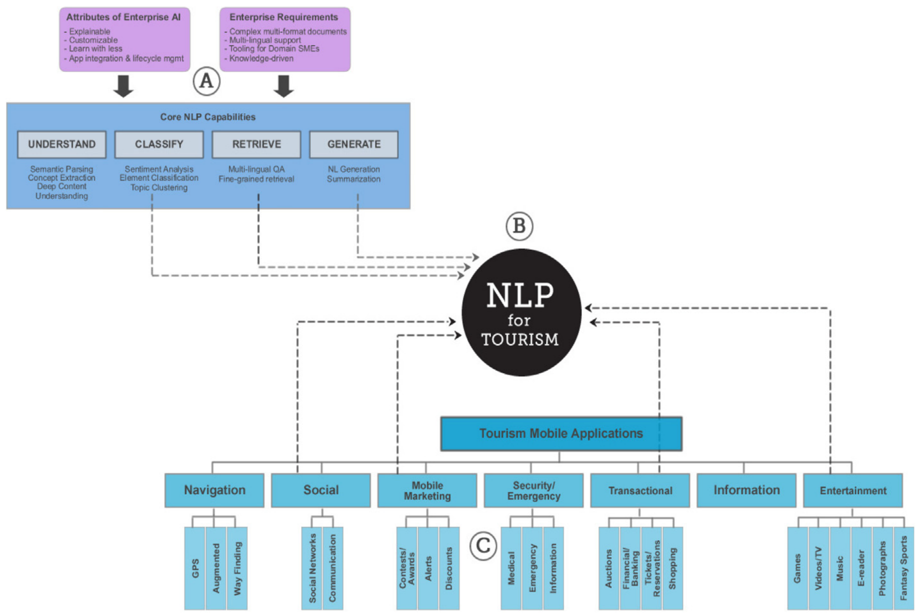

Fig. 15. Genealogy of our proposed taxonomy. In (A), a taxonomy for NLP capabilities, proposed by Roukos and Soffer (2020) is presented. In (B), the black circle represents our proposed taxonomy. In (C) is the taxonomy for tourism mobile application of Kennedy-Eden and Gretzel (2012).

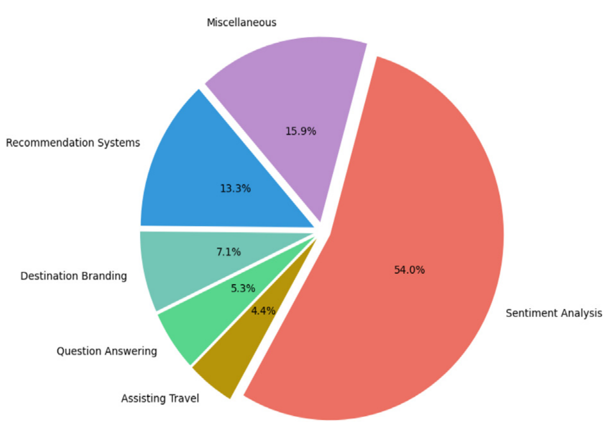

Fig. 16. Distribution of the studies according to the taxonomy.

wish to emphasize that this taxonomy, the first in our knowledge of NLP applications in tourism, could help organize future efforts in this field and give an overview of the area's landscape to expert researchers who could help expand the taxonomy. By building this categorization, we responded to the *first research question* of this study.

3.1. Issues definitions

We describe a few relevant terms below to provide complete information about the tourism issue found in this work.

3.1.1. Question–answering

This approach provides the appropriate text to answer a question. For example, Moubaiddin et al. (2015) introduced a chatbot to interact with hotel customers and help them book a room and other services; Gal-Tzur et al. (2014) presented a methodology for automatically categorizing questions related to transportation in Q&A forums (like TripAdvisor) with data in Hebrew; Gkritzali et al. (2018) analyzed TripAdvisor messages to understand the economic recession's impact on Athens' tourist perception.

3.1.2. Sentiment analysis

The aim is to classify, according to their positive/negative polarity, the opinions of customers (Anis et al., 2020). For example, Al-Smadi et al. (2019) and Rybakov and Malafeev (2018) presented aspect-based sentiment analysis classification tasks for hotels reviews in Arabic and Russian, respectively. Duan et al. (2016) conducted sentiment analysis in the hotel industry to decompose user reviews into different dimensions to measure hotel service quality; Luo and Zhai (2017) analyzed opinions on a web platform (Weibo); their goal was to understand the evolution of people's reactions in a political crisis that becomes a threat to tourism. Kim et al. (2017) presented a Big Data approach to understand the perception of the service in different categories. García-Pablos et al. (2018) proposed a non-supervised classification method for aspect-based sentiment analysis. The method is carried out simultaneously in three subtasks: domain aspect classification, aspect/opinion separation, and sentiment polarity classification. Cheng and Jin (2019) identified, through sentiment analysis, three key attributes that influence the reviews on the Airbnb platform. Guerreiro and Rita (2020), investigated an analysis of travelers' opinions on online platforms to identify triggers for direct negative recommendations.

3.1.3. Branding design or destination branding

With automated content analysis, institutions can understand how a tourist destination is perceived. For example, Uchiyama et al. (2017) analyzed the perception of tourists to recognize regional products in the Niigata prefecture (Japan); Költringer and Dickinger (2015) proposed an approach to web content extraction to obtain information about the perceived destination image from different online sources; Hu and Trivedi (2020), analyzed hotel reviews to understand the coincidences and differences in customers' expectations.

3.1.4. Recommendation systems

Using language analysis and UGC, the system offers user-friendly recommendations. Suzuki et al. (2013) have proposed a method to visualize the weighting of the service (a star rating) in different categories in a hotel recommendation system based on trends in user reviews; Lin et al. (2015) have proposed an app for hotel recommendations. Navigation behavior (gestures, keys, among others.) was used to learn user patterns and recommend hotels; Wang et al. (2019) have proposed a method to assist hotel guests. The approach has extracted Chinese characteristics and opinions online. Lee et al. (2018) investigated the usefulness of TripAdvisor reviews. Their results showed that the characteristics of reviewers are good predictors of the review's usefulness. Moro et al. (2019) studied the effects of gamification to motivate and attract travelers to contribute to online platforms. They found three characteristics that can explain the relationship between the gamification characteristics and the traveler's behavior while writing the reviews. Francesco and Roberta (2019) have analyzed 9000 online reviews to understand the differences in how travelers from different countries perceive the hotel's attributes. A better understanding of these differences makes better recommendations.

While extracting the information, we have established a classification system for the documents gathered. The system was based on studies' ability to respond to the research questions, and the ranking scale varied from five to one. We have given five points to studies that provided profuse information to answer all research questions and one point to documents that provided sufficient information to answer at least one question. Despite the subjective nature of this classification, we thought it would be a good guideline for presenting the most important contributions. We have summarized all the articles with scores of five and four (see Tables 4–9), for a total of fifty-three, close to a quarter of the literature analyzed.

4. Discussion

The main objective of this systematic review was to answer three interesting questions from the NLP in tourism. The discussion of each question is presented in this section, with their respective answers:

- In which tourism issues are the NLP techniques applied?: To respond to the first question, a taxonomy was manually constructed according to the authors' expertise. Subsequently, an automatic analysis of the articles' abstracts was conducted to discover hidden topics. From both approaches, there was a coincidence in four topics: Sentiment Analysis, Recommendation systems, Destination branding, and Travel or assisting travel. The remaining two topics discovered by the automatic approach stress the preponderance of sentiment analysis and recommendation systems in the literature, where most articles were focused. Regarding missing topics (miscellaneous and question–answering), this situation emphasizes human reasoning's capacity to detect the subtleties among topics.

Table 4
Summary of the most relevant studies for Sentiment Analysis task.

Work	Task	Data representation	Algorithms or Informatic tools	Data source	Num. records	Country
García et al. (2012)	Sentiment Analysis	Bag of Words, Bag of N-grams.	Lexicon	TripAdvisor	1994	Spain
De Núñez et al. (2018)	Sentiment Analysis	N-grams of words	Naive Bayes	Twitter	N/A	Spain
Shi and Li (2011)	Sentiment Analysis	TF-IDF	SVM	Hotel reviews	4000	China
Xiao et al. (2018)	Sentiment Analysis	N/A	Convolution Control Block	Hotel reviews	192,848	China
Zvarevashe and Olugbara (2018)	Sentiment Analysis	Distributional term representation	Naive Bayes, Sequential minimal optimization	Hotel reviews	259,000	South Africa
Sanchez-Franco et al. (2019)	Sentiment Analysis	N/A	Latent Dirichlet Allocation	Yelp	47,172	Spain
Ye et al. (2012)	Sentiment Analysis	Bag of Words, Bag of N-grams.	Lexicon	Microblogging Comments	24,850	China
Kirilenko et al. (2018)	Sentiment Analysis	Lexicon	Bayesian Networks, SVM, SentiStrength, Deeply Moving	Twitter, Hotel reviews, Surveys with qualitative data	765,039	USA
Putri and Kusumaningrum (2017)	Sentiment Analysis	Bag of Words, Bag of N-grams, LDA, LSA, LSI	Latent Dirichlet Allocation	TripAdvisor	100	Indonesia
de Souza et al. (2018)	Sentiment Analysis	N/A	Deep Learning, Hierarchical learning	TripAdvisor	69,075	Brazil
Agarwal et al. (2018)	Sentiment Analysis	Word2Vec	k–Means	Hotel Reviews	N/A	India
Chatterjee (2020)	Sentiment Analysis	Bag of Words, Bag of N-grams	Random Forest, SVM, ANN, Poisson regression, and Negative binomial models	TripAdvisor	942	India
Zhang and Yu (2017)	Sentiment Analysis	Word2Vec, Bag of clusters (k-means and ISODATA)	XGBoost	Hotel Reviews	100,000	China
Chang et al. (2019)	Sentiment Analysis	Bag of Words, Bag of N-grams	SVM	TripAdvisor	6,090,584	China
Kurniawan et al. (2018)	Sentiment Analysis	Term weighting	Bayesian Network	Traveloka	1,720	Indonesia
Elnagar et al. (2018)	Sentiment Analysis	Bag of Words, Bag of N-grams	Logistic Regression, SVM, Random Forest, Perceptron, AdaBoost	Booking.com	373,772	United Arab Emirates
Ren and Hong (2017)	Sentiment Analysis	Bag of Words, Bag of N-grams	Latent Dirichlet Allocation	Ctrip.com	2000	South Korea
Barbosa et al. (2015)	Sentiment Analysis	POS tagging with WSentUAH	Unsupervised Naive Bayes, Bayesian Networks, Lexicon, OpinionFinder, and CoreNLP	TripAdvisor	1,335,781	

Table 5
Summary of the most relevant studies for Assisting Travel task.

Work	Task	Data representation	Algorithms or Informatic tools	Data source	Num. records	Country
Gal-Tzur et al. (2014)	Assisting Travel	Bag of Words, Bag of N-grams.	SVM	Twitter	1500	Israel
Ahn et al. (2017)	Assisting Travel	TF-IDF	Semantic network	Booking.com	125,076	South Korea
Ali et al. (2017)	Assisting Travel	N-grams of words	Fuzzy ontology, fuzzy rules	Twitter, FaceBook	3,255	South Korea
Argal et al. (2018)	Assisting Travel	N/A	Deep Learning, hierarchical learning	Hotel reviews, Airline rec.	230,000	U.S.A.
Ali et al. (2019)	Assisting Travel	N-grams, Word2Vec, SentiWordNet	Deep Learning, fuzzy rules, fuzzy Ontology, SVM	Twitter, FaceBook, Hotel reviews, TripAdvisor, Etc.	500,000	South Korea
Serna et al. (2017)	Assisting Travel	N/A	WordNet aligned with SUMO ontology, Freeling	Minube (online travel social network)	43,251	Spain
Poernomo (2019)	Assisting Travel	Unigram feature	k-NN, Bayesian Networks, SVM	Facebook	N/A	Indonesia
Dardas et al. (2020)	Assisting Travel	N/A	R, k-means	Questionnaire	25 participants	Canada

Table 6
Summary of the most relevant studies for Recommendation Systems task.

Work	Task	Data representation	Algorithms or Informatic tools	Data source	Num. records	Country
Fukumoto et al. (2012)	Recommendation Systems	Bag of words, n-grams	SVM	Hotel reviews	11,800	Japan
Fukumoto et al. (2014)	Recommendation Systems	Word triplets	SVM	Hotel reviews	348,564	Japan
Hu and Chen (2016)	Recommendation Systems	Bag of words	Linear regression, decision tree (M5P), SVR	TripAdvisor	573,527	China
Bhardwaj et al. (2017)	Recommendation Systems	N/A	Naive Bayes and Lexicon combination	Twitter	500	India
Siering et al. (2018)	Recommendation Systems	Bag of words, bag of n-grams	Logistic regression, Naive Bayes, SVM, neural network	airlinequality.com	3,000	Germany
Chen et al. (2019)	Recommendation Systems	bag of concepts	text mining, link analysis	Hotel reviews	2,552	China
Sun et al. (2018)	Recommendation Systems	Distributional term representation	k-NN	Hotel reviews	69,774	China
Al-Smadi et al. (2018)	Recommendation Systems	Lexical N-grams, Morphological features, Syntactic features, Semantic features	Deep learning, SVM	Hotel Reviews	24,028	Jordan
Meehan et al. (2016)	Recommendation Systems	N/A	Alchemy API	Twitter	5,370	United Kingdom
Capdevila et al. (2016)	Recommendation Systems	LDA, LSA, LSI, SLDA	Logistic Regression, Bayesian Networks,	Foursquare	309,640	Spain

Table 7
Summary of the most relevant studies for Destination branding task.

Work	Task	Data representation	Algorithms or Informatic tools	Data source	Num. records	Country
Költringer and Dickinger (2015)	Destination branding	N-grams of words	N/A	Websites of DMOs, UGC of several blogs, etc.	5,719	Austria
Shin et al. (2017)	Destination branding	Bag of words, bag of n-grams	The R statistical language	TripAdvisor	7,563	South Korea
Bigné et al. (2019)	Destination branding	N/A	Multi-Layer Perceptron	Twitter	N/A	Spain
De Lucia et al. (2020)	Destination branding	N/A	Ordered logistic regression model	Survey	223	Italy
Hu and Trivedi (2020)	Destination branding	part-of-speech (PoS)	KH-Coder, Idiogrid	Hotel reviews, TripAdvisor	111,986	China
Liu et al. (2020)	Destination branding	PoS, frequency and ratio word	NVivo 11	Ctrip, TripAdvisor	51,191	China

Table 8
Summary of the most relevant studies for Question–Answering task.

Work	Task	Data representation	Algorithms or Informatic tools	Data source	Num. records	Country
Moubaiddin et al. (2015)	Question–Answering	Semantic trees	Government and Binding-Based Parser	N/A	500	Jordan
Gal-Tzur et al. (2018)	Question–Answering	N-grams of characters	Decision tree, Naive Bayes, Random forest, SVM	TripAdvisor	760	Israel
Gkritzali et al. (2018)	Question–Answering	N/A	N/A	TripAdvisor	109,460	United Kingdom
Sano et al. (2018)	Question–Answering	N/A	AGNES algorithm	Data of hotels	39	Indonesia
Tussyadiah and Park (2018)	Question–Answering	Tokenization	Clustering, Jaccard distance	Airbnb	31,119	United Kingdom
Varga et al. (2014)	Question–Answering	N-ary ontology design pattern	Conditional Random Field	Museum Guides, Wikipedia, DBpedia, Geonames, Freebase,	3,200	Romania

Table 9
Summary of the most relevant studies (Miscellaneous).

Work	Task	Data representation	Algorithms or Informatic tools	Data source	Num. records	Country
Tsou (2010)	Miscellaneous	N/A	N/A	Google search queries	N/A	China
Lee and Wang (2012)	Miscellaneous	Bag-of-keypoints, Fuzzy ARTMAP, Adaptive Affinity Propagation	SVM Self Organizing Maps, Fuzzy ARTMAP, and Neural Networks	Tourism Bureau Kaohsiung website, Flickr, and blogs	3,600	China
Menchavez and Espinosa (2015)	Miscellaneous	Bag of words, bag of n-grams, n-grams of words	Logistic regression, Naive Bayes, SVM	Twitter	392	Philippines
Kim and Lee (2019)	Miscellaneous	Distributional term representation	Clustering, CONCOR analysis	Medical newspapers	4,735	South Korea
Wang et al. (2020)	Miscellaneous	PoS and Word Frequency	ROST News Analysis Tool	Newspapers	15,894	China

Table 10
Considered studies in this systematic review classified by the authors according to the taxonomy.

Category	Studies
Sentiment Analysis	Zvarevashe and Olugbara (2018), García et al. (2012), De Núñez et al. (2018), Ye et al. (2012), Shi and Li (2011), Anis et al. (2020), Akhtar et al. (2017), Rybakov and Malafeev (2018), Agarwal et al. (2018), Kirilenko et al. (2018), Chatterjee (2020), Al-Smadi et al. (2019), Barbosa et al. (2015), Duan et al. (2016), Kurniawan et al. (2018), Elnagar et al. (2018), Zhang and Yu (2017), Guerreiro and Rita (2020), Ren and Hong (2017), Putri and Kusumaningrum (2017), Prameswari et al. (2017b), Irawan et al. (2019), Chang et al. (2019), Campos et al. (2019), Sanchez-Franco et al. (2019), Xiao et al. (2018), García-Pablos et al. (2018), Cheng and Jin (2019), Kim et al. (2017), Zhang (2019), Luo and Zhai (2017), Agüero-Torales et al. (2019), Moon and Kamakura (2017), Tolkach (2018), Thelwall (2019), Alaei et al. (2017), Priyantina and Sarno (2019), Lai and Raheem (2020), Devika et al. (2016), Nisar and Prabhakar (2018), Nakayama and Wan (2018), AL-Smadi et al. (2016), Jianlei et al. (2019), Linfeng et al. (2017), Park et al. (2016), Micu et al. (2017), Bueno et al. (2019), He et al. (2017), Godnov and Redek (2016), Panigrahi and Asha (2018), Becken et al. (2019), Zapata et al. (2019), Mishra et al. (2016), Li et al. (2018), Yadav and Roychoudhury (2019), Xu (2020), Kuhzady and Ghasemi (2019), Thelwall (2018), Wang and Hao (2018), Antonio et al. (2018), Huang et al. (2017), Sutabri et al. (2018), Pham et al. (2018), Liu et al. (2019b), Tran et al. (2019), Duan et al. (2013), Tao et al. (2019), Becken et al. (2017), Gonzalez-Rodriguez et al. (2014), Sodanil (2016), González-Rodríguez et al. (2016), Fu et al. (2018), Afzaal et al. (2018), Kim and Im (2018), Colhon et al. (2014), Geetha et al. (2017), Panigrahi et al. (2018), Lee et al. (2017), Yates et al. (2013), Gu et al. (2018), Martins et al. (2017), Ma et al. (2015), Sutabri et al. (2018), Liu et al. (2019a), Kuhamanee et al. (2017), Khotimah and Sarno (2019), Markopoulos et al. (2015), Yu et al. (2019), Song and Wang (2017), Bădică et al. (2014), Shuai et al. (2018), Farisi et al. (2019), Yergesh et al. (2017), Windasari and Eridani (2017), Suardika (2016), Aliandu (2015), Gao et al. (2019), Ma et al. (2018), Yang and Chao (2018), Zhang et al. (2016), Mao et al. (2018), Salur et al. (2019), Mena-Maldonado et al. (2016), Wu and Zhang (2016), Xu and Li (2016), Gao et al. (2015), Hu et al. (2017), Xu et al. (2018), Phillips et al. (2019), Zhu et al. (2017), Gupta and Gupta (2019), Hermanto et al. (Dec 2018), Ramanathan and Meyyappan (2019), Park et al. (2018), Berezina et al. (2015), Mathayomchan and Sripanidkulchai (2019), Nguyen-Thanh and Tran (2019), Schwartz et al. (2019), Hu et al. (2019), Nawangsari et al. (2019), Cruz et al. (2013), Buzova et al. (2018), Guerrero-Rodriguez et al. (2021), Álvarez-Carmona et al. (2022b), Carmona-Sánchez et al. (2021), Álvarez-Carmona et al. (2021), Romero-Cantón and Aranda (2021), and Álvarez-Carmona et al. (2022)
Miscellaneous	Lee and Wang (2012), Menchavez and Espinosa (2015), Tsou (2010), Kim and Lee (2019), Wang et al. (2020), Colladon et al. (2019), Young and Gavade (2018), Serna and Gasparovic (2018), Punel et al. (2018), Chanwisitkul et al. (2018), Sanz-Blas et al. (2016), Rahmani et al. (2018), Sekar et al. (2017), Moreno-Ortiz et al. (2019), Fukui et al. (2019), Barbado et al. (2019), Cheng (2017), Lee et al. (2011), Tao et al. (2019), Zeng et al. (2018), Tjahyanto and Sisephaputra (2017), Panawong and Sittisaman (2019), Wong and Qi (2017), Xu et al. (2017), Qi et al. (2017), Linares et al. (2015), **?**), Zapata et al. (2017), Zhang et al. (2019), Tsujii et al. (2013), Yan et al. (2018), Cheng and Foley (2018), Ruhanen et al. (2019), Moro et al. (2019), Biehl et al. (2019), Mazanec (2017), Liu et al. (2018), Pacheco et al. (2021), Álvarez-Carmona et al. (2022a)
Recommendation Systems	Meehan et al., 2016, Chen et al. (2019), Lee et al. (2018), Moro et al. (2019), Fukumoto et al. (2012), Francesco and Roberta (2019), Al-Smadi et al. (2018), Siering et al. (2018), Fukumoto et al. (2014), Capdevila et al. (2016), Suzuki et al. (2013), Bhardwaj et al. (2017), Hu et al. (2017), Lin et al. (2015), Hu and Chen (2016), Wang et al. (2019), Sun et al. (2018), Arteaga et al. (2019), Arote and Paikrao (2018), Yang and Lin (2016), An et al. (2019), Ravindran and Rejikumar (2017), Tsujii et al. (2015a), Bucur (2015), Sezgen et al. (2019), Tsujii et al. (2015b), Sun et al. (2018), Tazl et al. (YYYY), Liang et al. (2019), Bhargav et al. (2019), Arce-Cardenas et al. (2021), Álvarez-Carmona et al. (2021), and Álvarez-Carmona et al. (2022)
Destination Branding	Költringer and Dickinger (2015), Shin et al. (2017), Uchiyama et al. (2017), De Lucia et al. (2020), Bigné et al. (2019), Hu and Trivedi (2020), Liu et al. (2020), Scorrano et al. (2019), Moro and Rita (2018), Micera and Crispino (2017), Ying (2018), Loaiza et al. (2019), Zapata et al. (2018), Lee and Bradlow (2011), Lee et al. (2019), Athuraliya and Farook (2018), and Álvarez-Carmona et al. (2022)
Question Answering	Gal-Tzur et al. (2018), Moubaiddin et al. (2015), Gkritzali et al. (2018), Varga et al. (2014), Sano et al. (2018), Tussyadiah and Park (2018), Neidhardt et al. (2017), Becheru and Bădică (2016), Buzova et al. (2016), Çalı and Balaman (2019), Srivastava and Prabhakar (2019), and Siow et al. (2015)
Assisting Travel	Masrury et al. (2019), Ali et al. (2019), Ali et al. (2017), Poernomo (2019), Argal et al. (2018), Serna et al. (2017), Gal-Tzur et al. (2014), Ahn et al. (2017), Dardas et al. (2020), and Maghrebi et al. (2015)

- Which state-of-the-art NLP algorithms/methods are applied for tourism issues?
  Different artificial intelligence algorithms are employed in the literature concerning cutting-edge techniques. The following stages are typically followed for NLP machine learning processes: preprocessing, representation, and classification. The most used data preprocessing technique is removing stop words, tokenization, and a steaming procedure. Bags of n-grams and words became the most popular schemes to represent text. Finally, decision trees and SVM are the most used algorithms for data classification since these techniques were applied in 40.8 % of the reviewed works. In this regard, it is important to highlight China's research efforts, the most numerous, and performing the most exhaustive exploration of AI methods. Although China's economic income is not the best, their tourism potential, combined with NLP techniques investigation, makes foreseeable their elevated possibilities for growth in the number of visitors and income.
  Something essential to note about machine learning algorithms is the difficulties that occur during the process. According to our experience in tourism research, difficulties arise in two main fields: methodological approaches and technical problems. Regarding data research projects, there are two principal methodological approaches: top-down and bottom-up (Wirsch, 2014). In a top-down methodology, we start with a research question and using domain-specific knowledge, the hypothesis is proposed and tested with the gathered data. This approach is traditionally preferred in tourism research.
  On the other hand, a button-up methodology performs an exploratory analysis of all the available data. These explorations can find interesting patterns that can lead to meaningful discoveries that can be or cannot be under the objective of the research. These methodologies can be seen as complements without one exceeding the other. However, the difference between approaches and preferences entails difficulties for multidisciplinary teams. The technical problems relate to the quality or nature of the data. The quality refers to noisy data, missing values, class unbalance, and other defects corrected during the pre-processing stages (Hardeniya et al., 2016). By the nature of data, we refer to privacy-related concerns on the examined data, such as sensitive information of tourists. Considering the above, different encrypting strategies must be contemplated (Geyer et al., 2017).
  The languages that were analyzed through NLP were examined. The status of English as a lingua franca among travelers or the wider variety of frameworks for its preprocessing may be why many studies focused on this language. On the other hand, the Chinese language took second place in the comparison. This could be attributed to the significant quantity of Chinese tourists and the importance of Tourism-NLP research in China.
  Another interesting aspect is the lower participation of deep learning techniques that have shown outstanding results in other NLP areas such as transformers.
- What information sources (data) are required to apply NLP techniques for tourism issues?
  The third research question found that hotel reviews and data from Tripadvisor were the preferred source. We infer this is due to social media's gradual acceptance and other travel platforms during the current decade. There are higher possibilities of future growth of studies using these platforms since they have become trendy in the last decade, and their use is positive. However, much of the data used in the research is taken from online platforms. The specific query (keywords search, dates, cleaning data process, among others) or date base used is not provided.

Additionally, connections among countries were investigated to understand similarities in their issues and the algorithms employed. Varying the connectivity criterion in the range of $[5, 15]$ words, four countries stood out: China, the USA, Spain, and Thailand. Such countries addressed tourism issues like customer satisfaction/dissatisfaction, hospitality, chatbots, sentiment analyses, and service development. The principal techniques employed were artificial neural networks, big data analysis, data mining, and topic modeling. The statistical and semantic network analyses agreed on the three countries' relevance and highlighted the similarities Thailand shares with the other three.

Something important to note is that many authors do not share the entire process of learning and classifying in their research work. 37.5% did not share the data preprocessing, the text representation by 36.3%, the classification algorithm by 51.19%, and the database's description 50%. Most studies find it difficult to replicate the result for possible comparison. Despite all the facts presented in this study, we found that text analysis was performed superficially in many studies. Moreover, the increasing dependency on electronic means in tourism makes it relevant to invest in NLP research, which is not regarded as necessary by many countries with high potential for the tourism business but with lower academic production in this scope.

5. Opportunity areas and challenges

In previous sections, we have shown an enormous gap between developing NLP research and its practical use to address tourism issues. Research in NLP is evolving rapidly, and our results show that the practical applications in the tourism investigation still rely on antiquated techniques such as the ones displayed above. Bearing this in mind, our study mostly shows the significant delay we usually observe between practitioners and specialists. In this section, we analyze the recent advances in NLP and the challenges of the tourism scope. We believe this analysis could bring valuable information and insights to specialists looking for new directions and opportunities.

5.1. Challenges

As shown in previous sections, adopting NLP techniques and methods has been fruitful within the tourism sector. However, new developments and paradigms of NLP may provide further improvements in this industry, opening up new challenges that should be addressed in future research. Some of the main challenges related to these technological advances are summarized below:

5.1.1. Deep learning architectures

In recent years, one of the most successful tools within NLP has been architectures based on deep learning (Tay et al., 2020). Architectures such as Bert (Boukkouri et al., 2020) or GPT-3 (Dale, 2021) have represented a significant advance in various tasks such as textual classification to address sentiment analysis, recommendation systems, or transport recommendation, or to generate text and develop chatbots or an advertising generator automatically. However, in the last decade, where the NLP has been employed in tourism research, various architectures that have shown fascinating results in various tasks have not been used in this field. It is clear that sooner or later, the most complex deep learning architectures will gain ground in tourist tasks. However, some problems will have to be solved, such as:

1. The amount of the data. It is well known that these architectures work better the greater the amount of data, but for them to work reasonably well, the data also needs to be sufficient. For this reason, it is essential to start generating databases large enough for good use in the area.
2. Many languages are involved in tourism. For a single place, it is reasonable to consider that it would be visited by people from many parts of the world who speak different languages, so it is expected that many reviews can be written in more than one language. In this way, building architectures for multilanguage scenarios may be necessary.
3. Also, many people speak the same language but not from the same place. There is significant variability within languages. For example, the Spanish spoken in Spain is not the same as that of Mexico or the English of the United States as that of Australia. Therefore, a solution may be building architectures that know how to identify the differences in the variability of languages.

5.2. Directions for future research works

5.2.1. Uncertainty

The Merriam-Webster Dictionary defines uncertainty as the quality or status of being uncertain or something uncertain. As Celikyilmaz and Türksen (2009) have said, uncertainty exists in nearly all real-world problems and is inseparable from measurement. Uncertainty arises in cognitive problems because of the ambiguity inherent in natural languages, and it may emerge in social relationships from the shared meanings obtained by people. In almost all the studies analyzed, opinions written by ordinary people are the basis of NLP approaches. Given the spontaneous nature of the impressions shared online by enthusiasts without training in writing (mostly), it is highly likely that vagueness in language arises. Although various sophisticated techniques have been used in tourism research, the studies examined in this paper have not proposed alternatives to address uncertainty. In order to get more accurate results or more complete models, it is essential to address this challenge.

5.2.2. Explanatory power

Machine learning has proven to be a powerful tool in different areas, providing high precision and reliability models. Although it is a desirable scenario for computer scientists, a learning model is just a black box that receives data and produces accurate outputs without understandable reasons to trust these predictions. According to Dolce et al. (2020), statistical modeling can be classified into two different approaches. One is explanation-oriented, known as "data modeling culture". The other is a prediction-oriented approach or "algorithmic modeling culture". In the first approach, data is assumed to be drawn from a given stochastic model, and researchers are interested in testing the hypothesized "true" relationship between two or more variables and the gears governing their interconnection. The major objective of this approach is to reproduce the parameters of the model using statistical inference to improve the explanatory power of models. For the second approach, the data production process is unknown, and the goal is to find an algorithm that can recognize hidden patterns in the data, which then provides the best prediction for output values through input values of new observations. Many approaches analyzed in this work belong to the second category, and for tourism researchers, it is also desirable to understand how and why the process is carried out. There are various techniques to achieve a higher power of explanation, but they are beyond the scope of this work. However, some popular approaches include interpretable AI, explainable ML, causality, safe AI, computational social science, and automatic scientific discovery (Gilpin et al., 2018).

5.3. Stream-learning on Big Data

Nowadays, a famous slogan says, "data is the new oil, and to make a profit, you need to analyze it." Internet of things, new storage technologies, and social networks have made it possible to generate tremendous amounts of data, or Big Data as it is known. This paradigm refers to data objects that are too large to be processed by conventional hardware and software, and a broad definition describes the concept in terms of three characteristics of information: Volume, Velocity, and Variety. Volume is for the vast quantities of data for analysis. Velocity represents the frequency at which data is generated, captured, and shared. Variety shows that information can be found in different formats such as text, images, audio, and video (Díaz-Pacheco and Reyes-Garcia, 2021). While other vs. have been added to the definition (Validity and Volatility), we are interested in the first three in this section. However, many manuscripts in this review employed Big Data (Alaei et al., 2017; Bhardwaj et al., 2017; Çalı and Balaman, 2019) in online comments or Internet posts in tourism forums, the vs. for Velocity and Variety were absent. We believe that a significant challenge for tourism research is using methods to get real-time indicators of the factors of interest for intelligent decision-making. During various "events", pleasant like the Olympics or negative ones like meteorological crises, data is massively generated and can be collected for DMOs and governments to design strategies.

5.4. Techniques/approaches

5.4.1. Incorporation of interval type-2 fuzzy sets to address the uncertainty

The theory of fuzzy sets covers many methods and techniques to capture human knowledge and address uncertainty in a wide range of problems. The fuzzy logic was introduced by Zadeh in 1965 and uses crisp and precise type-1 fuzzy sets (T1FS) to model human behavior under certain conditions (Zadeh, 1975a; Zadeh, 1975b). A type-2 fuzzy (T2FS) set is a set that also has uncertainty about the membership function (Castillo, 2012). The approach presented by Bi et al. (2019) proposes to use interval type-2 fuzzy numbers. The authors employed the sentiment analysis task's accuracy to decide the outcome's grade of fuzziness. If the prediction accuracy was 100%, they presented the conclusions as a triangular fuzzy number. With limited accuracy, they converted the outcome into a T2FS. The method can be employed in recommendation systems to improve the ranking of the alternatives. T2FS has also been employed to get ontology-based systems. In the work of Ali et al. (2015), they employed opinion mining in combination with T2FS to produce a new type of ontology (T2FO). According to the study, hotel reviews have many uncertainties, and most featured opinions are based on complex linguistic wording (small, big, very good, and very bad). To provide better solutions, available ontology-based systems cannot extract blurred information from reviews. T2FO was able to capture user, hotel, and provider descriptions and combine them in a global ontology employed to retrieve the exact hotel information and high sentiment score according to the users' preferences. Type 2 fuzzy sets are a great alternative to face this challenge in an area where uncertainty is the norm, not the exception. A better understanding of vagueness can improve the decision-making process for entities in the field.

5.5. Approaches to perform stream-learning on Big Data

In our experience, at least three machine learning sub-areas will help extract information from data streams in the tourism context: stream mining, Active learning, and Multimodal sentiment analysis. According to Kranjc et al. (2015), stream mining is an online learning paradigm aimed at integrating information from the evolving data stream without having to re-learn the model. On the other hand, active learning addresses data mining scenarios where a learning algorithm can periodically select new examples labeled by a human annotator and add them to a training data set to improve the model performance on new data. Multimodal learning is a set of techniques that analyzes the emotions, attitudes, and opinion of audiovisual data (Yadav et al., 2015). Many short videos and regular videos about different aspects that travelers want to share during their holidays make this type of material an essential source of data that tourist over writing sometimes prefers because of its practicality and its spontaneous nature. An important branch of multimodal learning is known as Visual Sentiment Analysis. This approach relies on deep learning techniques to determine feelings expressed in visual content, such as social media. In Chen et al. (2020), visual sentiment analysis and active learning were merged to achieve a robust classification performance. In reference Smailović et al. (2014), a methodology for stream-based active learning was proposed to analyze the relationship between the feelings expressed in the Twitter feeds of different companies and their changes in stock prices. This set of techniques could significantly benefit decision-making and is an exciting area to explore in tourism research.

6. Conclusions and considerations for future research

This work presented a systematic review of 227 studies that used NLP techniques in tourism research. The PRISMA methodology has been used to organize the process and allow the reproducibility of our results. The following research questions guided this study: (a) In which tourism issues are the NLP techniques applied? (b) Which are the state-of-the-art algorithms/methods applied for tourism issues? (c) What information sources (data) are required to apply NLP techniques for tourism issues?.

A taxonomy was built manually according to the authors' expertise to answer the first question. This taxonomy was confirmed by utilizing Topic Modeling Techniques. Of these two approaches, there was a coincidence in four topics: Sentiment Analysis, Recommendation systems, Destination branding, and assisting travel. We believe this taxonomy can be a valuable tool for future reviews, helping organize future research in the field.

For the second question, we found that tourism investigation still relies on antiquated techniques such as Lexicon, traditional neural networks, and Bayesian classifiers, among others. The major languages analyzed with NLP techniques were English and Chinese. It is essential to note that deep learning techniques, despite its proved qualities and soundness for NLP tasks, were scarcely employed in the literature (about a 6%) in comparison to the top 5 techniques (Decision tree, Suppor Vectors Machine, Bayesian Networks, Lexicon, and Logistic Regression) that accounted by around of 75% of the techniques identified. Although techniques such as SVM and the other shallow approaches are excellent, the state-of-the-art is currently dominated by robust deep learning architectures due to their outstanding results.

In the third question, we noted that hotel reviews were the most important data source for the studies, followed closely by web platforms like TripAdvisor. It is an essential indicator of the inexorable transformation of the field because of our growing digital involvement.

The authors conducted a bibliometric analysis of the keywords and other metadata with a tool designed for this purpose. Connections between countries were investigated to understand similarities in their problems and techniques. Four countries were highlighted by changing the connectivity criterion in the range of [5,15] keywords: China, USA, Spain, and Thailand. These countries have addressed tourism issues such as client satisfaction/unsatisfaction, hospitality, chatbots, sentiment analysis, and service development. The main techniques used were artificial neural networks, Big Data, data mining, and topic modeling.

As all these data presented only the current state of research, we wanted to explore the possible progress of the field by analyzing open challenges, possible techniques to address them, and new data sources. As established in the answer to the second research question, deep learning architectures were scarcely employed, and today we can take advantage of many advanced techniques in this field to manage more data, analyze more languages, and surgically differentiate variations in the same language (e.g., Spanish from Spain or Mexico).

Many guidelines have been presented for future research in addressing the uncertainty inherent to language, exploring different approaches to increase the explanatory power of the machine learning techniques in tourism research, and using Big Data techniques to analyze data flow to obtain real-time indicators for decision-making. In the same spirit, one of the approaches that can be used in future work is interval type-2 fuzzy sets to manage the uncertainty and the active learning and multimodal sentiment analysis paradigms to analyze data streams. In answer to the third research question, major information sources were plain text from surveys or web platforms. However, many developments allow the exploitation of multimedia data on the internet in online videos, podcasts, and audio. These new formats will expand the data available for analysis and get richer information.

It is important to note that many authors do not share the entire learning and classification process in their research. 37.5% did not share the data pre-processing, the text representation by 36.3%, the classification algorithm by 51.19%, and the database's description 50%. This makes it difficult to replicate the results of these studies. Despite this situation, we believe new paths will be opened for research in tourism using computer techniques.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Afzaal, M., Usman, M., Fong, A., 2018. Predictive aspect-based sentiment classification of online tourist reviews. J. Inf. Sci. 45 (3), 341–363. https://doi.org/10.1177/0165551518789872.

Agarwal, V., Aher, P., Sawant, V., 2018. Automated aspect extraction and aspect oriented sentiment analysis on hotel review datasets. In: 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), IEEE, pp. 1–4.

Agüero-Torales, M., Cobo, M., Herrera-Viedma, E., López-Herrera, A., 2019. A cloud-based tool for sentiment analysis in reviews about restaurants on TripAdvisor. Proc. Comput. Sci. 162, 392–399. https://doi.org/10.1016/j.procs.2019.12.002.

Ahn, D., Park, H., Yoo, B., 2017. Which group do you want to travel with? a study of rating differences among groups in online travel reviews. Electron. Commer. Res. Appl. 25, 105–114.

Akhtar, N., Zubair, N., Kumar, A., Ahmad, T., 2017. Aspect based sentiment oriented summarization of hotel reviews. Proc. Comput. Sci. 115, 563–571.

Akopyants, A.M., Kiseleva, A.V., Getmanskaya, M.Y., Vartanova, K.Y., Kicheva, I.V., 2017. Assimilation of the english language vocabulary in the tourism sphere. In: Perspectives on the use of New Information and Communication Technology (ICT) in the Modern Economy. Springer, pp. 839–845.

Al-Smadi, M., Al-Ayyoub, M., Jararweh, Y., Qawasmeh, O., 2019. Enhancing aspect-based sentiment analysis of arabic hotels' reviews using morphological, syntactic and semantic features. Inf. Process. Manage. 56 (2), 308–319.

Al-Smadi, M., Qawasmeh, O., Al-Ayyoub, M., Jararweh, Y., Gupta, B., 2018. Deep recurrent neural network vs. support vector machine for aspect-based sentiment analysis of arabic hotels' reviews. J. Comput. Sci. 27, 386–393.

AL-Smadi, M., Qwasmeh, O., Talafha, B., Al-Ayyoub, M., Jararweh, Y., Benkhelifa, E., Dec. 2016. An enhanced framework for aspect-based sentiment analysis of hotels' reviews: Arabic reviews case study. In: 2016 11th International Conference for Internet Technology and Secured Transactions (ICITST). IEEE. doi: 10.1109/icitst.2016.7856675

Alaei, A.R., Becken, S., Stantic, B., 2017. Sentiment analysis in tourism: Capitalizing on big data. J. Travel Res. 58 (2), 175–191. https://doi.org/10.1177/0047287517747753.

Ali, F., El-Sappagh, S., Kwak, D., 2019. Fuzzy ontology and lstm-based text mining: A transportation network monitoring system for assisting travel. Sensors 19 (2), 234.

Ali, F., Kim, E.K., Kim, Y.-G., 2015. Type-2 fuzzy ontology-based opinion mining and information extraction: A proposal to automate the hotel reservation system. Appl. Intell. 42 (3), 481–500.

Ali, F., Kwak, D., Khan, P., Islam, S.R., Kim, K.H., Kwak, K.S., 2017. Fuzzy ontology-based sentiment analysis of transportation and city feature reviews for safe traveling. Transp. Res. Part C 77, 33–48.

Aliandu, P., 2015. Sentiment analysis to determine accommodation, shopping and culinary location on foursquare in kupang city. Proc. Comput. Sci. 72, 300–305. https://doi.org/10.1016/j.procs.2015.12.144.

Álvarez-Carmona, M. Á., Aranda, R., Arce-Cardenas, S., Fajardo-Delgado, D., Guerrero-Rodríguez, R., López-Monroy, A.P., Martínez-Miranda, J., Pérez-Espinosa, H., Rodríguez-González, A.Y., 2021. Overview of rest-mex at iberlef 2021: Recommendation system for text mexican tourism.

Álvarez-Carmona, M. Á., Aranda, R., Diaz-Pacheco, Á., de Jesús Ceballos-Mejía, J., et al., 2022a. Generador automático de resúmenes científicos en investigación turística.

Álvarez-Carmona, M.Á., Aranda, R., Guerrero-Rodríguez, R., Rodríguez-González, A. Y., López-Monroy, A.P., 2022b. A combination of sentiment analysis systems for the study of online travel reviews: Many heads are better than one. Computación y Sistemas 26 (2).

Alvarez-Carmona, M. Á., Aranda, R., et al., 2022. Determinación automática del color del semáforo mexicano del covid-19 a partir de las noticias.

Álvarez-Carmona, M.Á., Díaz-Pacheco, Á., Aranda, R., Rodríguez-González, A.Y., Fajardo-Delgado, D., Guerrero-Rodríguez, R., Bustio-Martínez, L., 2022. Overview of rest-mex at iberlef 2022: Recommendation system, sentiment analysis and covid semaphore prediction for mexican tourist texts. In: Procesamiento del Lenguaje Natural 69.

Álvarez-Carmona, M. Á., Villatoro-Tello, E., Villaseñor-Pineda, L., Montes-y Gómez, M., 2022b. Classifying the social media author profile through a multimodal representation. In: Intelligent Technologies: Concepts, Applications, and Future Directions. Springer, pp. 57–81.

Anderson, L.G., Rocliffe, S., Haddaway, N.R., Dunn, A.M., 2015. The role of tourism and recreation in the spread of non-native species: a systematic review and meta-analysis. PloS one 10, (10) e0140833.

Anis, S., Saad, S., Aref, M., 2020. A survey on sentiment analysis in tourism. Int. J. Intell. Comput. Inf. Sci. 20 (1), 1–15.

Antonio, N., de Almeida, A.M., Nunes, L., Batista, F., Ribeiro, R., 2018. Hotel online reviews: creating a multi-source aggregated index. Int. J. Contemporary Hospitality Manage. 30 (12), 3574–3591. https://doi.org/10.1108/ijchm-05-2017-0302.

Arce-Cardenas, S., Fajardo-Delgado, D., Álvarez-Carmona, M.Á., Ramírez-Silva, J.P., 2021. A tourist recommendation system: A study case in mexico. In: Mexican International Conference on Artificial Intelligence. Springer, pp. 184–195.

Argal, A., Gupta, S., Modi, A., Pandey, P., Shim, S., Choo, C., 2018. Intelligent travel chatbot for predictive recommendation in echo platform. In: 2018 IEEE 8th annual computing and communication workshop and conference (CCWC). IEEE, pp. 176–183.

Arote, S.S., Paikrao, R.L., 2018. A modified approach towards personalized travel recommendation system using sentiment analysis. In: 2018 International Conference On Advances in Communication and Computing Technology (ICACCT). IEEE. https://doi.org/10.1109/icacct.2018.8529327.

Arteaga, D., Arenas, J., Paz, F., Tupia, M., Bruzza, M., Jun. 2019. Design of information system architecture for the recommendation of tourist sites in the city of manta, ecuador through a chatbot. In: 2019 14th Iberian Conference on Information Systems and Technologies (CISTI). IEEE. doi: 10.23919/cisti.2019.8760669.

Athuraliya, B., Farook, C., 2018. Revyew hotel maintenance issue classifier and analyzer using machine learning and natural language processing. In: 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON). IEEE. https://doi.org/10.1109/iemcon.2018.8615075.

Bădică, C., Colhon, M., Şendre, A., Iftene, A., Mititelu, V.B., Cristea, D., Tufiş, D., 2014. Sentiment analysis of tourist reviews: Data preparation and preliminary results. In: CONFERENCE "LINGUISTIC RESOURCES AND TOOLS FOR PROCESSING THE ROMANIAN LANGUAGE" 18-19 SEPTEMBER 2014. p. 135. URL:http://consilr.info.uaic.ro/2014/Consilr_2014.pdf#page=143.

Barbado, R., Araque, O., Iglesias, C.A., 2019. A framework for fake review detection in online consumer electronics retailers. Inf. Process. Manage. 56 (4), 1234–1244. https://doi.org/10.1016/j.ipm.2019.03.002.

Barbosa, R.R.L., Sánchez-Alonso, S., Sicilia-Urban, M.A., 2015. Evaluating hotels rating prediction based on sentiment analysis services. Aslib J. Inf. Manage.

Becheru, A., Bădică, C., 2016. A deeper perspective of online tourism reviews analysis using natural language processing and complex networks techniques. In: Proceedings of the 12th International Conference Linguistic Resources and Tools for Processing the Romanian Language, ConsILR, pp. 189–192.

Becken, S., Alaei, A.R., Wang, Y., 2019. Benefits and pitfalls of using tweets to assess destination sentiment. J. Hospitality Tourism Technol. 11 (1), 19–34. https://doi.org/10.1108/jhtt-09-2017-0090.

Becken, S., Stantic, B., Chen, J., Alaei, A.R., Connolly, R.M., 2017. Monitoring the environment and human sentiment on the great barrier reef: Assessing the potential of collective sensing. J. Environ. Manage. 203, 87–97. https://doi.org/10.1016/j.jenvman.2017.07.007.

Berezina, K., Bilgihan, A., Cobanoglu, C., Okumus, F., 2015. Understanding satisfied and dissatisfied hotel customers: Text mining of online hotel reviews. J. Hospitality Market. Manage. 25 (1), 1–24. https://doi.org/10.1080/19368623.2015.983631.

Bhardwaj, P., Gautam, S., Pahwa, P., 2017. Opinion mining and sentiment analysis of travel websites through twitter. Int. J. Appl. Eng. Res. 12 (22), 12431–12439.

Bhargav, P.S., Reddy, G.N., Chand, R.R., Pujitha, K., Mathur, A., 2019. Sentiment analysis for hotel rating using machine learning algorithms. Int. J. Innov. Technol. Explor. Eng. 8 (6). URL: https://www.ijitee.org/wp-content/uploads/papers/v8i6/F3835048619.pdf.

Bi, J.-W., Liu, Y., Fan, Z.-P., 2019. Representing sentiment analysis results of online reviews using interval type-2 fuzzy numbers and its application to product ranking. Inf. Sci. 504, 293–307.

Biehl, A., Chen, Y., Sanabria-Véaz, K., Uttal, D., Stathopoulos, A., 2019. Where does active travel fit within local community narratives of mobility space and place? Transp. Res. Part A 123, 269–287. https://doi.org/10.1016/j.tra.2018.10.023.

Bigné, E., Oltra, E., Andreu, L., 2019. Harnessing stakeholder input on twitter: A case study of short breaks in spanish tourist cities. Tourism Manage. 71, 490–503.

Botchkarev, A., 2019. A new typology design of performance metrics to measure errors in machine learning regression algorithms. Interdiscip. J. Inf. Knowl. Manage. 14, 045–076.

Boukkouri, H.E., Ferret, O., Lavergne, T., Noji, H., Zweigenbaum, P., Tsujii, J., 2020. Characterbert: Reconciling elmo and bert for word-level open-vocabulary representations from characters. arXiv preprint arXiv:2010.10392.

Bucur, C., 2015. Using opinion mining techniques in tourism. Proc. Econ. Finance 23, 1666–1673. https://doi.org/10.1016/s2212-5671(15)00471-2.

Bueno, I., Carrasco, R.A., Ureña, R., Herrera-Viedma, E., 2019. Application of an opinion consensus aggregation model based on OWA operators to the recommendation of tourist sites. Proc. Comput. Sci. 162, 539–546. https://doi.org/10.1016/j.procs.2019.12.021.

Bulchand-Gidumal, J., 2020. Impact of artificial intelligence in travel, tourism, and hospitality. Handbook of e-Tourism; Xiang, Z., Fuchs, M., Gretzel, U., Höpken, W., Eds, 1–20.

Buzova, D., Sanz-Blas, S., Cervera-Taulet, A., 2016. Cross-cultural perceptions of onshore guided tours: A qualitative approach based on eWOM. Psychol. Market. 33 (12), 1054–1061. https://doi.org/10.1002/mar.20939.

Buzova, D., Sanz-Blas, S., Cervera-Taulet, A., 2018. Tour me onshore: understanding cruise tourists evaluation of shore excursions through text mining. J. Tourism Cultural Change 17 (3), 356–373. https://doi.org/10.1080/14766825.2018.1552277.

Cai, T., Giannopoulos, A.A., Yu, S., Kelil, T., Ripley, B., Kumamaru, K.K., Rybicki, F.J., Mitsouras, D., 2016. Natural language processing technologies in radiology research and clinical applications. Radiographics 36 (1), 176–191.

Calderón, F.A.C., Blanco, M.V.V., 2017. Impacto de internet en el sector turístico. Revista UNIANDES Episteme 4 (4), 477–490.

Çali, S., Balaman, Ş.Y., 2019. Improved decisions for marketing, supply and purchasing: Mining big data through an integration of sentiment analysis and intuitionistic fuzzy multi criteria assessment. Comput. Ind. Eng. 129, 315–332. https://doi.org/10.1016/j.cie.2019.01.051.

Campos, D., Silva, R.R., Bernardino, J., 2019. Text mining in hotel reviews: Impact of words restriction in text classification. In: KDIR. pp. 442–449.

Capdevila, J., Arias, M., Arratia, A., 2016. Geosrs: A hybrid social recommender system for geolocated data. Inf. Syst. 57, 111–128.

Carmona-Sánchez, G., Carmona, A., Álvarez-Carmona, M.A., 2021. Naive features for sentiment analysis on mexican touristic opinions texts. In: Proceedings of the Third Workshop for Iberian Languages Evaluation Forum (IberLEF 2021), CEUR WS Proceedings.

Castillo, O., 2012. Type-2 Fuzzy Logic in Intelligent Control Applications. Springer, Berlin Heidelberg. https://doi.org/10.1007/978-3-642-24663-0.

Celikyilmaz, A., Türksen, I., 2009. Modeling Uncertainty with Fuzzy Logic: With Recent Theory and Applications. Studies in Fuzziness and Soft Computing. Springer, Berlin Heidelberg. URL: https://books.google.com.mx/books?id=XNHzqfvMTbwC.

Chang, Y.-C., Ku, C.-H., Chen, C.-H., 2019. Social media analytics: Extracting and visualizing hilton hotel ratings and reviews from tripadvisor. Int. J. Inf. Manage. 48, 263–279.

Chanwisitkul, P., Shahgholian, A., Mehandjiev, N., 2018. The reason behind the rating: Text mining of online hotel reviews. In: 2018 IEEE 20th Conference on Business Informatics (CBI). IEEE. https://doi.org/10.1109/cbi.2018.00025.

Chatterjee, S., 2020. Drivers of helpfulness of online hotel reviews: A sentiment and emotion mining approach. Int. J. Hospitality Manage. 85, 102356.

Chen, J., Mao, Q., Xue, L., 2020. Visual sentiment analysis with active learning. IEEE Access 8, 185899–185908. https://doi.org/10.1109/access.2020.3024948.

Chen, M.-C., Hsiao, Y.-H., Chang, K.-C., Lin, M.-K., 2019. Applying big data analytics to support kansei engineering for hotel service development. Data Technol. Appl.

Cheng, M., 2017. A cross-cultural comparison of east and western academic literature on adventure tourism. Tourist Stud. 18 (4), 357–374. https://doi.org/10.1177/1468797617723472.

Cheng, M., Foley, C., 2018. The sharing economy and digital discrimination: The case of airbnb. Int. J. Hospitality Manage. 70, 95–98. https://doi.org/10.1016/j.ijhm.2017.11.002.

Cheng, M., Jin, X., 2019. What do airbnb users care about? An analysis of online review comments. Int. J. Hospitality Manage. 76, 58–70.
Chi, O.H., Denton, G., Gursoy, D., 2020. Artificially intelligent device use in service delivery: a systematic review, synthesis, and research agenda. J. Hospitality Market. Manage., 1–30
Chowdhury, G.G., 2003. Natural language processing. Annu. Rev. Inf. Sci. Technol. 37 (1), 51–89.
Colhon, M., Bădică, C., Şendre, A., 2014. Relating the opinion holder and the review accuracy in sentiment analysis of tourist reviews. In: Knowledge Science, Engineering and Management. Springer International Publishing, pp. 246–257. https://doi.org/10.1007/978-3-319-12096-6_22.
Colladon, A.F., Guardabascio, B., Innarella, R., 2019. Using social network and semantic analysis to analyze online travel forums and forecast tourism demand. Decis. Support Syst. 123,. https://doi.org/10.1016/j.dss.2019.113075 113075.
Cruz, F.L., Troyano, J.A., Enríquez, F., Ortega, F.J., Vallejo, C.G., 2013. 'Long autonomy or long delay?' The importance of domain in opinion mining. Expert Syst. Appl. 40 (8), 3174–3184. https://doi.org/10.1016/j.eswa.2012.12.031.
Dale, R., 2021. Gpt-3: What's it good for? Natural Lang. Eng. 27 (1), 113–118.
Dardas, A.Z., Williams, A., Scott, D., 2020. Carer-employees' travel behaviour: Assisted-transport in time and space. J. Transp. Geogr. 82, 102558.
De Lucia, C., Pazienza, P., Balena, P., Caporale, D., 2020. Exploring local knowledge and socio-economic factors for touristic attractiveness and sustainability. Int. J. Tourism Res. 22 (1), 81–99.
De Núñez, G., Maria, X., Núñez-Valdez, E.R., Pascual Espada, J., González-Crespo, R., Garcia-Díaz, V., 2018. A proposal for sentiment analysis on twitter for tourism-based applications. In: New Trends in Intelligent Software Methodologies, Tools and Techniques. Frontiers in Artificial Intelligence and Applications, pp. 713–722.
de Souza, J.G.R., de Paiva Oliveira, A., de Andrade, G.C., Moreira, A., 2018. A deep learning approach for sentiment analysis applied to hotel's reviews. In: International Conference on Applications of Natural Language to Information Systems. Springer, pp. 48–56.
Devika, M., Sunitha, C., Ganesh, A., 2016. Sentiment analysis: A comparative study on different approaches. Proc. Comput. Sci. 87, 44–49. https://doi.org/10.1016/j.procs.2016.05.124.
Díaz-Pacheco, A., Reyes-Garcia, C.A., 2021. A classification-based fuzzy-rules proxy model to assist in the full model selection problem in high volume datasets. J. Exp. Theor. Artif. Intell., 1–30
Dolce, P., Marocco, D., Maldonato, M.N., Sperandeo, R., Mar. 2020. Toward a machine learning predictive-oriented approach to complement explanatory modeling. an application for evaluating psychopathological traits based on affective neurosciences and phenomenology. Frontiers in Psychology 11. doi: 10.3389/fpsyg.2020.00446.
Duan, W., Cao, Q., Yu, Y., Levy, S., 2013. Mining online user-generated content: Using sentiment analysis technique to study hotel service quality. In: 2013 46th Hawaii International Conference on System Sciences. IEEE. https://doi.org/10.1109/hicss.2013.400.
Duan, W., Yu, Y., Cao, Q., Levy, S., 2016. Exploring the impact of social media on hotel service performance: A sentimental analysis approach. Cornell Hospitality Q. 57 (3), 282–296.
Elnagar, A., Khalifa, Y.S., Einea, A., 2018. Hotel arabic-reviews dataset construction for sentiment analysis applications. In: Intelligent Natural Language Processing: Trends and Applications. Springer, pp. 35–52.
Ert, E., Fleischer, A., 2016. Mere position effect in booking hotels online. J. Travel Res. 55 (3), 311–321.
Farisi, A.A., Sibaroni, Y., Faraby, S.A., 2019. Sentiment analysis on hotel reviews using multinomial naïve bayes classifier. J. Phys.: Conf. Ser. 1192,. https://doi.org/10.1088/1742-6596/1192/1/012024 012024.
Francesco, G., Roberta, G., 2019. Cross-country analysis of perception and emphasis of hotel attributes. Tourism Manage. 74, 24–42.
Fu, Y., Hao, J.-X., Li, X.R., Hsu, C.H., 2018. Predictive accuracy of sentiment analytics for tourism: A metalearning perspective on chinese travel news. J. Travel Res. 58 (4), 666–679. https://doi.org/10.1177/0047287518772361.
Fukui, M., Ohe, Y., 2019. Assessing the role of social media in tourism recovery in tsunami-hit coastal areas in tohoku, japan. Tourism Econ. 26 (5), 776–791. https://doi.org/10.1177/1354816618825014.
Fukumoto, F., Motegi, C., Matsuyoshi, S., 2012. Collaborative filtering based on sentiment analysis of guest reviews for hotel recommendation. In: KDIR. pp. 193–198.
Fukumoto, F., Sugiyama, H., Suzuki, Y., Matsuyoshi, S., 2014. Exploiting guest preferences with aspect-based sentiment analysis for hotel recommendation. In: International Joint Conference on Knowledge Discovery, Knowledge Engineering, and Knowledge Management. Springer, pp. 34–49.
Gal-Tzur, A., Grant-Muller, S.M., Kuflik, T., Minkov, E., Nocera, S., Shoor, I., 2014. The potential of social media in delivering transport policy goals. Transp. Policy 32, 115–123.
Gal-Tzur, A., Rechavi, A., Beimel, D., Freund, S., 2018. An improved methodology for extracting information required for transport-related decisions from q&a forums: A case study of tripadvisor. Travel Behaviour and Society 10, 1–9.
Gao, J., Yao, R., Lai, H., Chang, T.-C., 2019. Sentiment analysis with CNNs built on LSTM on tourists comments. In: 2019 IEEE Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS). IEEE. https://doi.org/10.1109/ecbios.2019.8807844.
Gao, S., Hao, J., Fu, Y., 2015. The application and comparison of web services for sentiment analysis in tourism. In: 2015 12th International Conference on Service Systems and Service Management (ICSSSM). IEEE. https://doi.org/10.1109/icsssm.2015.7170341.
García, A., Gaines, S., Linaza, M.T., et al., 2012. A lexicon based sentiment analysis retrieval system for tourism domain. Expert Syst. Appl. Int. J. 39 (10), 9166–9180.
García-Pablos, A., Cuadros, M., Linaza, M.T., 2016. Automatic analysis of textual hotel reviews. Inf. Technol. Tourism 16 (1), 45–69.
García-Pablos, A., Cuadros, M., Rigau, G., 2018. W2vlda: almost unsupervised system for aspect based sentiment analysis. Expert Syst. Appl. 91, 127–137.
Geetha, M., Singha, P., Sinha, S., 2017. Relationship between customer sentiment and online customer ratings for hotels - an empirical analysis. Tourism Manage. 61, 43–54. https://doi.org/10.1016/j.tourman.2016.12.022.
Geyer, R.C., Klein, T., Nabi, M., 2017. Differentially private federated learning: A client level perspective.
Gilpin, L.H., Bau, D., Yuan, B.Z., Bajwa, A., Specter, M., Kagal, L., 2018. Explaining explanations: An overview of interpretability of machine learning. In: 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA). IEEE. https://doi.org/10.1109/dsaa.2018.00018.
Gkritzali, A., Gritzalis, D., Stavrou, V., 2018. Is xenios zeus still alive? Destination image of athens in the years of recession. J. Travel Res. 57 (4), 540–554.
Godnov, U., Redek, T., 2016. Application of text mining in tourism: Case of croatia. Ann. Tourism Res. 58, 162–166.
Gomezelj, D.O., 2016. A systematic review of research on innovation in hospitality and tourism. Int. J. Contemporary Hospitality Manage.
Gonzalez-Rodriguez, M., Martinez-Torres, M., Toral, S., Dec. 2014. Monitoring travel-related information on social media through sentiment analysis. In: 2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing. IEEE. doi: 10.1109/ucc.2014.102.
González-Rodríguez, M.R., Martínez-Torres, R., Toral, S., 2016. Post-visit and pre-visit tourist destination image through eWOM sentiment analysis and perceived helpfulness. Int. J. Contemporary Hospitality Manage. 28 (11), 2609–2627. https://doi.org/10.1108/ijchm-02-2015-0057.
Gu, Y.H., Yoo, S.J., Jiang, Z., Lee, Y.J., Piao, Z., Yin, H., Jeon, S., Jan. 2018. Sentiment analysis and visualization of chinese tourism blogs and reviews. In: 2018 International Conference on Electronics, Information, and Communication (ICEIC). IEEE. doi: 10.23919/elinfocom.2018.8330589.
Guerreiro, C., Viegas, M., Guerreiro, M., 2019. Social networks and digital influencers: Their role in customer decision journey in tourism. J. Spatial Organiz. Dyn. 7 (3), 240–260.
Guerreiro, J., Rita, P., 2020. How to predict explicit recommendations in online reviews using text mining and sentiment analysis. J. Hospitality Tourism Manage. 43, 269–272.
Guerrero-Rodriguez, R., Álvarez-Carmona, M.Á., Aranda, R., López-Monroy, A.P., 2021. Studying online travel reviews related to tourist attractions using nlp methods: the case of guanajuato, mexico. Curr. Issues Tourism, 1–16.
Gupta, G., Gupta, P., 2019. Twitter mining for sentiment analysis in tourism industry. In: 2019 Third World Conference on Smart Trends in Systems Security and Sustainablity (WorldS4). IEEE. doi: 10.1109/worlds4.2019.8903940.
Gupta, K., Gauba, T., Jain, S., 2017. Big data in hospitality industry: A survey. Int. Res. J. Eng. Technol. 4 (11), 476–479.
Hardeniya, N., Perkins, J., Chopra, D., Joshi, N., Mathur, I., 2016. Natural language processing: python and NLTK. Packt Publishing Ltd..
He, W., Tian, X., Tao, R., Zhang, W., Yan, G., Akula, V., 2017. Application of social media analytics: a case of analyzing online hotel reviews. Online Inf. Rev. 41 (7), 921–935. https://doi.org/10.1108/oir-07-2016-0201.
Hermanto, D.T., Ziaurrahman, M., Bianto, M.A., Setyanto, A., Dec 2018. Twitter social media sentiment analysis in tourist destinations using algorithms naive bayes classifier. J. Phys: Conf. Ser. 1140,. https://doi.org/10.1088/1742-6596/1140/1/012037 012037.
Hu, F., Trivedi, R.H., 2020. Mapping hotel brand positioning and competitive landscapes by text-mining user-generated content. Int. J. Hospitality Manage. 84, 102317.
Hu, N., Zhang, T., Gao, B., Bose, I., 2019. What do hotel customers complain about? text analysis using structural topic model. Tourism Manage. 72, 417–426. https://doi.org/10.1016/j.tourman.2019.01.002.
Hu, Y.-H., Chen, K., 2016. Predicting hotel review helpfulness: The impact of review visibility, and interaction between hotel stars and review ratings. Int. J. Inf. Manage. 36 (6), 929–944.
Hu, Y.-H., Chen, K., Lee, P.-J., 2017. The effect of user-controllable filters on the prediction of online hotel reviews. Inf. Manage. 54 (6), 728–744. https://doi.org/10.1016/j.im.2016.12.009.
Hu, Y.-H., Chen, Y.-L., Chou, H.-L., 2017. Opinion mining from online hotel reviews–a text summarization approach. Inf. Process. Manage. 53 (2), 436–449.
Huang, H.-H., Wang, J.-J., Chen, H.-H., 2017. Implicit opinion analysis: Extraction and polarity labelling. J. Assoc. Inf. Sci. Technol. 68 (9), 2076–2087. https://doi.org/10.1002/asi.23835.
Irawan, H., Akmalia, G., Masrury, R.A., 2019. Mining tourist's perception toward indonesia tourism destination using sentiment analysis and topic modelling. In: Proceedings of the 2019 4th International Conference on Cloud Computing and Internet of Things. pp. 7–12.
Jianlei, H., Qingzhong, M., Changwen, X., Pengfei, S., Xiaotong, W., Keying, Y., 2019. Analysis and implications of travel experience and perception in border ethnic areas based on online text mining——a case study of xishuangbanna in yunnan province. J. Phys: Conf. Ser. 1325,. https://doi.org/10.1088/1742-6596/1325/1/012053 012053.

Jocuns, A., 2018. English in thai tourism: Global english as a nexus of practice. In: Language and Literature in a Glocal World. Springer, pp. 57–76.

Kennedy-Eden, H., Gretzel, U., 2012. A taxonomy of mobile applications in tourism. pp. 47–50. URL: https://ro.uow.edu.au/commpapers/2510/.

Khotimah, D., Sarno, R.S., 2019. Sentiment analysis of hotel aspect using probabilistic latent semantic analysis, word embedding and LSTM. Int. J. Intell. Eng. Syst. 12 (4), 275–290. doi: 10.22266/ijies2019.0831.26

Kim, J., Im, J., 2018. Proposing a missing data method for hospitality research on online customer reviews. Int. J. Contemporary Hospitality Manage. 30 (11), 3250–3267. https://doi.org/10.1108/ijchm-10-2017-0708.

Kim, K., Park, O.-J., Yun, S., Yun, H., 2017. What makes tourists feel negatively about tourism destinations? application of hybrid text mining methodology to smart destination management. Technol. Forecast. Soc. Chang. 123, 362–369.

Kim, S., Lee, W.S., 2019. Network text analysis of medical tourism in newspapers using text mining: The south korea case. Tourism Manage. Perspect. 31, 332–339.

Kirilenko, A.P., Stepchenkova, S.O., Kim, H., Li, X., 2018. Automated sentiment analysis in tourism: Comparison of approaches. J. Travel Res. 57 (8), 1012–1025.

Költringer, C., Dickinger, A., 2015. Analyzing destination branding and image from online sources: A web content mining approach. J. Business Res. 68 (9), 1836–1843.

Kranjc, J., Smailović, J., Podpečan, V., Grčar, M., Žnidaršič, M., Lavrač, N., 2015. Active learning for sentiment analysis on data streams: Methodology and workflow implementation in the ClowdFlows platform. Inf. Process. Manage. 51 (2), 187–203. https://doi.org/10.1016/j.ipm.2014.04.001.

Kuhamanee, T., Talmongkol, N., Chaisuriyakul, K., San-Um, W., Pongpisuttinun, N., Pongyupinpanich, S., 2017. Sentiment analysis of foreign tourists to bangkok using data mining through online social network. In: 2017 IEEE 15th International Conference on Industrial Informatics (INDIN). IEEE. doi: 10.1109/indin.2017.8104921.

Kuhzady, S., Ghasemi, V., 2019. Factors influencing customers' satisfaction and dissatisfaction with hotels: A text-mining approach. Tourism Anal. 24 (1), 69–79. https://doi.org/10.3727/108354219x15458295631972.

Kurniawan, S., Kusumaningrum, R., Timu, M.E., 2018. Hierarchical sentence sentiment analysis of hotel reviews using the naïve bayes classifier. In: 2018 2nd International Conference on Informatics and Computational Sciences (ICICoS). IEEE, pp. 1–5.

Lai, S.T., Raheem, M., 2020. Sentiment analysis of online customer reviews for hotel industry: An appraisal of hybrid approach. International Research Journal of Engineering and Technology (IRJET) 7 (12), 1355–1359. URL: https://www.irjet.net/archives/V7/i12/IRJET-V7I12238.pdf.

Landauer, T.K., McNamara, D.S., Dennis, S., Kintsch, W., 2013. Handbook of latent semantic analysis. Psychology Press.

Lee, C.-H., Wang, S.-H., 2012. An information fusion approach to integrate image annotation and text mining methods for geographic knowledge discovery. Expert Syst. Appl. 39 (10), 8954–8967.

Lee, M., Jeong, M., Lee, J., 2017. Roles of negative emotions in customers' perceived helpfulness of hotel reviews on a user-generated review website. Int. J. Contemporary Hospitality Manage. 29 (2), 762–783. https://doi.org/10.1108/ijchm-10-2015-0626.

Lee, M., Lee, S.A., Koh, Y., 2019. Multisensory experience for enhancing hotel guest experience. Int. J. Contemporary Hospitality Manage. 31 (11), 4313–4337. https://doi.org/10.1108/ijchm-03-2018-0263.

Lee, M.J., Singh, N., Chan, E.S., 2011. Service failures and recovery actions in the hotel industry: A text-mining approach. J. Vacation Market. 17 (3), 197–207. https://doi.org/10.1177/1356766711409182.

Lee, P.-J., Hu, Y.-H., Lu, K.-T., 2018. Assessing the helpfulness of online hotel reviews: A classification-based approach. Telematics Inform. 35 (2), 436–445.

Lee, T.Y., Bradlow, E.T., 2011. Automated marketing research using online customer reviews. J. Mark. Res. 48 (5), 881–894. https://doi.org/10.1509/jmkr.48.5.881.

Li, J., Xu, L., Tang, L., Wang, S., Li, L., 2018. Big data in tourism research: A literature review. Tourism Manage. 68, 301–323.

Li, Q., Li, S., Zhang, S., Hu, J., Hu, J., 2019. A review of text corpus-based tourism big data mining. Appl. Sci. 9 (16), 3300.

Li, W., Guo, K., Shi, Y., Zhu, L., Zheng, Y., 2018. DWWP: Domain-specific new words detection and word propagation system for sentiment analysis in the tourism domain. Knowl.-Based Syst. 146, 203–214. https://doi.org/10.1016/j.knosys.2018.02.004.

Song, W.l., Wang, J.h., 2017. Sentiment analysis of tourism micro-blog comments. DEStech Transactions on Computer Science and Engineering (aita). doi: 10.12783/dtcse/aita2016/7594.

Liang, X., Liu, P., Wang, Z., 2019. Hotel selection utilizing online reviews: a novel decision support model based on sentiment analysis and DL-VIKOR method. Technol. Econ. Develop. Econ. 25 (6), 1139–1161. https://doi.org/10.3846/tede.2019.10766.

Liberati, A., Altman, D.G., Tetzlaff, J., Mulrow, C., Gøtzsche, P.C., Ioannidis, J.P., Clarke, M., Devereaux, P.J., Kleijnen, J., Moher, D., 2009. The prisma statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J. Clin. Epidemiol. 62 (10), e1–e34.

Liddy, E. D. (2001). Natural language processing in Encyclopedia of Library and Information Science 2nd ed., New York: Marcel Decker.

Lin, K.-P., Lai, C.-Y., Chen, P.-C., Hwang, S.-Y., 2015. Personalized hotel recommendation using text mining and mobile browsing tracking. In: 2015 IEEE International Conference on Systems, Man, and Cybernetics. IEEE, pp. 191–196.

Linares, R., Herrera, J., Cuadros, A., Alfaro, L., Oct. 2015. Prediction of tourist traffic to peru by using sentiment analysis in twitter social network. In: 2015 Latin American Computing Conference (CLEI). IEEE. doi: 10.1109/clei.2015.7360051.

Linfeng, Dong, Haoyu, Wang, Chenglin, Zhang, Analysis of sanya's visitors preference based on text mining of tourism forum. In: Proceedings of 2017 Symposium on Service Innovation in Big Data Era.

Liu, M.T., Liu, Y., Mo, Z., Ng, K.L., 2020. Using text mining to track changes in travel destination image: the case of macau. Asia Pacific J. Market. Logist.

Liu, P., Nie, D., He, X., Zhang, W., Huang, Z., He, K., Jul. 2019a. Sentiment analysis of chinese tourism review based on boosting and LSTM. In: 2019 International Conference on Communications, Information System and Computer Engineering (CISCE). IEEE. doi: 10.1109/cisce.2019.00154.

Liu, S., Yuan, T., Feng, Y., Zhuang, Y., 2018. Comparison of tourist thematic sentiment analysis methods based on weibo data. Acta Scientiarum Naturalium Universitatis Pekinensis.

Liu, Y., Huang, K., Bao, J., Chen, K., 2019b. Listen to the voices from home: An analysis of chinese tourists' sentiments regarding australian destinations. Tourism Manage. 71, 337–347. https://doi.org/10.1016/j.tourman.2018.10.004.

Loaiza, J.V., Pérez-Torres, A., Contreras, K.D., 2019. Semantic icons: A sentiment analysis as a contribution to sustainable tourism. Sustainability 11 (17), 4655. https://doi.org/10.3390/su11174655.

Loyola-González, O., Medina-Pérez, M.A., Choo, K.-K.R., 2020. A review of supervised classification based on contrast patterns: Applications, trends, and challenges. J. Grid Comput., 1–49

Lu, W., Liu, Z., Huang, Y., Bu, Y., Li, X., Cheng, Q., 2020. How do authors select keywords? a preliminary study of author keyword selection behavior. J. Inf. 14, (4) 101066.

Luo, Q., Zhai, X., 2017. i will never go to hong kong again! how the secondary crisis communication of occupy central on weibo shifted to a tourism boycott. Tourism Manage. 62, 159–172.

Ma, E., Cheng, M., Hsiao, A., 2018. Sentiment analysis – a review and agenda for future research in hospitality contexts. Int. J. Contemporary Hospitality Manage. 30 (11), 3287–3308. https://doi.org/10.1108/ijchm-10-2017-0704.

Ma, Z., Du, J., Zhou, Y., 2015. Sentiment analysis based on evaluation of tourist attractions. In: Proceedings of the 2015 Chinese Intelligent Systems Conference. Springer, Berlin Heidelberg, pp. 375–382. doi: 10.1007/978-3-662-48386-2_39.

Maghrebi, M., Abbasi, A., Rashidi, T.H., Waller, S.T., Sep. 2015. Complementing travel diary surveys with twitter data: Application of text mining techniques on activity location, type and time. In: 2015 IEEE 18th International Conference on Intelligent Transportation Systems. IEEE. doi: 10.1109/itsc.2015.43.

Mao, Z., Yang, Y., Wang, M., 2018. Sleepless nights in hotels? understanding factors that influence hotel sleep quality. International Journal of Hospitality Management 74, 189–201. https://doi.org/10.1016/j.ijhm.2018.05.002.

Marasco, A., De Martino, M., Magnotti, F., Morvillo, A., 2018. Collaborative innovation in tourism and hospitality: a systematic review of the literature. Int. J. Contemporary Hospitality Manage.

Markopoulos, G., Mikros, G., Iliadi, A., Liontos, M., 2015. Sentiment analysis of hotel reviews in greek: A comparison of unigram features. In: Cultural Tourism in a Digital Era. Springer International Publishing, pp. 373–383. https://doi.org/10.1007/978-3-319-15859-4_31.

Marr, B., 2016. How big data and analytics are changing hotels and the hospitality industry. Forbes. com, January 26.

Martins, G.S., de Paiva Oliveira, A., Moreira, A., 2017. Sentiment analysis applied to hotels evaluation. In: Computational Science and Its Applications – ICCSA 2017. Springer International Publishing, pp. 710–716. https://doi.org/10.1007/978-3-319-62407-5_52.

Masrury, R.A., Alamsyah, A., et al., 2019. Analyzing tourism mobile applications perceived quality using sentiment analysis and topic modeling. In: 2019 7th International Conference on Information and Communication Technology (ICoICT). IEEE, pp. 1–6.

Mathayomchan, B., Sripanidkulchai, K., 2019. Utilizing google translated reviews from google maps in sentiment analysis for phuket tourist attractions. In: 2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE). IEEE. doi: 10.1109/jcsse.2019.8864150.

Mazanec, J.A., 2017. Determining long-term change in tourism research language with text-mining methods. Tourism Analysis 22 (1), 75–83. https://doi.org/10.3727/108354217x14828625279771.

Meehan, K., Lunney, T., Curran, K., McCaughey, A., 2016. Aggregating social media data with temporal and environmental context for recommendation in a mobile tour guide system. J. Hospitality Tourism Technol.

Mena-Maldonado, E., Sinnott, R., Tenemaza, M., 2016. Study of the viability of using twitter sentiment analysis in the hotel industry. pp. 13–19, cited By 1. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85040251043&partnerID=40&md5=62e074f7ccb31cb403ef49cf584f0f89.

Menchavez, J.C.L., Espinosa, K.J.P., 2015. Fun in the philippines: Automatic identification and sentiment analysis of tourism-related tweets. In: 2015 IEEE International Conference on Data Mining Workshop (ICDMW). IEEE, pp. 660–667.

Micera, R., Crispino, R., 2017. Destination web reputation as "smart tool for image building: the case analysis of naples city-destination. International Journal of Tourism Cities 3 (4), 406–423. https://doi.org/10.1108/ijtc-11-2016-0048.

Micu, A., Micu, A.E., Geru, M., Lixandroiu, R.C., 2017. Analyzing user sentiment in social media: Implications for online marketing strategy. Psychol. Market. 34 (12), 1094–1100. https://doi.org/10.1002/mar.21049.

Mishra, D., Venugopalan, M., Gupta, D., 2016. Context specific lexicon for hindi reviews. Proc. Comput. Sci. 93, 554–563. https://doi.org/10.1016/j.procs.2016.07.283.
Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., Group, P., et al., 2009. Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. PLoS med 6 (7), e1000097.
Moon, S., Kamakura, W.A., 2017. A picture is worth a thousand words: Translating product reviews into a product positioning map. Int. J. Res. Mark. 34 (1), 265–285. https://doi.org/10.1016/j.ijresmar.2016.05.007.
Moreno-Ortiz, A., Salles-Bernal, S., Orrequia-Barea, A., 2019. Design and validation of annotation schemas for aspect-based sentiment analysis in the tourism sector. Inf. Technol. Tourism 21 (4), 535–557. https://doi.org/10.1007/s40558-019-00155-0.
Moro, S., Ramos, P., Esmerado, J., Jalali, S.M.J., 2019. Can we trace back hotel online reviews' characteristics using gamification features? Int. J. Inf. Manage. 44, 88–95.
Moro, S., Rita, P., 2018. Brand strategies in social media in hospitality and tourism. Int. J. Contemporary Hospitality Manage. 30 (1), 343–364. https://doi.org/10.1108/ijchm-07-2016-0340.
Moro, S., Rita, P., Ramos, P., Esmerado, J., 2019. Analysing recent augmented and virtual reality developments in tourism. J. Hospitality Tourism Technol. 10 (4), 571–586. https://doi.org/10.1108/jhtt-07-2018-0059.
Moubaiddin, A., Shalbak, O., Hammo, B., Obeid, N., 2015. Arabic dialogue system for hotel reservation based on natural language processing techniques. Computación y Sistemas 19 (1), 119–134.
Nadkarni, P.M., Ohno-Machado, L., Chapman, W.W., 2011. Natural language processing: an introduction. J. Am. Med. Inform. Assoc. 18 (5), 544–551.
Nadkarni, S., Kriechbaumer, F., Rothenberger, M., Christodoulidou, N., 2019. The path to the hotel of things: Internet of things and big data converging in hospitality. J. Hospitality Tourism Technol.
Nakayama, M., Wan, Y., 2018. Is culture of origin associated with more expressions? an analysis of yelp reviews on japanese restaurants. Tourism Manage. 66, 329–338. https://doi.org/10.1016/j.tourman.2017.10.019.
Nawangsari, R.P., Kusumaningrum, R., Wibowo, A., 2019. Word2vec for indonesian sentiment analysis towards hotel reviews: An evaluation study. Proc. Comput. Sci. 157, 360–366. https://doi.org/10.1016/j.procs.2019.08.178.
Neidhardt, J., Rümmele, N., Werthner, H., 2017. Predicting happiness: user interactions and sentiment analysis in an online travel forum. Inf. Technol. Tourism 17 (1), 101–119. https://doi.org/10.1007/s40558-017-0079-2.
Nguyen-Thanh, T., Tran, G.T., 2019. Vietnamese sentiment analysis for hotel review based on overfitting training and ensemble learning. In: Proceedings of the Tenth International Symposium on Information and Communication Technology - SoICT 2019. ACM Press. doi: 10.1145/3368926.3369675.
Nica, I., Tazl, O.A., Wotawa, F., Chatbot-based tourist recommendations using model-based reasoning. pp. 25–30. URL:http://ceur-ws.org/Vol-2220/#05_CONFWS18_paper_31.
Nisar, T.M., Prabhakar, G., 2018. Trains and twitter: Firm generated content, consumer relationship management and message framing. Transp. Res. Part A 113, 318–334. https://doi.org/10.1016/j.tra.2018.04.026.
Pacheco, F.S., Jiménez, A.R., Campos, Á.R.A., Carmona, M.Á.Á., 2021. Técnicas automatizadas para la adquisición de datos turísticos en comparación con técnicas tradicionales. KIKAME 11 (11), 101–115.
Panawong, N., Sittisaman, A., 2019. Tourism web filtering and analysis using naïve bay with boundary values and text mining. In: Advances in Intelligent Systems and Computing. Springer Singapore, pp. 535–547. https://doi.org/10.1007/978-981-13-6861-5_46.
Panigrahi, N., Asha, T., 2018. Aspect-level sentiment analysis on hotel reviews. In: Advances in Intelligent Systems and Computing. Springer Singapore, pp. 379–389. https://doi.org/10.1007/978-981-10-8055-5_34.
Panigrahi, N., T, A., Nov. 2018. RHALSA: Ranking hotels using aspect level sentiment analysis. Journal of Computer Science 14 (11), 1512–1520. doi: 10.3844/jcssp.2018.1512.1520
Park, E., Kang, J., Choi, D., Han, J., 2018. Understanding customers' hotel revisiting behaviour: a sentiment analysis of online feedback reviews. Curr. Issues Tourism 23 (5), 605–611. https://doi.org/10.1080/13683500.2018.1549025.
Park, S.B., Jang, J., Ok, C.M., 2016. Analyzing twitter to explore perceptions of asian restaurants. J. Hospitality Tourism Technol. 7 (4), 405–422. https://doi.org/10.1108/jhtt-08-2016-0042.
Pham, D.-H., Le, A.-C., 2018. Learning multiple layers of knowledge representation for aspect based sentiment analysis. Data Knowl. Eng. 114, 26–39. https://doi.org/10.1016/j.datak.2017.06.001.
Phillips, P., Antonio, N., de Almeida, A., Nunes, L., 2019. The influence of geographic and psychic distance on online hotel ratings. J. Travel Res. 59 (4), 722–741. https://doi.org/10.1177/0047287519858400.
Pinarbasi, F., Taskiran, M., 2020. Natural Language Processing for Global and Local Business. Advances in Business Information Systems and Analytics. IGI Global. URL: https://books.google.com.mx/books?id=ALkIEAAAQBAJ.
Poernomo, A., 2019. Indonesian online travel agent sentiment analysis using machine learning methods. Indones. J. Electron. Eng. Comput. Sci. 14 (1), 117–122.
Prameswari, P., Surjandari, I., Laoh, E., 2017. Opinion mining from online reviews in bali tourist area. In: 2017 3rd International Conference on Science in Information Technology (ICSITech). IEEE, pp. 226–230.
Prameswari, P., Surjandari, I., Laoh, E., et al., 2017. Mining online reviews in indonesia's priority tourist destinations using sentiment analysis and text summarization approach. In: 2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST). IEEE, pp. 121–126.
Priyantina, R., Sarno, R., 2019. Sentiment analysis of hotel reviews using latent dirichlet allocation, semantic similarity and LSTM. International Journal of Intelligent Engineering and Systems 12 (4), 142–155. doi: 10.22266/ijies2019.0831.14
Punel, A., Ermagun, A., 2018. Using twitter network to detect market segments in the airline industry. Journal of Air Transport Management 73, 67–76. https://doi.org/10.1016/j.jairtraman.2018.08.004.
Putri, F.P., Meidia, H., Gunawan, D., 2019. Designing intelligent personalized chatbot for hotel services. In: Proceedings of the 2019 2nd International Conference on Algorithms, Computing and Artificial Intelligence. pp. 468–472.
Putri, I., Kusumaningrum, R., 2017. Latent dirichlet allocation (lda) for sentiment analysis toward tourism review in indonesia. In: Journal of Physics: Conference Series. Vol. 801. IOP Publishing, p. 012073.
Qi, M., Li, X., Zhu, E., Shi, Y., 2017. Evaluation of perceived indoor environmental quality of five-star hotels in china: An application of online review analysis. Build. Environ. 111, 1–9. https://doi.org/10.1016/j.buildenv.2016.09.027.
Rahmani, K., Gnoth, J., Mather, D., 2018. Hedonic and eudaimonic well-being: A psycholinguistic view. Tourism Management 69, 155–166. https://doi.org/10.1016/j.tourman.2018.06.008.
Ramanathan, V., Meyyappan, T., 2019. Twitter text mining for sentiment analysis on people's feedback about oman tourism. In: 2019 4th MEC International Conference on Big Data and Smart City (ICBDSC). IEEE. doi: 10.1109/icbdsc.2019.8645596.
Ravindran, A., Rejikumar, G., 2017. An investigation on service quality attributes of indian hotels by text mining. In: 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC). IEEE. doi: 10.1109/iccic.2017.8524566.
Ren, G., Hong, T., 2017. Investigating online destination images using a topic-based sentiment analysis approach. Sustainability 9 (10), 1765.
Reyes-Menendez, A., Saura, J.R., Filipe, F., 2019. The importance of behavioral data to identify online fake reviews for tourism businesses: A systematic review. PeerJ Computer Science 5, e219.
Rodríguez-Ruiz, J., Medina-Pérez, M.A., Monroy, R., Loyola-González, O., 2019. A survey on minutiae-based palmprint feature representations, and a full analysis of palmprint feature representation role in latent identification performance. Expert Syst. Appl. 131, 30–44.
Romero-Cantón, A., Aranda, R., 2021. Sentiment classification for mexican tourist reviews based on k-nn and jaccard distance. In: Proceedings of the third workshop for iberian languages evaluation forum (iberlef 2021), ceur ws proceedings.
Roukos, S., Soffer, A., 2020. IBM Research addressing Enterprise NLP challenges in 2020 |IBM Research Blog. [Online; accessed 14. Aug. 2022]. URL: https://www.ibm.com/blogs/research/2020/06/advancingnlp2020.
Ruhanen, L., lee Moyle, C., Moyle, B., Apr. 2019. New directions in sustainable tourism research. Tourism Review 74 (2), 138–149. doi: 10.1108/tr-12-2017-0196
Rybakov, V., Malafeev, A., 2018. Aspect-based sentiment analysis of russian hotel reviews. In: CEUR Workshop Proceedings. pp. 75–84.
Salur, M.U., Aydin, I., Alghrsi, S.A., Sep. 2019. SmartSenti: A twitter-based sentiment analysis system for the smart tourism in turkey. In: 2019 International Artificial Intelligence and Data Processing Symposium (IDAP). IEEE. doi: 10.1109/idap.2019.8875922.
Samara, D., Magnisalis, I., Peristeras, V., 2020. Artificial intelligence and big data in tourism: a systematic literature review. J. Hospitality Tourism Technol.
Sanchez-Franco, M.J., Cepeda-Carrion, G., Roldán, J.L., 2019. Understanding relationship quality in hospitality services. Internet Research.
Sano, A.V.D., Imanuel, T.D., Calista, M.I., Nindito, H., Condrobimo, A.R., 2018. The application of agnes algorithm to optimize knowledge base for tourism chatbot. In: 2018 International Conference on Information Management and Technology (ICIMTech). IEEE, pp. 65–68.
Santhappan, J., Chokkalingam, P., 2020. A comparative analysis of predictive modeling techniques: A case study of device failure. In: Machine Learning and Information Processing. Springer, pp. 223–233.
Sanz-Blas, S., Buzova, D., 2016. Guided tour influence on cruise tourist experience in a port of call: An eWOM and questionnaire-based approach. International Journal of Tourism Research 18 (6), 558–566. https://doi.org/10.1002/jtr.2073.
Schwartz, A.J., Dodds, P.S., O'Neil-Dunne, J.P.M., Danforth, C.M., Ricketts, T.H., 2019. Visitors to urban greenspace have higher sentiment and lower negativity on twitter. People and Nature 1 (4), 476–485. https://doi.org/10.1002/pan3.10045.
Scorrano, P., Fait, M., Maizza, A., Vrontis, D., 2019. Online branding strategy for wine tourism competitiveness. International Journal of Wine Business Research 31 (2), 130–150. https://doi.org/10.1108/ijwbr-06-2017-0043.
Seifert, A., Rössel, J., 2019. Digital Participation. Springer International Publishing, Cham, pp. 1–5. https://doi.org/10.1007/978-3-319-69892-2_1017-1.
Sekar, A., Chen, R.B., Cruzat, A., Nagappan, M., 2017. Digital narratives of place: Learning about neighborhood sense of place and travel through online responses. Transportation Research Record: Journal of the Transportation Research Board 2666 (1), 10–18. https://doi.org/10.3141/2666-02.
Serna, A., Gasparovic, S., 2018. Transport analysis approach based on big data and text mining analysis from social media. Transportation Research Procedia 33, 291–298. https://doi.org/10.1016/j.trpro.2018.10.105.
Serna, A., Gerrikagoitia, J.K., Bernabé, U., Ruiz, T., 2017. Sustainability analysis on urban mobility based on social media content. Transportation Research Procedia 24, 1–8.

Sezgen, E., Mason, K.J., Mayer, R., 2019. Voice of airline passenger: A text mining approach to understand customer satisfaction. Journal of Air Transport Management 77, 65–74. https://doi.org/10.1016/j.jairtraman.2019.04.001.

Shi, H.-X., Li, X.-J., 2011. A sentiment analysis model for hotel reviews based on supervised learning. 2011 International Conference on Machine Learning and Cybernetics, vol. 3. IEEE, pp. 950–954.

Shin, S.-H., Yang, S.-B., Nam, K., Koo, C., 2017. Conceptual foundations of a landmark personality scale based on a destination personality scale: Text mining of online reviews. Information Systems Frontiers 19 (4), 743–752.

Shuai, Q., Huang, Y., Jin, L., Pang, L., Oct. 2018. Sentiment analysis on chinese hotel reviews with doc2vec and classifiers. In: 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC). IEEE. doi: 10.1109/iaeac.2018.8577581.

Siddaway, A.P., 2014. What is a systematic literature review and how do i do one. In: Pueda. pp. 1–13.

Siering, M., Deokar, A.V., Janze, C., 2018. Disentangling consumer recommendations: Explaining and predicting airline recommendations based on online reviews. Decis. Support Syst. 107, 52–63.

Siow, M.-L., Ramachandran, S., Shuib, A., Afandi, S.H.M., 2015. Adapting evidence-based intervention in rural tourism policies. Worldwide Hospitality and Tourism Themes 7 (5), 473–485. https://doi.org/10.1108/whatt-06-2015-0031.

Smailović, J., Grčar, M., Lavrač, N., Žnidaršič, M., 2014. Stream-based active learning for sentiment analysis in the financial domain. Inf. Sci. 285, 181–203. https://doi.org/10.1016/j.ins.2014.04.034.

Sodanil, M., 2016. Multi-language sentiment analysis for hotel reviews. MATEC Web of Conferences 75, 03002. https://doi.org/10.1051/matecconf/20167503002.

Sokolova, M., Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. Information processing & management 45 (4), 427–437.

Srivastava, S., Prabhakar, T., 2019. Hospitality of chatbot building platforms. In: Proceedings of the 2nd ACM SIGSOFT International Workshop on Software Qualities and Their Dependencies - SQUADE 2019. ACM Press. doi: 10.1145/3340495.3342751.

Suardika, I.G., 2016. Sentiment analysis system and correlation analysis on hospitality in bali. Journal of Theoretical and Applied Information Technology 84 (1), 88. URL: http://www.jatit.org/volumes/Vol84No1/10Vol84No1.pdf.

Sun, L., Guo, J., Zhu, Y., 2018. Applying uncertainty theory into the restaurant recommender system based on sentiment analysis of online chinese reviews. World Wide Web 22 (1), 83–100. https://doi.org/10.1007/s11280-018-0533-x.

Sun, L., Guo, J., Zhu, Y., 2018. A multi-aspect user-interest model based on sentiment analysis and uncertainty theory for recommender systems. Electronic Commerce Research, 1–26.

Sutabri, T., Putra, S.J., Effendi, M.R., Gunawan, M.N., Napitupulu, D., Aug. 2018a. Sentiment analysis for popular e-traveling sites in indonesia using naive bayes. In: 2018 6th International Conference on Cyber and IT Service Management (CITSM). IEEE. doi: 10.1109/citsm.2018.8674262.

Sutabri, T., Suryatno, A., Setiadi, D., Negara, E.S., Oct. 2018b. Improving naïve bayes in sentiment analysis for hotel industry in indonesia. In: 2018 Third International Conference on Informatics and Computing (ICIC). IEEE. doi: 10.1109/iac.2018.8780444.

Suzuki, T., Gemba, K., Aoyama, A., 2013. Hotel classification visualization using natural language processing of user reviews. In: 2013 IEEE International Conference on Industrial Engineering and Engineering Management. IEEE, pp. 892–895.

Tao, W., Zhang, Q., Zhang, M., Li, Y., May 2019a. Mining pain points from hotel online comments based on sentiment analysis. In: 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC). IEEE. doi: 10.1109/itaic.2019.8785893.

Tao, Y., Zhang, F., Shi, C., Chen, Y., 2019. Social media data-based sentiment analysis of tourists' air quality perceptions. Sustainability 11 (18), 5070. https://doi.org/10.3390/su11185070.

Tay, Y., Dehghani, M., Bahri, D., Metzler, D., 2020. Efficient transformers: A survey. arXiv preprint arXiv:2009.06732.

Thelwall, M., 2018. Gender bias in sentiment analysis. Online Information Review 42 (1), 45–57. https://doi.org/10.1108/oir-05-2017-0139.

Thelwall, M., 2019. Sentiment analysis for tourism. In: Big Data and Innovation in Tourism, Travel, and Hospitality. Springer Singapore, pp. 87–104. https://doi.org/10.1007/978-981-13-6339-9_6.

Tjahyanto, A., Sisephaputra, B., 2017. The utilization of filter on object-based opinion mining in tourism product reviews. Procedia Computer Science 124, 38–45. https://doi.org/10.1016/j.procs.2017.12.127.

Tolkach, D., 2018. A view of occupy central impacts on hong kong tourism from the other side of the great firewall: A rejoinder to luo & zhai. Tourism Management 67, 307–311. https://doi.org/10.1016/j.tourman.2018.02.011.

Tran, T., Ba, H., Huynh, V.-N., 2019. Measuring hotel review sentiment: An aspect-based sentiment analysis approach. In: Lect. Notes Comput. Sci.. Springer International Publishing, pp. 393–405. https://doi.org/10.1007/978-3-030-14815-7_33.

Tranfield, D., Denyer, D., Smart, P., 2003. Towards a methodology for developing evidence-informed management knowledge by means of systematic review. British journal of management 14 (3), 207–222.

Tsao, W.-C., Hsieh, M.-T., Shih, L.-W., Lin, T.M., 2015. Compliance with ewom: The influence of hotel reviews on booking intention from the perspective of consumer conformity. International Journal of Hospitality Management 46, 99–111.

Tsou, M.-C., 2010. Geographic information retrieval and text mining on chinese tourism web pages. International Journal of Information Technology and Web Engineering (IJITWE) 5 (1), 56–75.

Tsujii, K., Fujita, Y., Tsuda, K., 2013. The extraction method of the service improvement information from guests' review. Procedia Computer Science 22, 1303–1310. https://doi.org/10.1016/j.procs.2013.09.218.

Tsujii, K., Takahashi, M., Tsuda, K., 2015a. Feature extraction from numerical evaluation in online hotel reviews. Procedia Computer Science 60, 1138–1145. https://doi.org/10.1016/j.procs.2015.08.164.

Tsujii, K., Tsuda, K., Takahashi, M., Jul. 2015b. Towards extracting the hotel evaluations from the comments by the foreign tourists with text mining. In: 2015 IIAI 4th International Congress on Advanced Applied Informatics. IEEE. doi: 10.1109/iiai-aai.2015.172.

Tussyadiah, I.P., Park, S., 2018. When guests trust hosts for their words: Host description and trust in sharing economy. Tourism Management 67, 261–272.

Uchiyama, Y., Tanaka, Y., Matsuoka, H., Kohsaka, R., 2017. Expectations of residents and tourists of agriculture-related certification systems: analysis of public perceptions. Journal of Ethnic Foods 4 (2), 110–117.

Uppink Calderwood, L., Soshkin, M., Jul. 2022. The travel and tourism competitiveness report 2019. [Online; accessed 1. Jul. 2022]. URL: https://www.weforum.org/reports/the-travel-tourism-competitiveness-report-2019.

Varga, B., Trambitas-Miron, A.D., Roth, A., Marginean, A., Slavescu, R.R., Groza, A., 2014. Lela-a natural language processing system for romanian tourism. In: 2014 Federated Conference on Computer Science and Information Systems. IEEE, pp. 281–288.

Wang, C., Guo, N., Lian, M., Xiao, X., 2020. Differences in regional media responses to china's holistic tourism: Big data analysis based on newspaper text. IEEE Access 8, 135050–135058.

Wang, H.C., Chiang, Y.H., Sun, Y.F., 2019. Use of multi-lexicons to analyse semantic features for summarization of touring reviews. The Electronic Library.

Wang, R., Hao, J.-X., 2018. Gender difference on destination image and travel options: An exploratory text-mining study. In: 2018 15th International Conference on Service Systems and Service Management (ICSSSM). IEEE. doi: 10.1109/icsssm.2018.8465084.

Wattanacharoensil, W., La-ornual, D., 2019. A systematic review of cognitive biases in tourist decisions. Tourism Management 75, 353–369.

Weed, M., 2006. Sports tourism research 2000–2004: A systematic review of knowledge and a meta-evaluation of methods. Journal of Sport & Tourism 11 (1), 5–30.

Windasari, I.P., Eridani, D., 2017. Sentiment analysis on travel destination in indonesia. In: 2017 4th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE). IEEE. doi: 10.1109/icitacee.2017.8257717.

Wirsch, A., 2014. Analysis of a top-down bottom-up data analysis framework and software architecture design (Ph.D. thesis). Massachusetts Institute of Technology. URL: https://web.mit.edu/smadnick/www/wp/2014-08.pdf.

Wong, C.U.I., Qi, S., 2017. Tracking the evolution of a destination's image by text-mining online reviews - the case of macau. Tourism Management Perspectives 23, 19–29. https://doi.org/10.1016/j.tmp.2017.03.009.

woo An, H., Moon, N., Oct. 2019. Design of recommendation system for tourist spot using sentiment analysis based on CNN-LSTM. Journal of Ambient Intelligence and Humanized Computing. doi: 10.1007/s12652-019-01521-w

Wu, X., Zhang, Y., 2016. Study on the rural tourism image of she ethnic group based on network text mining analysis. In: Proceedings of the 2016 International Conference on Economics and Management Innovations. Atlantis Press. doi: 10.2991/icemi-16.2016.40.

Xiang, Z., Schwartz, Z., Gerdes Jr, J.H., Uysal, M., 2015. What can big data and text analytics tell us about hotel guest experience and satisfaction? International Journal of Hospitality Management 44, 120–130.

Xiang, Z., Wang, D., O'Leary, J.T., Fesenmaier, D.R., 2015. Adapting to the internet: trends in travelers' use of the web for trip planning. Journal of travel research 54 (4), 511–527.

Xiao, Z., Li, X., Wang, L., Yang, Q., Du, J., Sangaiah, A.K., 2018. Using convolution control block for chinese sentiment analysis. Journal of Parallel and Distributed Computing 116, 18–26.

Xu, X., 2020. Examining an asymmetric effect between online customer reviews emphasis and overall satisfaction determinants. Journal of Business Research 106, 196–210. https://doi.org/10.1016/j.jbusres.2018.07.022.

Xu, X., Li, Y., 2016. The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: A text mining approach. International Journal of Hospitality Management 55, 57–69. https://doi.org/10.1016/j.ijhm.2016.03.003.

Xu, X., Liu, W., Gursoy, D., 2018. The impacts of service failure and recovery efforts on airline customers' emotions and satisfaction. Journal of Travel Research 58 (6), 1034–1051. https://doi.org/10.1177/0047287518789285.

Xu, X., Wang, X., Li, Y., Haghighi, M., 2017. Business intelligence in online customer textual reviews: Understanding consumer perceptions and influential factors. Int. J. Inf. Manage. 37 (6), 673–683. https://doi.org/10.1016/j.ijinfomgt.2017.06.004.

Yadav, M.L., Roychoudhury, B., 2019. Effect of trip mode on opinion about hotel aspects: A social media analysis approach. International Journal of Hospitality Management 80, 155–165. https://doi.org/10.1016/j.ijhm.2019.02.002.

Yadav, S.K., Bhushan, M., Gupta, S., 2015. Multimodal sentiment analysis: Sentiment analysis using audiovisual format. In: 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom). pp. 1415–1419.

Yan, Q., Zhou, S., Wu, S., 2018. The influences of tourists' emotions on the selection of electronic word of mouth platforms. Tourism Management 66, 348–363. https://doi.org/10.1016/j.tourman.2017.12.015.

Yang, H.-L., Chao, A.F., 2018. Sentiment annotations for reviews: an information quality perspective. Online Information Review 42 (5), 579–594. https://doi.org/10.1108/oir-04-2017-0114.

Yang, Y., Lin, X., 2016. Tourism network comments sentiment analysis and early warning system based on ontology. In: Intelligent Computing Theories and Application. Springer International Publishing, pp. 863–870. https://doi.org/10.1007/978-3-319-42294-7_76.

Yates, A., Goharian, N., Yee, W.G., 2013. Semi-supervised probabilistic sentiment analysis: Merging labeled sentences with unlabeled reviews to identify sentiment. Proceedings of the American Society for Information Science and Technology 50 (1), 1–10. doi: 10.1002/meet.14505001031.

Ye, C., Wu, J., Li, M., Tu, L., Shi, H., 2012. A quantitative analysis of tourist destination images based on ecspt model and sentiment analysis. In: 2012 IEEE Symposium on Robotics and Applications (ISRA). IEEE, pp. 115–117.

Yergesh, B., Bekmanova, G., Sharipbay, A., Oct. 2017. Sentiment analysis on the hotel reviews in the kazakh language. In: 2017 International Conference on Computer Science and Engineering (UBMK). IEEE. doi: 10.1109/ubmk.2017.8093531.

Ying, Z., 2018. Research on the promotion of word of mouth in tourist scenic spots based on web text mining——the case study of wanlu valley in guangdong province. MATEC Web of Conferences 173, 03060. https://doi.org/10.1051/matecconf/201817303060.

Young, L.M., Gavade, S.R., 2018. Translating emotional insights from hospitality employees' comments. International Hospitality Review 32 (1), 75–92. https://doi.org/10.1108/ihr-08-2018-0007.

Yu, C., Zhu, X., Feng, B., Cai, L., An, L., 2019. Sentiment analysis of japanese tourism online reviews. Journal of Data and Information Science 4 (1), 89–113. https://doi.org/10.2478/jdis-2019-0005.

Zadeh, L., 1975a. The concept of a linguistic variable and its application to approximate reasoning—i. Inf. Sci. 8 (3), 199–249. https://doi.org/10.1016/0020-0255(75)90036-5.

Zadeh, L., 1975b. The concept of a linguistic variable and its application to approximate reasoning—II. Inf. Sci. 8 (4), 301–357. https://doi.org/10.1016/0020-0255(75)90046-8.

Zahedpisheh, N., Bakar, A., Zulqarnain, B., Saffari, N., 2017. English for tourism and hospitality purposes (etp). English Language Teaching 10 (9), 86–94.

Zapata, G., Murga, J., Raymundo, C., Alvarez, J., Dominguez, F., 2017. Predictive model based on sentiment analysis for peruvian SMEs in the sustainable tourist sector. In: Proceedings of the 9th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management. SCITEPRESS - Science and Technology Publications. https://doi.org/10.5220/0006583302320240.

Zapata, G., Murga, J., Raymundo, C., Dominguez, F., Mogerza, J., 2018. Sentiment analysis based information architecture model for peruvian sustainable tourism smes. In: International Conference on Tourism Research. Academic Conferences International Limited, pp. 176–X.

Zapata, G., Murga, J., Raymundo, C., Dominguez, F., Moguerza, J.M., Alvarez, J.M., 2019. Business information architecture for successful project implementation based on sentiment analysis in the tourist sector. Journal of Intelligent Information Systems 53 (3), 563–585. https://doi.org/10.1007/s10844-019-00564-x.

Zeng, C., Nakatoh, T., Hirokawa, S., Eguchi, M., 2018. Text mining of tourism preference in a multilingual site. IEEJ Transactions on Electrical and Electronic Engineering 14 (4), 590–596. https://doi.org/10.1002/tee.22841.

Zhang, J., 2019. What's yours is mine: exploring customer voice on airbnb using text-mining approaches. Journal of Consumer Marketing.

Zhang, X., Yu, Q., 2017. Hotel reviews sentiment analysis based on word vector clustering. In: 2017 2nd IEEE International Conference on Computational Intelligence and Applications (ICCIA). IEEE, pp. 260–264.

Zhang, X., Yu, Y., Li, H., Lin, Z., 2016. Sentimental interplay between structured and unstructured user-generated contents. Online Information Review 40 (1), 119–145. https://doi.org/10.1108/oir-04-2015-0101.

Zhang, Z., Li, H., Meng, F., Li, Y., 2019. The effect of management response similarity on online hotel booking. Int. J. Contemporary Hospitality Manage. 31 (7), 2739–2758. https://doi.org/10.1108/ijchm-09-2018-0740.

Zhu, L., Li, W., Guo, K., Shi, Y., Zheng, Y., 2017. The tourism-specific sentiment vector construction based on kernel optimization function. Procedia Computer Science 122, 1162–1167. https://doi.org/10.1016/j.procs.2017.11.487.

Zvarevashe, K., Olugbara, O.O., 2018. A framework for sentiment analysis with opinion mining of hotel reviews. In: 2018 Conference on Information Communications Technology and Society (ICTAS). IEEE, pp. 1–4.