

# Topic Classification in the Tourism Field

Subjects: **Hospitality, Leisure, Sport & Tourism**

Contributor: Tai-Yi Yu , I-Cheng Chang , Jeou-Shyan Horng , Chih-Hsing Liu ,

In the past, the process of classifying, deconstructing, and analyzing relevant documents in tourism area required significant time and resources from domain experts. As this body of work continues to grow and diversify, text mining technology can better comprehend and promote the leisure industry, such that the general public becomes willing to understand, recognize, support, and participate in achieving the goals of sustainable development. Text mining can provide valuable knowledge from a large number of unstructured texts. Early text mining techniques were used in file classification.

cluster analysis

text mining

word cloud

co-word analysis

strategic diagram

## 1. Introduction

In recent years, smart technology, including artificial intelligence, big data, and the sharing economy, has become an important trend leading to the development of the global smart industry. In particular, big data and artificial intelligence have become dominant in various industries, especially knowledge-intensive ones such as tourism <sup>[1]</sup>. In the era of big data, firms use artificial intelligence to analyze the huge amounts of messy data they capture to identify useful knowledge that can help them innovate business models and value propositions. By utilizing big data analysis and artificial intelligence, the tourism and catering industry can provide real-time feedback, as well as improved transparency, market segmentation, decision-making, and product and service innovation, among other aspects <sup>[2]</sup>, and thereby increase the value of the industry.

Tourism is generally defined as persons traveling to and staying in places outside their usual environment, for not more than one consecutive year, for leisure, business, or other purposes <sup>[3]</sup>. The 2030 Agenda for Sustainable Development SDG target 8.9 states the following target to achieve by 2030: “devise and implement policies to promote sustainable tourism that creates jobs and promotes local culture and products” <sup>[4]</sup>. The connotation of sustainable tourism is frequently enriched: initial attention focused on environmental issues, while more definitions denote the importance of working towards the balanced development of economic, social, and environmental aspects. Most people believe that sustainable tourism emphasizes the connection between tourism activities and society with respect to the long-term coordinated development of the economy, resources, and the environment <sup>[5]</sup>. Goals are aimed at both economic development and a reduction in the negative impact of tourism activities, which includes continued development of the tourism industry while protecting natural and cultural resources. It is vital to coordinate and balance the relationships between different stakeholders in the process of tourism development <sup>[6]</sup>.

Information technology is part of the lifeblood of the tourism industry [7]. Combining knowledge gathered through statistics and domain experts from the tourism industry can help verify the results of visualization analysis. Information technology can apply an automatic topic classification to natural language processing documents to classify representative documents quickly and objectively; co-word analysis and association rule analysis can then be used to analyze the importance and relevance of specific words. There are four main research aims for this entry: (1) carry out the subject classification process of academic articles in the tourism field to assess the consistency and characteristics of the topic classification; (2) assess the characteristics of the subject classification and confirm its consistency; (3) use co-word analysis and strategic diagram to understand the importance and relevance of specific marketing strategy vocabulary; and (4) recognize the research tendencies of distinct topics in tourism field.

## 2. Topic Classification in the Tourism Industry

In the past, the process of classifying, deconstructing, and analyzing relevant documents in this area required significant time and resources from domain experts. As this body of work continues to grow and diversify, text mining technology can better comprehend and promote the leisure industry, such that the general public becomes willing to understand, recognize, support, and participate in achieving the goals of sustainable development. Text mining can provide valuable knowledge from a large number of unstructured texts. Early text mining techniques were used in file classification [8]. As the various types of text information keep increasing, including e-books, web pages, online news pages, blog articles, images, sounds, and videos, manual capture becomes impossible, and the need for topic models for automatic classification becomes apparent. Regarding the application of text mining in the tourism field, Okumus et al. [9] investigated the catering and tourism industry from 1976 to 2016: they analyzed the evolution of food and gastronomy research and identified emerging research topics, methods, and areas of national or interdisciplinary cooperation. Most of the 462 articles centered on gourmet, quantitative, and practical topics. Sainaghi et al. [10] used a cross-reference network analysis to evaluate the literature on hotel performance published between 1996 and 2015 to identify the most cross-cited papers, authors, and journals. Their sample analysis included 734 papers and demonstrated a spectacular growth of outputs, with the last time period (2011–2015) contributing 56% of output; in total, 1% of the sample accounted for 14% of the cross-references.

The topic model can use text mining algorithms, keyword libraries, and keyword occurrence ratios from a large amount of unstructured text data to define the subjective or objective category of the documents. Topic classification algorithms commonly used in text mining include cluster analysis, logistic regression, boost tree regression, hierarchical K-Means, K-means, latent Dirichlet allocation (LDA), and support vector machine (SVM) [11][12]. The hierarchical K-Means is very commonly applied in tourism management, mainly for classification issues, such as the attribute classification of tourists [13][14][15] and motivation classification of tourists [16][17]. With the hierarchical K-means cluster analysis, Suni and Komppula [16] used 30 motivational statements to classify respondents into one of five groups: controllers, indifferent, nostalgia, comfort seekers, and novelty seekers. Lee and Kim [14] divided the older adult volunteers in an international sporting event into two distinct segments of

serious leisure characteristics, while Michèle et al. [15] analyzed the activity profiles of social and leisure activities among older adults and divided the respondents into seven clusters. Finally, Jiao et al. [17] classified cruise ship tourists into four main categories: psychometric tourists, traditional tourists, pioneers, and sightseers.

With the help of the LDA model, Jia [18] randomly selected 100 yoga centers in Shanghai, identified 15 topics with the top 10 words from review comments. Vu et al. [19] utilized topic modeling to perform a travel itinerary analysis using the LDA model to clarify information on itineraries and tourist preferences. The optimal number of topics was decided as 24 using validation perplexity and computation times for different topic numbers with a total of 12,446 daily itineraries; each topic was visualized using the word cloud method and a heat map diagram. Shafqat and Byun [20] applied the LDA model to a travel blog database, extracted the top 150 blogs on tourism in Jeju, Korea, and identified the top 11 topics: location, timing, food, weather, entertainment, environment, accommodations, transportation, expense, services, and rental cars. Sutherland and Kiatkawsin [21] applied the LDA approach to identify 43 topics of interest that drive customer experience and satisfaction within a dataset of 1,086,800 Airbnb reviews; they grouped them into four topics: evaluation, location, unit, and management characteristics. The number of suitable classifications and sound interpretations of the topic categories are important for topic modeling.

Pleumarom [22] studied the tourism industry in the Mekong region and stated that the local government must adopt a cohesive management approach to achieve sustainable tourism in areas where multiculturalism and government institutions coexist. Sustainable tourism is an important topic in the tourism and hospitality industries; it can improve organizational performance, help gain competitive advantage, and be used as a commercial marketing topic. There are many research themes and influencing factors of sustainable tourism, such as sense of place, pro-environmental behaviors [23][24][25], and human health [26]. The interpretation ability of the local tourism industry could increase the income of sustainable tourism, and local interpreters could meet customer needs and create local employment, promote economic sustainability, and also act as on-site supervisors of visitors to influence their understanding of local perspectives, social protection, and environmental issues [27][28].

### 3. Marketing Strategy in the Tourism Industry

Marketing strategy plays a very important role in tourism. The traditional 4P marketing strategy proposed by McCarthy [29], focused on promotion, place, product, and price, has been widely applied in various business fields. Booms and Bitner [30] revised McCarthy's [29] 4P to 7P, adding people, physical evidence, and process. Kotler [31] revised McCarthy's [29] 4P to 6P, adding politics and public opinion to provide marketing strategies for a complex and diverse society. Pomeroy, Noble, and Johnson [32] applied 10 marketing foundations to the marketing model of sustainable tourism: promotion, place, product, price, physical evidence, process, packaging, participation, programming, and partnership while considering economic, environmental, and social issues. Dudensing et al. [33] indicated that the different marketing strategies of stakeholders in the tourism industry led to heterogeneity or conflict in terms of stakeholder expectations. Wray [34] emphasized the interactive nature of sustainable tourism—ensuring that profits remain with local operators and sites. Lozano-Oyola et al. [35] also supported the aforementioned view, advocating that those stakeholders review economic indicators before making sustainable tourism decisions. Different expectations of various stakeholder groups may cause conflicts: destination marketing

must take into account the views of numerous stakeholders [36]. Generally, marketing and sustainability can work together through the development of a sustainable tourism marketing model, such as managing the travel route of the tourism industry through an ecological footprint.

Big data plays a catalytic role in the process of determining consumer preferences. By obtaining correct data, meaningful analysis can take place, leading to structural changes in consumer behavior models and marketing strategies [37]. Through appropriately identified disseminations, Samara et al. [38] noted that the benefits of adopting big data and artificial intelligence strategies include increased efficiency, productivity, and profitability for tourism suppliers, combined with an extremely rich and personalized experience for travelers. Katsikari et al. [39] proposed that the rapid expansion of the Internet and social media provides marketers with simple and cost-effective ways and opportunities to reach potential tourists; their study investigated which destination elements are deemed attractive by tourists who use social media.

Through the text mining results of a large number of academic articles, important words related to tourism marketing can be obtained and, at the same time, their relevance and relative importance can be understood. Based on the existing information technology, the tourism industry can expand consumerism-based IT and tourism research in order to participate in a wider dialogue; emphasize the use of technology to achieve a better quality of life, economic prosperity, social well-being, and sustainability; and use open data and shared social knowledge as a basis for tourism experience and innovation [40].

## 4. Co-Word Analysis and Strategic Diagram

Co-word Analysis is a quantitative technique for scanning document content to denote when and where a defined specific word co-occurs. Through co-word analysis, various co-occurrence relationships in a specific object can be expressed, such as co-quotation, co-author, and co-word characteristics [41]. A strategic diagram is developed from the analysis, which assists in identifying evolutionary trends and relationships between thematic groups [42].

In the application of co-word analysis, Guo et al. [43] surveyed 1138 articles and reviews from 1980 to 2016 and used 52 high-frequency keywords related to company restrictions to investigate the current situation and trends of company restrictions. The central terms were “restrictions”, “learning”, “institutions”, and “behavior”; the results show that “restrictions” had the highest degree of importance. The aforementioned 52 high-frequency keywords could be divided into six categories, and the indicators of company development (such as innovation, supply chain, decision-making, performance, sustainability, and employee behavior) were significantly related to company restrictions. Khasseh et al. [44] used co-word analysis to describe the topic characteristics within two journals, *Scientometrics* and *Journal of Informetrics*, from 1978 to 2014; they then divided them into 11 representative topics using hierarchical cluster analysis and utilized a strategy diagram to illustrate the structure, maturity, and cohesion of each topic. Corrales-Garay et al. [45] applied co-word analysis to create a map of the main themes identified in the knowledge areas and determined their importance and relevance.

Leung et al. [46] sampled 406 publications related to social media from 2007 to 2016 across 16 business and hospitality/tourism journals and applied co-word analysis to identify the evolution of research themes over time. Shen et al. [47] collected 29 years of online database data of academic journals and then used co-word analysis and bibliographic analysis techniques to analyze trends, core authors, degrees of cooperation, core journal analysis, and distribution of publishing institutions. This allowed them to establish 10 important and unevenly distributed research trends and then propose a new potential research theme of information search and information security. De la Hoz-Correa et al. [48] utilized co-word analysis to denote six clusters of themes in published research listed in the Web of Science (WoS) and Scopus database; this type of analysis offers powerful insights into the conceptual structure of medical tourism research. The co-word analysis is an effective manner to identify the content, importance, and relevance of different themes from the aforementioned references.

Rodríguez-López et al. [49] applied a strategic diagram to present the importance of topic themes from a bibliometric analysis of published academic research dealing with restaurants in the fields of hospitality, leisure, sport, and tourism ring the period from 2000 to 2018. Muñoz-Leiva et al. [50] conducted data mining on 759 papers related to blockchain technology in the financial field by employing co-word analysis and strategic diagrams to explore hot topics and predict future development trends. Rodríguez-López et al. [51] selected documents whose titles included specific terms from two online databases, Web of Science (WoS) and Scopus, and utilized a keyword strategic diagram to determine the importance of keywords and their levels of development. Terán-Yépez et al. [52] sampled 216 articles from sustainable entrepreneurship and identified the most significant research tendencies, enabling the proposal of several future research directions through graphic mapping of strategic diagrams. Finally, Jiménez-García et al. [53] applied bibliometric techniques to investigate research trends in 214 articles related to sports tourism and sustainability and used strategic diagrams to identify the most significant research tendencies across distinct topics.

## References

1. Buhalis, D. Technology in tourism-from information communication technologies to eTourism and smart tourism towards ambient intelligence tourism: A perspective article. *Tour. Rev.* 2019, 75, 267–272.
2. Brown, B.; Chui, M.; Manyika, J. Are you ready for the era of 'big data'. *McKinsey Q.* 2011, 4, 24–35.
3. Björk, P. Definition paradoxes: From concept to definition. In *Critical Issues in Ecotourism: Understanding a Complex Tourism Phenomenon*; Butterworth-Heinemann: Oxford, UK, 2007; pp. 23–45.
4. United Nations. General Assembly Resolution A/RES/70/1 Transforming Our World the 2030 Agenda for Sustainable Development. 2015. Available online: <https://sdgsunorg/2030agenda> (accessed on 28 August 2021).

5. Torres-Delgado, A.; Palomeque, F.L. The growth and spread of the concept of sustainable tourism: The contribution of institutional initiatives to tourism policy. *Tour. Manag. Perspect.* 2012, 4, 1–10.
6. Herrera, M.R.G.; Sasidharan, V.; Hernández, J.A.Á.; Herrera, L.D.A. Quality and sustainability of tourism development in Copper Canyon Mexico: Perceptions of community stakeholders and visitors. *Tour. Manag. Perspect.* 2018, 27, 91–103.
7. Sigala, M. New technologies in tourism: From multi-disciplinary to anti-disciplinary advances and trajectories. *Tour. Manag. Perspect.* 2018, 25, 151–155.
8. Miner, G.; Elder, I.V.J.; Fast, A.; Hill, T.; Nisbet, R.; Delen, D. *Practical Text Mining and Statistical Analysis for Non-Structured Text Data Applications*; Academic Press: Cambridge, MA, USA, 2012.
9. Okumus, B.; Koseoglu, M.A.; Ma, F. Food and gastronomy research in tourism and hospitality: A bibliometric analysis. *Int. J. Hosp. Manag.* 2018, 73, 64–74.
10. Sainaghi, R.; Phillips, P.; Baggio, R.; Mauri, A. Cross-citation and authorship analysis of hotel performance studies. *Int. J. Hosp. Manag.* 2018, 73, 75–84.
11. Garcia, S.; Derrac, J.; Cano, J.; Herrera, F. Prototype selection for nearest neighbor classification: Taxonomy and empirical study. *IEEE Trans. Pattern. Anal. Mach. Intell.* 2012, 34, 417–435.
12. Sohrabi, B.; Vanani, I.R.; Shineh, M.B. Topic modeling and classification of cyberspace papers using text mining. *J. Cyber. Stud.* 2018, 2, 103–125.
13. Veisten, K.; Haukeland, J.V.; Baardsen, S.; Degnes-Ødemark, H.; Grue, B. Tourist segments for new facilities in national park areas: Profiling tourists in Norway based on psychographics and demographics. *J. Hosp. Mark. Manag.* 2015, 24, 486–510.
14. Lee, Y.; Kim, M. Serious leisure characteristics of older adult volunteers: The case of an international sporting event. *World Leis. J.* 2018, 60, 45–57.
15. Michèle, J.; Guillaume, M.; Alain, T.; Nathalie, B.; Claude, F.; Kamel, G. Social and leisure activity profiles and well-being among the older adults: A longitudinal study. *Aging Ment. Health* 2019, 23, 77–83.
16. Suni, J.; Komppula, R. SF-Film village as a movie tourism destination—a case study of movie tourist push motivations. *J. Travel. Tour. Mark.* 2012, 29, 460–471.
17. Jiao, Y.; Hou, Y.; Lau, Y.Y. Segmenting Cruise Consumers by Motivation for an Emerging Market: A Case of China. *Front. Psychol.* 2021, 12, 634.
18. Jia, S.S. Leisure motivation and satisfaction: A text mining of yoga centres yoga consumers and their interactions. *Sustainability* 2018, 10, 4458.

19. Vu, H.Q.; Li, G.; Law, R. Discovering implicit activity preferences in travel itineraries by topic modeling. *Tour. Manag.* 2019, 75, 435–446.
20. Shafqat, W.; Byun, Y.C. A recommendation mechanism for under-emphasized tourist spots using topic modeling and sentiment analysis. *Sustainability* 2020, 12, 320.
21. Sutherland, I.; Kiatkawsin, K. Determinants of guest experience in Airbnb: A topic modeling approach using LDA. *Sustainability* 2020, 12, 3402.
22. Pleumarom, A. How sustainable is Mekong tourism? In *Sustainable Tourism: A Global Perspective*; Routledge: Oxfordshire, UK, 2012; pp. 140–166.
23. Vaske, J.J.; Kobrin, K.C. Place attachment and environmentally responsible behavior. *J. Environ. Educ.* 2001, 32, 16–21.
24. Stedman, R.C. Toward a social psychology of place: Predicting behavior from place-based cognitions attitude and identity. *Environ. Behav.* 2002, 34, 561–581.
25. Kudryavtsev, A.; Stedman, R.C.; Krasny, M.E. Sense of place in environmental education. *Environ. Educ. Res.* 2012, 18, 229–250.
26. Sampson, R.; Gifford, S.M. Place-making settlement and well-being: The therapeutic landscapes of recently arrived youth with refugee backgrounds. *Health Place* 2010, 16, 116–131.
27. Dodds, R.; Graci, S. *Sustainable Tourism in Island Destinations*; Routledge: Oxfordshire, UK, 2012.
28. Ham, S.H.; Weiler, B. Interpretation as the centerpiece of sustainable wildlife tourism. In *Sustainable Tourism*; Butterworth-Heinemann Oxford: Oxford, UK, 2012; pp. 35–44.
29. McCarthy, E.J.; Shapiro, S.J.; Perreault, W.D. *Basic Marketing*; Irwin-Dorsey: Toronto, ON, Canada, 1979; pp. 29–33.
30. Booms, B.H.; Bitner, M.J. Marketing strategies and organization structures for service firms. In *Marketing of Services*; Donnelly, J.H., George, W.R., Eds.; American Marketing Association: Chicago, IL, USA, 1981; pp. 47–51.
31. Kotler, P. *Marketing Management: Analysis Planning, Implementation and Control*; Prentice-Hall: Hoboken, NJ, USA, 1999.
32. Pomering, A.; Noble, G.; Johnson, L.W. Conceptualising a contemporary marketing mix for sustainable tourism. *J. Sustain. Tour.* 2011, 19, 953–969.
33. Dudensing, R.M.; Hughes, D.W.; Shields, M. Perceptions of tourism promotion and business challenges: A survey-based comparison of tourism businesses and promotion organizations. *Tour. Manag.* 2011, 32, 1453–1462.

34. Wray, M. Adopting and implementing a transactive approach to sustainable tourism planning: Translating theory into practice. *J. Sustain. Tour.* 2011, 19, 605–627.
35. Lozano-Oyola, M.; Blancas, F.J.; González, M.; Caballero, R. Sustainable tourism indicators as planning tools in cultural destinations. *Ecol. Indic.* 2012, 18, 659–675.
36. Getz, D.; Timur, S. Stakeholder involvement in sustainable tourism: Balancing the voices. In *Global Tourism*; Routledge: Oxfordshire, UK, 2012; pp. 247–264.
37. Inanc–Demir, M.; Kozak, M. Big data and its supporting elements: Implications for tourism and hospitality marketing. In *Big Data and Innovation in Tourism, Travel and Hospitality*; Springer: Singapore, 2019; pp. 213–223.
38. Samara, D.; Magnisalis, I.; Peristeras, V. Artificial intelligence and big data in tourism: A systematic literature review. *J. Hosp. Tour. Technol.* 2020, 11, 343–367.
39. Katsikari, C.; Hatzithomas, L.; Fotiadis, T.; Folinas, D. Push and Pull Travel Motivation: Segmentation of the Greek Market for Social Media Marketing in Tourism. *Sustainability* 2020, 12, 4770.
40. Xiang, Z. From digitization to the age of acceleration: On information technology and tourism. *Tour. Manag. Perspect.* 2018, 25, 147–150.
41. Ding, Y.; Chowdhury, G.G.; Foo, S. Bibliometric cartography of information retrieval research by using co-word analysis. *Inf. Process. Manag.* 2001, 37, 817–842.
42. Yang, Y.; Wu, M.; Cui, L. Integration of three visualization methods based on co-word analysis. *Scientometrics* 2012, 90, 659–673.
43. Guo, D.; Chen, H.; Long, R.; Lu, H.; Long, Q. A co-word analysis of organizational constraints for maintaining sustainability. *Sustainability* 2017, 9, 1928.
44. Khasseh, A.A.; Soheili, F.; Moghaddam, H.S.; Chelak, A.M. Intellectual structure of knowledge in iMetrics: A co-word analysis. *Inf. Process. Manag.* 2017, 53, 705–720.
45. Corrales-Garay, D.; Ortiz-de-Urbina-Criado, M.; Mora-Valentín, E.M. Knowledge areasthemes and future research on open data: A co-word analysis. *Gov. Inf. Q.* 2019, 36, 77–87.
46. Leung, X.Y.; Sun, J.; Bai, B. Bibliometrics of social media research: A co-citation and co-word analysis. *Int. J. Hosp. Manag.* 2017, 66, 35–45.
47. Shen, L.; Xiong, B.; Hu, J. Research status hot spots and trends for information behavior in China using bibliometric and co-word analysis. *J. Doc.* 2017, 73, 618–633.
48. De la Hoz-Correa, A.; Muñoz-Leiva, F.; Bakucz, M. Past themes and future trends in medical tourism research: A co-word analysis. *Tour. Manag.* 2018, 65, 200–211.



49. Rodríguez-López, M.E.; Alcántara-Pilar, J.M.; del Barrio-García, S.; Muñoz-Leiva, F. A review of restaurant research in the last two decades: A bibliometric analysis. *Int. J. Hosp. Manag.* 2020, 87, 102387.
50. Muñoz-Leiva, F.; Porcu, L.; Barrio-García, S.D. Discovering prominent themes in integrated marketing communication research from 1991 to 2012: A co-word analytic approach. *Int. J. Advert.* 2015, 34, 678–701.
51. Rodríguez-López, N.; Diéguez-Castrillón, M.I.; Gueimonde-Canto, A. Sustainability and tourism competitiveness in protected areas: State of art and future lines of research. *Sustainability* 2019, 11, 6296.
52. Terán-Yépez, E.; Marín-Carrillo, G.M.; del Pilar Casado-Belmonte, M.; de las Mercedes Capobianco-Uriarte, M. Sustainable entrepreneurship: Review of its evolution and new trends. *J. Clean. Prod.* 2020, 252, 119742.
53. Jiménez-García, M.; Ruiz-Chico, J.; Peña-Sánchez, A.R.; López-Sánchez, J.A. A bibliometric analysis of sports tourism and sustainability (2002–2019). *Sustainability* 2020, 12, 2840.

---

Retrieved from <https://www.encyclopedia.pub/entry/history/show/54184>