

# Applying Creative Computing in the Analysis of Smart Tourism Research in China

Gongpeng Zhang<sup>a,b</sup>, Yuncai Ning<sup>a</sup>, Chi Zhang<sup>b,\*</sup>, Zhaogang Wang<sup>a</sup>, and Beu Lee<sup>c</sup>

<sup>a</sup>*School of Management, China University of Mining and Technology, Beijing, 100083, China*

<sup>b</sup>*Collaborative Innovation Center of eTourism, Beijing Union University, Beijing, 100101, China*

<sup>c</sup>*College of Business, Texas A&M University, San Antonio, Texas, 78224, USA*

---

## Abstract

Understanding the contents of smart tourism comprehensively can provide abundant guidance and reference for the practices of smart tourism. It also helps to identify and overcome the shortcomings of smart tourism research. Based on the core ideas of creative computing, this paper constructs a word frequency matrix and performs cluster analysis by using keywords from smart tourism research. Then, the relevant keywords are extracted by using chi-squared statistics to interpret the cluster results. The results show that many creative articles focus on tourism information, the status quo and countermeasures of smart tourism, tourism information technology, and smart tourism scenic area management, which are the hot topics of domestic smart tourism research. Furthermore, tourism big data, tourism service supply chain, Internet + tourism, and tourism experience have recently attracted increasing attention. The impact of tourism industry policy and the application of new information technology have always been an important source of creativity in smart tourism research.

*Keywords:* smart tourism; creative computing; cluster; chi-squared statistics

(Submitted on November 12, 2018; Revised on December 13, 2018; Accepted on January 17, 2019)

© 2019 Totem Publisher, Inc. All rights reserved.

---

## 1. Foreword

Understanding the contents of smart tourism comprehensively can not only provide abundant guidance and reference for the practices of smart tourism, but also help to identify and overcome the shortcomings of smart tourism research. Smart tourism, as a new form of tourism, is applied in tourism experience, industry development, administration, and other aspects by Internet of things, cloud computing, smart data mining, and other technologies. The development process of smart tourism matches the core ideas of creative computing indicating the combination of knowledge from various disciplines and the integration between creativity and technology, which can provide a new way of thinking for computer-aided innovation [1]. It is believed that creative computing may be a good way of grasping the context, hotspots, and trends of the development of smart tourism by analyzing domestic research on smart tourism.

## 2. Literature Summary

By extracting data from smart tourism literature, it is found that most of the smart tourism research is descriptive and lacks in-depth exploration [2]. Seven papers and 69 reports on smart tourism at [www.cnki.net](http://www.cnki.net) concluded that smart tourism research remains at a stage where practice facilitates theoretical development [3]. Smart tourism research currently remains immature, lacking both depth and width [4]. Most academic smart tourism research remains at the level of concepts [5-6], connotations, frameworks, values, applications and development trends related to smart tourism [7], as well as the relation between smart tourism and tourism informatization [8-9]. However, a problem of the existing reviews of the smart tourism literature is the lack of adequate references and analysis of all data on smart tourism literature. As a result, the comprehensiveness of the research conclusions cannot be guaranteed. Most of the research adopts qualitative methods. Quantitative research methods have not been paid much attention in smart tourism literature review.

\* Corresponding author.

E-mail address: [tibuuzc@163.com](mailto:tibuuzc@163.com)

The quantitative research methods used to identify the main research contents in a field are mostly interdisciplinary knowledge combination, which is a typical case of creative computing. By using the literature research method for knowledge discovery, Gao et al. [10] designed a knowledge discovery model based on document data visualization and conducted cluster analysis and a semantic network with a co-occurrence matrix of high-frequency keywords. Aiming at the research status of Scientometrics in the context of visualization, starting from the spatial and temporal dimensions, Chen et al. [11] summarized the main contents of current research, using word frequency analysis and co-word analysis to identify, for example, subject topics and research fronts. Using the analysis integration software Ucinet [12], VOSViewer [13], CiteSpace [14], and other tools, some scholars conducted clustered analyses or drew visual network diagrams after constructing high-frequency keyword co-occurrence matrices [15]. All existing quantitative analyses on research hotspots screen high-frequency keywords, construct a co-occurrence matrix, and then perform cluster analysis or draw a knowledge map to identify hotspots in relevant fields based on keyword analysis. To an extent, such practices may lose important information implied in low-frequency keywords and result in one-sided research conclusions. Furthermore, the cluster results based on the co-occurrence matrix are subjective because they require manual identification of different categories of keywords for interpretation, which involves an enormous workload in situations involving large amounts of data and high-frequency keywords.

Covering all the journal papers dealing with smart tourism research on [www.cnki.net](http://www.cnki.net), this paper analyzes the cluster based on the word frequency matrix of the keywords and conducts an in-depth discovery on the cluster results by selecting highly relevant keywords based on the chi-squared statistics. In this way, the paper presents the characteristics of clustering results comprehensively and objectively and identifies the main research ideas of each category.

### 3. Data Collection and Pre-Processing

Keywords are important attributes of academic research that can reflect the essence of papers and play a critical role in research contents. Keywords therefore provide an important basis for research on the most popular research topics within an academic field.

The CNKI database is used as the data source of this paper. “Smart Tourism” is used as the keyword to search the CNKI journal database. The selected time point is before 2018. A total of 1219 related articles were retrieved, and the keywords of these articles constitute the data used in this study.

The data set is imported into Excel to construct a word frequency matrix for the keywords, and this matrix serves as the data source for the cluster analysis. The partial word frequency matrix is shown in Table 1 below.

Table 1. Word frequency matrix (partial)

	Smart tourism	Tourism management	Tourism information	...
Paper 1	1	1	0	...
Paper 2	1	0	1	...
Paper 3	1	0	0	...
...	...	...	...	...

As shown in Table 1, each line in the word frequency matrix represents a paper, while each column represents a keyword, where 1 indicates the keyword is among those of the corresponding paper and 0 indicates it is not among the keywords of the paper.

### 4. Cluster Analysis

The keyword word frequency matrix is imported into SPSS software for cluster analysis, and the intra-group connection method with better cluster effect is used for system cluster analysis. Because the word frequency matrix built in this paper is a 0-1 binary matrix, the simple matching coefficient is adopted to measure the distance or similarity between the papers. As for the keyword word frequency matrix, the simple matching coefficient is the frequency at which all keywords share the same values for the two papers.

The simple matching coefficient  $S$  may be expressed as

$$S = \frac{a + d}{a + b + c + d} \quad (1)$$

As shown in Table 2, “ $a$ ” represents the number of keywords that are included in both papers,  $d$  represents the number of keywords included in neither paper, and  $b$  and  $c$  represent the numbers of keywords included only in each one of the two papers, respectively. “ $a + b + c + d$ ” represents the total number of keywords, i.e., the number of variables in the word frequency matrix.

Table 2. Relevant parameters of simple matching coefficients

	Valve	Paper 1	
		1	0
	Paper 2		
	1	$a$	$b$
	0	$c$	$d$

The pedigree of the system cluster is shown in Figure 1. Based on the research purpose and classification effect, the current smart tourism research can be assigned among 18 categories according to the distance standard shown by the red line. The cluster results are shown in Table 3.

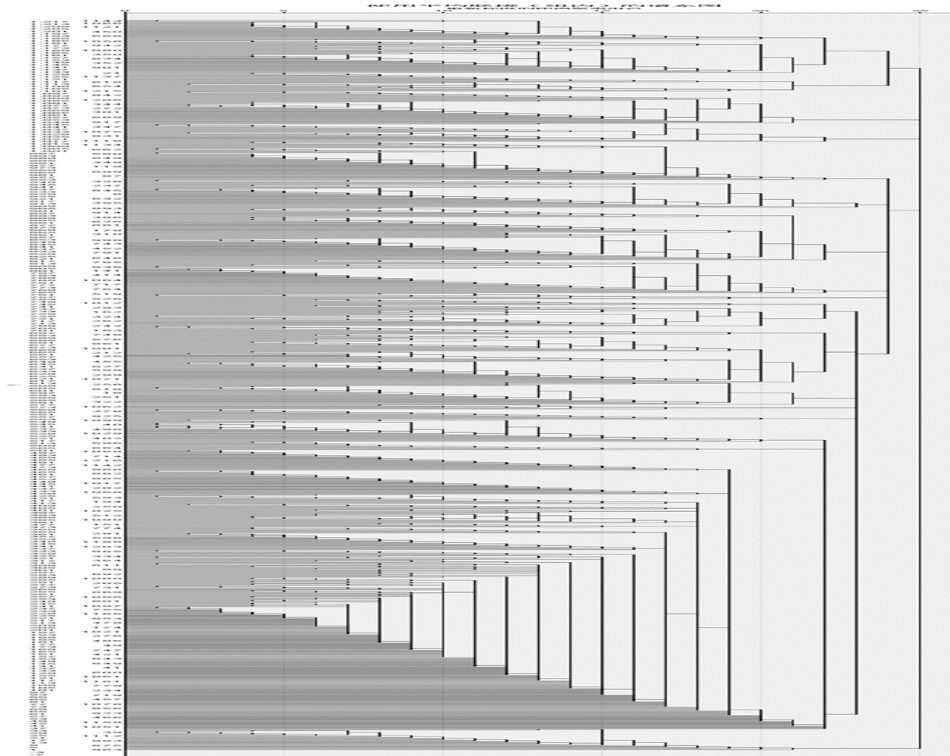


Figure 1. Cluster analysis in smart tourism literature during 2011-2017

Table 3. Results of clustering

Category	Number of papers	Category	Number of papers	Category	Number of papers
1	31	7	35	13	88
2	94	8	10	14	8
3	64	9	37	15	17
4	516	10	31	16	53
5	45	11	55	17	1
6	45	12	88	18	1

Table 3 shows a gap in the distribution of papers among the categories. Category 10 is the largest category, containing a total of 516 papers, and is followed by Categories 2, 12, and 13, each with 80-100 papers. The smallest categories are Categories 17 and 18, each of which contains just one paper, and because both these papers involve technological information they are ignored here. Each of the remaining categories contains only a small number of papers.

## 5. Interpretation of Clustering Results based on Creative Computing

The large number of papers in each category and the resultant large number of keywords create certain difficulties in the description of the cluster results. In this paper, the relevance between keywords and categories is calculated using chi-

squared statistics, and the cluster results are analyzed and described by selecting keywords with high relevance. The calculation formula is as follows:

$$\chi^2(t, c_i) = \frac{n \times (a \times d - c \times b)^2}{(a + c)(b + d)(a + b)(c + d)} \quad (2)$$

In the above formula, “ $n$ ” represents the total number of papers sampled from the relevant literature. “ $a$ ” represents the number of papers that belong to Category “ $c_i$ ” and include keyword “ $t$ ”, “ $b$ ” represents the number that do not belong to Category “ $c_i$ ” but include keyword “ $t$ ”, “ $c$ ” represents the number that belong to Category “ $c_i$ ” but exclude keyword “ $t$ ”, and “ $d$ ” represents the number of papers that neither belong to Category “ $c_i$ ” nor include the keyword “ $t$ ”.

The chi-squared value CHI between the keyword and the category is obtained based on chi-squared statistics, and a relevance ranking of the keywords and categories can be obtained by sorting keywords and the CHI value of each category in descending order. The higher the relevance, the stronger the ability to distinguish between Category “ $c_i$ ” and other categories. This paper also describes the ten most relevant characteristic words for each Category  $c_i$ . The extracted high-relevance keywords are shown in Table 4.

Table 4. Highly relevant keywords

Category	Highly relevant keywords	Number of papers
1	Tourism management, tourism management, management, training mode, talent cultivation mode, talent, talent cultivation, vocational tourism management, destination tourism management, applied talent	31
2	Development, countermeasures, development strategy, development model, tourism development, status quo, development strategy, tourism development, development status, tourism development model	94
3	Data, big data, databases, data mining, big data era, tourism-related big data, data centers, big data industry, big data platform, cloud data center	64
4	Information, scenic area, development, city, information, smart city, management, data, tourism information, service	516
5	Tourism services, tourism, service, service supply chain, supply chain, tour guide, tourism service supply chain, tourism public service, collaborative development, guide service	45
6	Internet, Internet+, mobile Internet, mobile Internet, Internet+, maker, APP, industry integration, Internet thinking, Internet+ tourism	45
7	Urban construction, smart city construction, construction, smart city, city, Wan Ning city, city informatization, Wan Ning, pilot demonstration, safe city	35
8	Public service system, service system, tourism public service system, public service, tourism public service, safety critical system, system, supply mode, tea culture background, construction mechanism	10
9	Experience, travel experience, tourist experience, user experience, Mount Lushan Scenic Area, tourist, park, guide, APP, route, smart tourism	37
10	Information service, tourism information service, public information, information, public information service, tourism information, service, tourism public information, visual threshold, tourism public information service	31
11	City, smart city, tourist city, smart tourism city, city tourism, regional coordinating, city agglomerations, urban agglomeration in southern Sichuan, Xianyang, integration	55
12	Information, informatization, tourism informatization, tourism information, information systems, information technology, geographic information systems, new generation of information technology, intelligence, technology	88
13	Scenic area, smart scenic area, scenic area management, tourist, tourist area, management, zoo, management, West Lake scenic area, relic	88
14	Public service platform, service platform, smart tourism public service platform, smart tourism public service, public service, TOP50, tourism vacation, Xiamen, talent training, performance evaluation	8
15	Enterprises, tourism enterprises, enterprise management, tourism enterprise management, business transformation, iSoftStone, Xinhua net, tourism enterprise intelligence, tourism value chain, package	17
16	Tourism industry, industry, tourism, tourism industry chain, industrial chain, tourists, international tourism area, modern tertiary industry, thinking form, Chi Zhou	53

Table 4 shows a clear difference between categories in terms of highly relevant keywords and degrees of similarity among keywords in the same category, which can better show the main creative content of each category.

The high relevance of the keywords in Category 1 indicates that the category includes research on smart tourism talent cultivation, which involves, for example, specialized settings and training patterns. The research papers in Category 2 focus on the current state of tourism and propose corresponding development strategies, models, and strategic responses, and so this category can be defined as the status quo and strategic responses category. The high relevance keywords in Category 3 indicate that this category deals with smart tourism big data research, covering, for example, databases, data mining, data

centers, and big data platform. Category 4 contains the largest number of papers, and high relevance keywords in this category involve information, informatization, smart city, data, and tourism information, indicating that this is the tourism informatization category. Category 5 focuses on tourism service, service supply chain, and tourism public service, and it can be described as the tourism service supply chain category. The high relevance keywords in Category 6 fully reflect the Internet, mobile Internet, and similar contents, so this category can be defined as the Internet + tourism research category. Category 7 includes keywords such as city construction, smart city, and pilot demonstration, and hence it can be considered the smart tourism city construction research category. Category 8 contains keywords related to tourism public service system, so it represents the smart tourism public service system category. The high relevance keywords in Category 9 include tourism experience, tourist experience, and user experience, making this the tourism experience research category. Category 10 comprises the smart tourism public information service research category. The high relevance keywords for Category 11 include smart tourism city, regional coordinating, urban agglomeration, and integration, so this is the smart tourism city cluster research category. The high relevance keywords in Category 12 include terms such as tourism informatization, information system, information technology, and wisdom, indicating that this is the tourism information technology research category. The high relevance keywords for Category 13 include terms such as smart scenic area, scenic area management, and tourist, indicating that this is the smart tourism scenic area management research category. Category 14 can be defined as the smart tourism public service platform research category. Category 15 contains keywords such as tourism enterprises, enterprise management, business transformation, and wisdom of tourism enterprises, indicating that this is the tourism enterprise wisdom category. Finally, the high relevance keywords for Category 16 include tourism industry, industrial chain, modern tertiary industry, and thinking form, making this the smart tourism industrial chain research category. For definitions of each category, please see Table 5 below.

Table 5. Definitions of relevant smart tourism researches

Category	Definition of creative content	Category	Definition of creative content
1	Talent cultivation	9	Tourism experience
2	Status quo and countermeasures	10	Smart tourism public information service
3	Tourism-related big data	11	Smart tourism city cluster
4	Tourism informatization	12	Tourism information technology
5	Tourism service supply chain	13	Smart tourism scenic area management
6	Internet+ tourism	14	Smart tourism public service platform
7	Smart tourism city construction	15	Wisdom of tourism enterprises
8	Smart tourism public service system	16	Smart tourism industrial chain

## 6. Analysis of Smart Tourism Research based on Creative Computing

Changes in the numbers of papers published in each creative content category over time are shown in Table 6.

Table 6. Volume of smart tourism research by year and category

Creative content category	Number of papers							
	2011	2012	2013	2014	2015	2016	2017	Total
Talent cultivation	0	0	5	3	7	9	7	31
Status quo and countermeasures	1	1	7	9	25	26	25	94
Tourism-related big data	0	1	2	10	17	16	18	64
Tourism informatization	5	18	52	88	132	110	111	516
Tourism service supply chain	0	0	2	8	8	10	17	45
Internet+ tourism	0	0	0	3	8	19	15	45
Smart tourism city construction	0	1	4	11	12	4	3	35
Smart tourism public service system	0	1	1	1	1	3	3	10
Tourism experience	0	0	0	6	11	11	9	37
Smart tourism public information service	1	3	1	5	9	2	10	31
Smart tourism city cluster	0	1	9	13	18	10	4	55
Tourism information technology	0	7	13	23	8	21	16	88
Smart tourism scenic area management	1	3	3	20	25	19	17	88
Smart tourism public service platform	0	0	0	4	1	0	3	8
Wisdom of tourism enterprises	0	0	1	4	5	5	2	17
Smart tourism industrial chain	0	5	4	12	14	13	5	53
Total	8	41	104	220	301	278	265	1217

### 6.1. Analysis of Popular Topics in Smart Tourism Research

In terms of the literature amount, research on tourism informatization ranks first. There are 516 references about tourism informatization, accounting for 42.4% of the total. Tourism informatization research grew rapidly before 2015. Although there was a downward trend in 2016 and 2017, the number is still large, and it is a hotspot in smart tourism research.

There are also many studies on the status quo and countermeasures of smart tourism, tourism informatization technology, and smart tourism scenic area management. There are over 88 articles about each issue. The number of articles studying the status quo and countermeasures has shown an upward trend in recent years while research on the smart tourism scenic area management has shown a downward trend; research on tourism informatization technology has shown volatile fluctuations and dramatic changes over time.

The total amount of research on tourism-related big data, tourism service supply chain, Internet + tourism, and tourism experience is not large. These issues were rarely studied before 2013. However, it has shown a clear upward trend since 2014, which is very worthy of attention.

There are few studies on talent cultivation, smart tourism city construction, smart tourism public information service, smart tourism city cluster, and smart tourism industrial chain. Among them, studies on smart tourism public information service are quite unstable. In addition, studies on smart tourism city construction, smart tourism city cluster, and smart tourism industrial chain all show a downward trend.

Studies on smart tourism public service system, smart tourism public service platform, and smart tourism enterprises are the most limited.

### 6.2. Analysis of Smart Tourism Research Trends

According to the literature time, articles related to smart tourism began to appear after 2011. In 2014, there were over 200 journal articles about smart tourism research, and it reached its peak in 2015 and developed stably in 2016 and 2017. The reason is that research on smart tourism is obviously influenced by industry policies: in 2010, China proposed the construction of smart cities, and smart tourism began to appear as a part of the smart city. The China National Tourism Administration proposed that the year 2014 should be regarded as the "Smart Tourism Year". The research results in 2014 and 2015 have reached a new level. In 2018, the Ministry of Culture and Tourism of the People's Republic of China was established. Research on culture and tourism is bound to be popular in the future.

As for the literature content, smart tourism research highlights multi-disciplinary integration and integrates the development of tourism industry creativity into the process of information technology development. Combining the tourism big data developed in 2014, the Internet + tourism proposed in 2015, the subsequent tourism service supply chain, and the tourism experience and tourism informatization technology, there were a total of 795 articles on tourism information research accounting for more than 65% of the total. The application of the emerging informatization technology in tourism will be an important creative content of smart tourism research.

## 7. Conclusions

By selecting journal articles about smart tourism research in China, this paper constructed a word frequency matrix of keywords and carried out the cluster analysis. It also introduced the chi-square statistic and selected keywords with high correlations to interpret the clustering results. The number of articles ensures the comprehensiveness of the analysis results. High-frequency keywords and low-frequency keywords are not distinguished and screened to avoid the omission of valuable and important information. Keywords with high correlations provide convenient methods for the description and deep analysis of the clustering results.

Based on the analysis of the literature amount, time, and creative content, it can be found that there is much research on tourism informatization, status quo and countermeasures of smart tourism, tourism information technology, and smart tourism scenic area management. Among them, tourism informatization, the status quo, and countermeasures have been hotspots in recent years. Research on tourism big data, tourism service supply chain, Internet + tourism, and tourism experience has attracted more attention and may be the research hotspots in the future. The impact of tourism industry policy and the application of new information technology have always been an important source of creativity in smart tourism research.

## References

1. H. J. Yang and L. Zhang, "Promoting Creative Computing: Origin, Scope, Research and Applications," *Digital Communications and Networks*, Vol. 2, pp. 84-91, 2016
2. G. B. Ji and L. J. Wang, "Research Progress on Smart Tourism Home and Abroad," *Ecological Economy*, Vol. 12, pp. 107-110, 2015
3. Y. Q. Fu and X. M. Zheng, "Research on Recent Development and Countermeasures of Smart Tourism in China," *Development Researches*, Vol. 4, pp. 62-65, 2013
4. S. S. Huang, "Summary on Domestic Smart Tourism Researches," *Geography and Geo-Information Science*, Vol. 2, pp. 97-101, 2014
5. Y. P. Li, Z. Z. Hu, C. Huang, and L. Q. Duan, "Discussion on Concept of Smart Tourism in the View of Tourism Information Service," *Tourism Journal*, Vol. 5, pp. 106-115, 2014
6. L. Y. Zhang, N. Li, and M. Liu, "Basic Concept and Theoretical System of Smart Tourism," *Tourism Journal*, Vol. 5, pp. 66-73, 2012
7. L. Y. Zhang, "Smart Tourism: Coming of a Time of Personalized Customization and Smart Public Service," *Tourism Journal*, Vol. 2, pp. 3-5, 2012
8. Z. Xiang, Y. P. Li, and R. F. Daniel, "Searching for the Future: Emergence of Smart Tourism," *Tourism Journal*, Vol. 12, pp. 8-12, 2015
9. B. L. Xu, D. H. Li, Y. L. Qian, and Y. T. Liu, "Smart Tourism: A New Trend of Tourism – Summary based on Existing Research Results," *Resource Development & Market*, Vol. 7, pp. 781-784, 2013
10. J. S. Gao, Y. Y. Li, Y. Q. Liang, and Y. Q. Sun, "Study on Knowledge Discovery Model based on Visualization of Literature Data," *Library Science Researches*, Vol. 1, pp. 49-56, 2016
11. B. K. Chen and Y. F. Wang, "Content Study for Analysis on Disciplinary Structure and Evolution Visualization," *Library Information Service*, Vol. 21, pp. 87-95, 2016
12. Y. Y. Chen, W. F. Tian, and J. H. Wu, "Research on Visualization Analysis Method of Tracking and Trend Prediction in Research Hotspots in Subject Area," *Information Theories and Practices*, Vol. 6, pp. 117-121, 2017
13. X. Liu and P. Z. Zhang, "Visualization Analysis of Knowledge Management Research based on CSSCI Data," *Information Journal*, Vol. 1, pp. 92-97, 2013
14. J. P. Qiu and H. Liu, "Visualization Analysis of Knowledge Network Research based on Knowledge Mapping," *Information Science*, Vol. 12, pp. 3-8, 2013
15. Y. Chen, W. Hu, and J. Wang, "Analysis on Hotspots and Trends of Smart Education Researches based on Knowledge Map," *China Distance Education*, Vol. 9, pp. 21-26, 2016

**Gongpeng Zhang** is a Ph.D. student in the School of Management at China University of Mining & Technology. His research interests include management science and engineering and smart tourism. E-mail: gongpeng@buaa.edu.cn.

**Yuncai Ning** is a deputy dean, professor, and doctoral tutor in the School of Management at China University of Mining & Technology. His research interests include technical and economic evaluation and project management.

**Chi Zhang** is a deputy dean and associate professor in the Tourism College at Beijing Union University. His research interests include tourism economics and tourism information. E-mail: tibuzc@163.com.

**Zhaogang Wang** is a Ph.D. student in the School of Management at China University of Mining & Technology. His research interests include management science and engineering and text mining.

**Beu Lee** is an assistant professor in the College of Business at Texas A&M University. Email: eve.lee@tamusa.edu.