

Repositório ISCTE-IUL

Deposited in *Repositório ISCTE-IUL*:

2018-03-20

Deposited version:

Post-print

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Moro, S., Rita, P. & Cortez, P. (2017). A text mining approach to analyzing Annals literature. *Annals of Tourism Research*. 66, 208-210

Further information on publisher's website:

[10.1016/j.annals.2017.07.011](https://doi.org/10.1016/j.annals.2017.07.011)

Publisher's copyright statement:

This is the peer reviewed version of the following article: Moro, S., Rita, P. & Cortez, P. (2017). A text mining approach to analyzing Annals literature. *Annals of Tourism Research*. 66, 208-210, which has been published in final form at <https://dx.doi.org/10.1016/j.annals.2017.07.011>. This article may be used for non-commercial purposes in accordance with the Publisher's Terms and Conditions for self-archiving.

Use policy

Creative Commons CC BY 4.0

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in the Repository
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

A text mining approach to analyzing Annals literature

Tourism has been researched for several decades, with some of the older journals dating back to the 1970s. While tourism has been a subject of several reviews in past years, an update may enlighten recent findings. Annals of Tourism Research (ATR) is a top journal for publishing tourism research results (McKercher et al., 2006). We analyzed two decades of publications between 1996 and 2016 in ATR. Such a time breadth combined with the journal's relevance justifies a critical novel analysis.

The distribution of ATR publications per year is expressed in Table 1. There is a significant increase in the more recent years after 2010, with the number of articles published rising clearly above sixty per year.

Table 1 – Number of articles analyzed.

Year	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
Nr.	46	45	39	39	45	44	50	46	48	51	52
Year	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	Total
Nr.	50	47	29	52	67	85	72	74	85	62	1,128

Text mining was performed over the whole textual contents of every article, excluding only the references and authors' affiliation, to analyze relations between research areas, tourism products, and geographic location of the studies using topic modeling, which groups articles in topics (Moro et al., 2015). Only articles identifying a location study within the text were considered, in a total of 811 from the 1,128 articles.

The ten topics found by crossing research areas and tourism products are shown in Table 2, presented in a descending order by the number of articles, including the term that best identifies each topic for the two categories, research area and tourism product, and the β distribution value (a smaller value represents a stronger relation to the topic). We have grouped articles in time slots of seven years each to simplify the visualization of research evolution across time.

The first topic identified a relation between culture tourism and education, with a steady growth over the studied period, and encompassing the largest number of articles. The airline industry is weakly related to satisfaction (second topic) and moderately associated with its impact (fifth topic), confirming the relevance of the air travel experience to tourism. Satisfaction is also the dominant research area of the last topic, with 21 articles, but in this case related to food. This appears as an emerging area of research in tourism, with the last timeframe showing more articles than the sum of the two previous periods, suggesting that forthcoming years are expected to see an increase in publications on this domain, an area traditionally more associated with psychology studies. The tightest relation is found on topic number three, through its similar β values, providing evidence that most of the studies on events are particularly focused on planning. Two other interesting findings involve technology studies on hotels (sixth topic) and

human resources in senior tourism (seventh topic). While the former is expected to continue growing, considering Big Data sources adopted in tourism such as social media and the Internet of Things for devices used in hotels, the latter has been a concern of tourism at least since 1996, showing the particular relevance of staff for the senior market.

Table 2 - Tourism products and areas

#	Area	β	Product	β	1996-2002	2003-2009	2010-2016
206	tourism education	2.93	culture	0.22	46	62	98
129	satisfaction	0.46	airline	5.38	39	36	54
113	planning	1.05	event	1.45	27	40	46
82	sustainable	0.28	gaming	3.12	20	28	34
80	impact	0.64	airline	2.41	27	20	33
49	technology	2.49	hotel	0.33	13	14	22
48	human resource	0.24	senior	3.44	12	17	19
47	advertising	3.47	urban	0.74	16	15	16
36	risk	0.19	event	3.44	9	12	15
21	satisfaction	3.99	food	0.91	5	3	13

We also analyzed results for discovering topics that relate research areas and the continents where the research took place. We focused on continents to have a resulting list of groups of summarized and easier to interpret articles (Table 3).

Table 3 - Tourism areas and continents

#	Area	β	Continent	β	1996-2002	2003-2009	2010-2016
150	tourism education	2.79	Europe	0.11	39	39	72
134	human resource	0.91	North America	4.69	40	46	48
124	sustainable	0.42	Oceania	1.60	32	37	55
108	risk	0.76	North America	0.90	31	37	40
95	tourism education	3.45	Asia	0.11	22	28	45
94	satisfaction	0.15	Europe	4.66	22	28	44
70	planning	0.62	South America	2.97	16	21	33
36	advertising	0.18	North America	2.71	12	11	13

Interestingly, tourism education is the most studied area in both Europe and Asia. The study by Dominici and Maitland (2016) reflects such finding, focusing on the development from educational to commercial tourism in Britain. In fact, the same study is also an example of the first topic from Table 2, unveiling a common line of research on the three dimensions of culture, education, and Europe. The second topic, including 134 articles, emphasizes human resources' management in North America, while the third topic shows a tighter relation between sustainability and Oceania when compared to the relations found in the first and second topics.

While sustainability has been addressed and studied in Oceania for a few decades, it is still a research area of interest and facing numerous challenges (Tolkach et al., 2016). Topic number four shows a very close association between North America and risk. South America is a region where there is still room for research concerning the need to further develop planning strategies (seventh topic). North America dominates as a region (three topics), followed by Europe (two topics). In comparison, Africa continues to be a continent with less appeal for tourism research.

Finally, we analyzed the topics uncovered when considering both tourism products and continents, in a search for relations between both categories (Table 4). The most relevant topic associates culture tourism with North America, accounting for 189 articles. The second topic groups 178 articles in a widely studied tourism product, hotels. North America is the continent most closely related. The third topic is highly linked to the European continent and weakly associated with culture. The fourth topic is related with religion and Asia. The fifth addresses cruise tourism and is associated mainly with Oceania. Both sixth and seventh topics are more concerned with Europe, involving studies on rural tourism and events. Research on airlines with a focus on South America closes the list.

Table 4 - Tourism products and continents

#	Product	β	Continent	β	1996-2002	2003-2009	2010-2016
189	culture	0.17	North America	3.17	40	52	97
178	hotel	0.68	North America	0.98	62	62	54
168	culture	4.13	Europe	0.07	45	45	78
114	religious	2.06	Asia	0.21	26	38	50
58	cruise	2.02	Oceania	0.73	15	16	27
45	rural	0.86	Europe	4.98	13	18	14
42	event	0.51	Europe	4.04	9	13	20
17	airline	0.85	South America	1.62	4	3	10

Culture tourism is the area standing out by being addressed twice in the top three with a total of 357 articles, and emphasizing two of the most developed regions in the world, Europe and North America. Considering the rich and ancient Asian culture as well as this region's impressive growth in tourism flows, there is a research potential for further studies in culture tourism in this continent.

Interestingly, whereas in tourism areas with more articles the North American continent is more often the focus of attention, when investigating tourism products Europe is more emphasized in diversity. In fact, whereas the former is more related with culture and hotels, the latter repeats culture but adds rural and events' tourism.

As tourism research continues to further develop, the adopted text mining procedure, with no precedent in tourism, offers a reusable guide to uncover trends and gaps. Future applications of the same method can provide insights on tourism evolution by analyzing a larger body of

knowledge. This is an important issue for researchers pushing the boundaries of current state-of-the-art in a world of Big Data where stacks of relevant literature keep piling up. Nevertheless, human language subjectivity poses limitations to a fully automated analysis. Also, the analysis of a single source limits the conclusions drawn to the selected journal, although it highlights both hot topics and ATR's lesser explored domains which may be compared with other sources to unveil research overlaps and the uniqueness of ATR.

References

Dominici, S., & Maitland, R. (2016). The PTA: Promoting Swiss tours, 1888–1939. *Annals of Tourism Research*, 60, 31-47.

McKercher, B., Law, R., & Lam, T. (2006). Rating tourism and hospitality journals. *Tourism Management*, 27, 1235–1252.

Moro, S., Cortez, P., & Rita, P. (2015). Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation. *Expert Systems with Applications*, 42(3), 1314-1324.

Tolkach, D., King, B., & Whitelaw, P. A. (2016). Creating Australia's National Landscapes: Issues of collaborative destination management. *Journal of Destination Marketing & Management*, 5(2), 117-132.