

Mining Meaning of Health in Baby Boomer Travel Blogs

Weixuan Wang

Follow this and additional works at: <https://scholarworks.umass.edu/ttra>

Wang, Weixuan, "Mining Meaning of Health in Baby Boomer Travel Blogs" (2021). *Travel and Tourism Research Association: Advancing Tourism Research Globally*. 54.
https://scholarworks.umass.edu/ttra/2021/research_papers/54

This Event is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Travel and Tourism Research Association: Advancing Tourism Research Globally by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

Mining Meaning of Health in Baby Boomer Travel Blogs

Introduction

Travel blogs are underestimated research resources for tourism scholars (Pan, MacLaurin, & Crotts, 2007). Statista projected there are going to be close to 32 million active bloggers (those who blog at least once a month) in the US alone by 2020, which increase 12% from 2015 in an industry that many considered a dying one (eMarketer & Squarespace, 2016). Among all the active blogs, travel is rated as the top 5 topics shared by bloggers (eMarketer & Squarespace, 2016). User generated contents such as travel blog is widely utilized in travel planning and sharing their travel experiences (Cox, Burgess, Sellitto, & Buultjens, 2009). These content rich blogs provide abundant insights to better understand blogger's travel experiences and behaviors (Banyai & Glover, 2011).

Baby Boomers are recognized as the most avid travelers. According to the AARP travel trend study, Boomers, on average, are expected to take four to five leisure travel in 2019 (Gelfeld, 2018). Many of the Baby Boomers are motivated to travel due to their health benefits (Chen & Petrick, 2013). H. L. Kim, Uysal, and Sirgy (2019) suggested that travel can bring health benefits to older adults, improving their overall well-being and quality of life. Chen and Petrick (2013) agree that health and wellness benefits are one of the top motivators for Boomer travelers.

Baby boomers are the major contributors to travel blogs online (Freespot, 2019). Blogs are robust resources for travel research, because of their storytelling narratives provide more insights, reflecting the travel experience of the bloggers (Bosangit, Hibbert, & McCabe, 2015). By mining meaning from the baby boomer blogs, this study aims to contribute to existing literature and provide a better understanding of baby boomers travel experience, and how travel is related to their health and overall well-being.

This research studied 57 Baby Boomers travel blogs and collected their health related posts to create a robust database. This mixed-methods research project employed Natural Language Processing (NLP) methods such as top modeling to understand the travel experience of Baby Boomers, following by a semi-structured interview with the Baby Boomer bloggers to better understand verify the results of topic modeling and better understand the travel relation with health in the view of Baby Boomer travelers.

Literature Review

Travel Blog in Tourism Research

The use of blogs began in the mid-1990s (Blood, 2002). The term "web log" was coined by Jon Barger in 1997 while the word "blog" was later coined by Peter Marshals in 1999 (Nagasundaram, 2008). Schmallegger and Carson (2008) defined travel blogging as publishing personal travel stories and recommendations in the form of travel diaries or product reviews. Travel blogs are personal online diaries that are made up of one or more individual entries with a common theme around travel (Pühringer & Taylor, 2008). The main themes shared in travel blogs are destination descriptions as the climate, cuisine, transportation, attraction and culture (Pan et al., 2007; Schmallegger & Carson, 2008). In their narratives, travel bloggers shared their experiences of different destinations and recommendations for their readers. The detailed information of travel

blogs has had a huge impact on the fundamental nature of travel planning (Xiang, Magnini, & Fesenmaier, 2015).

The rich contents in travel blogs have also drawn attention from the tourism scholars. Previous scholars have examined information content within travel blogs that is relevant to travel trends, destination image, evaluation of destinations and tourist behaviors (Pan et al., 2007; Schmallegger & Carson, 2008; Xu, Yuan, Ma, & Qian, 2015). Many studies use blog data to study the tourist behaviors in destinations (Fujii et al., 2016). For example, Yuan, Xu, Qian, and Li (2016) using travel blogs to study the travel patterns of tourists in Beijing, by analyzing the word cooccurrence and networks in the blog posts of travelers. Bosangit, Dulnuan, and Mena (2012) studies the post consumption behavior of tourists using discourse analysis on travel blogs. Sharing the experiences and images by blogging allows the tourist to recollect past experience at the post-visit stage. Since the travel blogs, especially for mirco-blogs such as twitter usually have locations tagged in the posts, travel patterns of tourists can also be predicted by analyzing the geo-tagged photos shared on blogs and social media (Yang, Wu, Liu, & Kang, 2017).

Travel blogs provide narratives and multiple interrelated information about destinations (Banyai & Glover, 2011). Wenger (2008), for example, analyzed 114 travel blogs related to trips to Austria to understand similarities and differences between the blog posts and Austria's tourism markets, and to identify positive and negative perceptions of Austria as a tourism destination. Svetlana Stepchenkov (2006) conducted their research using text mining to analyze the destination image of Russia by counting words in reviews. Also, a study by Choi et al. (2007) applied a similar technique to analyze the destination image of travelers to Macau.

Previous studies suggested that the patterns and trends identified through mining blog data can be used to amend and improve service offerings potentially forgoing stronger bonds with customers (Magnini, Crotts, & Zehrer, 2011). Travel blog data were often used in marketing research. The content of travel blogs can be used for various marketing strategies such as improving and monitoring destination images and products by responding to tourists' demands and expectations, and also adjusting competitive strategies (Pan et al., 2007; Yuan et al., 2016).

The findings of the American Travel Survey also clearly demonstrate that there are similarities as well as important differences in the use of Internet and travel blogs among generation of American consumers such as the Silent generation, Baby Boomer, Gen X and Gen Y (Xiang et al., 2015). However, few studies have focused on understand a specific generation and study their travel behavior by using blog data. However, travel blogs, unlike other social media where it is difficult to obtain personal information about the users, generally provide an introduction page, where the blogs provide some personal information including their age and occupation. Therefore, the travel blogs data has great potential for scholars to study the travel behaviors and experience of specific generations of travelers, such as Baby Boomer.

Travel and Health

The link between physical activity and health has long been established (Saunders, Green, Petticrew, Steinbach, & Roberts, 2013). Older adults are motivated to travel for increasing physical activities (Hsu, Cai, & Wong, 2007) and improving physical health. Travel for leisure purposes offered significant benefits to seniors' in terms of personal health and social well-being (Hunter - Jones et al., 2007; Tung and Ritchie, 2011). However, a result of aging is the continuous deterioration in health and physical abilities (Hsu et al., 2007). Declining health ability is recognized as one of the major travel constraints for older adults (Kazeminia, Del Chiappa, &

Jafari, 2013). While some older adults are discouraged from traveling because of their health conditions, other may be concerned about potential health risks during travel. However, the health risk or concerns of the older adults are less studies in the tourism literature.

Health risk is regarded as one of the factors that could endanger the safety and security of travelers (Jonas, Mansfeld, Paz, & Potasman, 2011). The WHO (2005) associates this type of risk with international travel, and stresses that the level of risk is determined by the travelers' characteristics, their travel behavior, and environmental conditions prevailing at the destination. Previous studies have suggested there are potential health risk associated with travel most of them are related to infectious diseases (Mangili & Gendreau, 2005). For senior population, because of their declining immune systems, they are more susceptible to these infectious diseases.

The purpose of this study was to understand travel blogs as a manifestation of Baby Boomers' travel experiences. The study aims to assess the health related posted on travel blogs to better understand how travel experience is related to health and overall well-being of Baby Boomer travelers.

Methodology

This project adapted the 4 steps text mining process designed by Abdous and He (2011) in combined with follow-up interviews to better understand the travel experience of Baby Boomers through their travel blogs with a focus on the relations of travel and health. The study first identified 57 most recognized Baby Boomer blogs from top 30 baby boomer blogs, baby boomer blog awards, and top 30 retirement blogs in the past 3 years (Freespot, 2019). The selected baby boomer travel blogs must be active with at least one post per month before February 2020. The study first identifies blogs that are written by blogger who self-identified as boomer or over 50 from these awards and ranking. The study excluded blogs from travel agency website. However, some travel blogs included in the study have sponsored posts, but the majority content of the blog are reflections of the experience of the travel blogger.

The study uses commercial web crawler Octoparse (Octoparse, 2019) to scrape blog posts that are related to health from the 57 blog sites. These blog posts are identified by tags such as health or wellness added by the bloggers themselves on their websites. Images and photography posts were neglected in the study. Octoparse is a powerful web scraping tools suggested by previous researcher for scraping blog data, as it can grab open data from almost all the websites (Ahamad, Mahmoud, & Akhtar, 2017), which allows us to modify the scraping task for each individually different blog sites. After the data collection, data cleaning, preprocessing and topic modeling are conducted using the using the Orange 3 text widget based on NLTK tool kit. Topic Modeling is a statistical form of NLP that uses algorithms to summarize large amounts of texts, called a corpus, into a group of topics. A topic is a theme that occurs frequently in a corpus. These topics are associated with a probabilistic generative model. Latent Dirichlet allocation is one of the most common algorithms for topic modeling (Mehrotra & Roberts, 2018). LDA is a mathematical method for estimating both at the same time: finding the mixture of words that is associated with each topic, while also determining the mixture of topics that describes each document (Silge & Robinson, 2017).

To validate the text mining results, semi-structured interviews has been conducted utilizing the top generated by the results. All of the 50 travel bloggers identified in the research will be contacted for interest in interview using contact information provided by the travel bloggers or their social

media accounts. Interviewees were asked to recall health related topics and posts they shared on their travel blogs.

Results

Out of the 57 blogs analyzed in the study, 32 of them are written by females and 5 are written by males. 15 blogs are run by couples and 5 of the 57 blogs have multiple contributors. 41 of the blogs are websites mainly focused on sharing travel experiences, 16 blogs are lifestyle blogs of Baby Boomers that included a travel section. Despite the effort of eliminating blogs of travel agencies and destinations from the lists of analyzed blogs, about 60% of the bloggers are sponsored or have sponsored posts declared by the bloggers. Out of 57 Baby Boomer travel blogs studies, 22 of these blogs have a health/wellness section/page or health and wellness tags. These health related posts from these 22 blogs were scraped. In total, over 1700 blog entries were scraped initially.

In the data cleaning step, we removed video and photography only posts. Promotional posts, which are posts only contains links to other websites or previous posts, were also deleted. All the textual data were transformed to lowercase with html parsed, punctuation and urls removed. After the initial preprocessing of text, we removed entries with less than 50 words, which resulted in 1372 usable blog posts left.

Textual data were then prepared for analysis with tokenization and WordNet Lemmatizer normalization. Before the topic modeling, we removed common English stop words such as “and”, “or” and “in” (Schütze, Manning, & Raghavan, 2008). A customized stop words list was added in addition including the most terms (with more than 1000 occurrence) and numbers as suggested by word counts (Figure. 1).

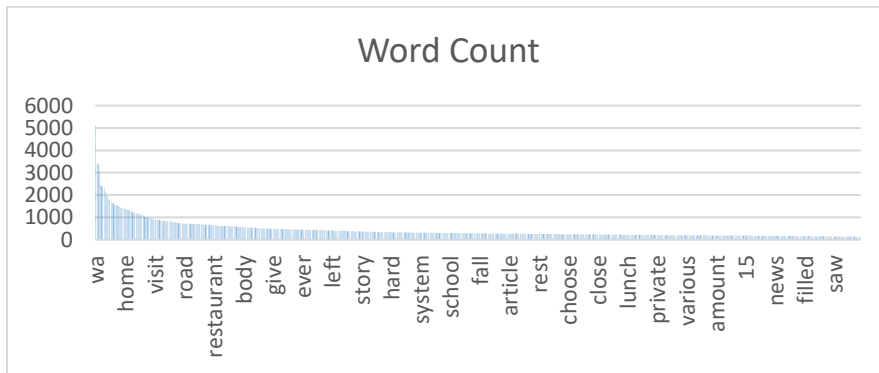


Figure 1. Word Counts of the Preprocessed Text

Topic Modeling

The selection of an appropriate topic model involves a variety of tradeoffs and judgments by the human researcher (Evans, 2014). To determine the number of topic models to select with LDA topic modeling approach, we tested various numbers of topics from 5 to 50 to find the most stable topic models. Wang, Feng, and Dai (2018) suggested the more topics generated the more stable the topics become; however, more topics adds to the model perplexity. To avoid over-clustering and improve clustering results, we set the number of topics to be 12.

Table 1. 12-topic model result of LDA analysis

Topics and Top 10 Keywords	Marginal Topic Probability
1: book, life, plan, happy, date, self, person, sleep, body, board	0.00
2: park, museum, property, center, english, white, mile, culture, shop, bag	0.02
3: home, health, care, medical, water, state, cost, insurance, stay, option	0.28
4: plan, care, benefit, eye, body, area, amount, health, insurance, happy	0.00
5: resort, room, air, hiking, flight, bar, airport, accommodation, tourism, spa	0.01
6: travel, cruise, book, happy, life, self, sleep, health, practice, plan	0.00
7: travel, trip, city, visit, tour, place, hotel, country, cruise, restaurant	0.17
8: site, town, island, point, group, guide, fun, review, photo, town	0.14
9: family, dog, life, thanks, woman, work, friend, car, picture	0.24
10: travel, cruise, experience, boomer, board, place, world, city, international	0.00
11: art, sit, path, far, social, famous, garden, dog, street, member	0.00
12: free, food, night, place, wine, eat, dinner, walking, kind, possible	0.13
Total	0.99

Table 2 displayed the 12-topic model generated by LDA with the top 10 keywords for each topic listed. To make sense of these topics, we plot the word cloud and labeled each topic based on the logical connection between their top words and the corresponding relative weights. These topic labels are verified by the researcher using concordance function to find the keywords in the text corpus and displayed the context in which the word is used. For example (see Figure 2), keywords from these topics are logically connected to different kind of travel attractions, which is verified by searching the corpus. Most blog posts that can be associated with this topic are posts about

specific destination or events such as ‐Fancy Food Show‐ and film festivals shared by these baby boomer bloggers. A full list of 12 topics and its labels are listed in Table 3.

Topic 2: Travel Attraction and Activities

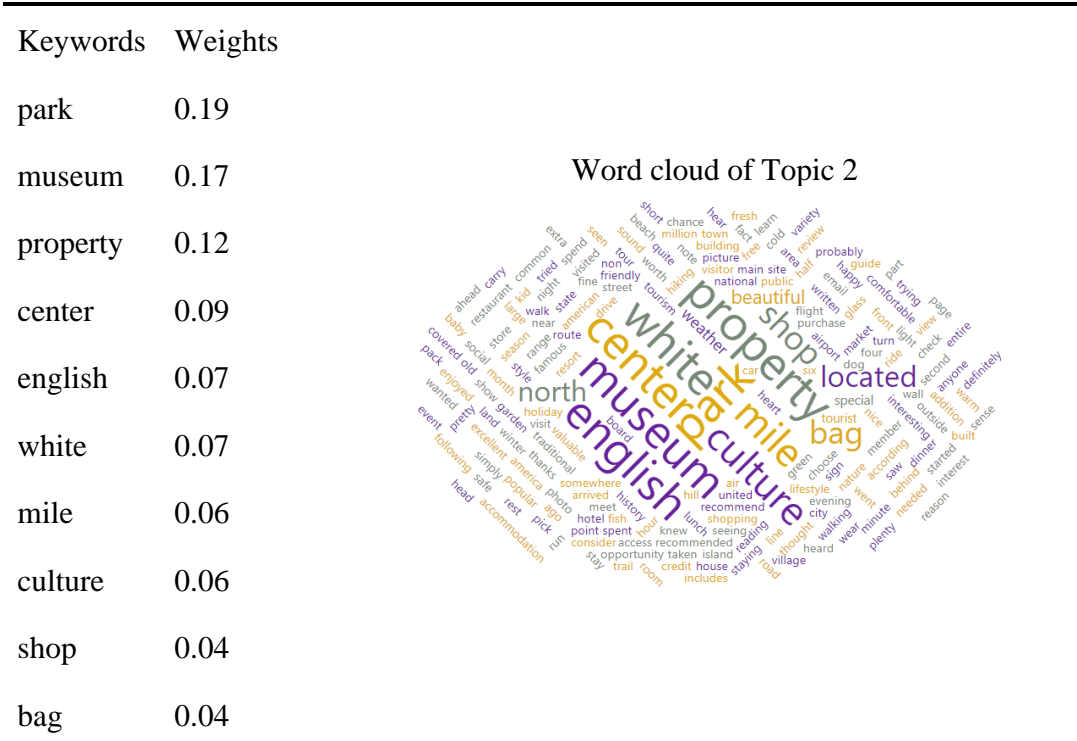


Figure 2. Example of naming Topic 2: Travel Attraction and Activities

Qualitative Validation

After the topics are generated and labeled (as shown in table 2), we conducted a brief interview with 10 travel bloggers to verify the finding of the topic models. The interviewees were able to verify most topics generated by the LDA model, except for cruise and arts.

Conclusion and Discussion

Topic modeling algorithms are a suitable method to unveil hidden thematic structures from large unstructured textual data (Blei, Ng, & Jordan, 2003). This study utilized Latent Dirichlet allocation topic modeling algorithm to mine meaning from Baby Boomer travel bloggers’ health related posts. The results suggested in the opinion of baby blogger, health is closely related to travel in different aspects, these 12 aspects (shown in Table 2) include travel planning, travel attraction and events, health care at home, health insurance and care during travel, resort, cruise, city trips, food and beverage, travel activities on-site, social well-being, healthy diet and lifestyle, and arts. The

qualitative validation interviews suggested that bloggers were able to verify most topic generated by the LDA model, except for art and cruise. We examined the two topics again. The “Art” related posts contain several series of art crafting posts by two different Boomer bloggers who tagged these posts with a “health” tag. These two bloggers were not included in the validation interview. Therefore, we decided to keep the art and social experience topics in the model.

Table 2. Labeling of the 12 topics and verification

Topics	Labeling of the Topics	Verification by Interviews
1: book, life, plan, happy, date, self, person, sleep, body, board	Travel planning	✓
2: park, museum, property, center, english, white, mile, culture, shop, bag	Travel attractions and events	✓
3: home, health, care, medical, water, state, cost, insurance, stay, option	Health condition and health care at home	✓
4: plan, care, benefit, eye, body, area, amount, health, insurance, happy	Health care/insurance for trips	✓
5: resort, room, air, hiking, flight, bar, airport, accommodation, tourism, spa	Resort experience	✓
6: travel, cruise, book, happy, life, self, sleep, health, practice, plan	Cruise experience	x
7: travel, trip, city, visit, tour, place, hotel, country, cruise, restaurant	City travel, food and beverage	✓
8: site, town, island, point, group, guide, fun, review, photo, town	Travel Activities on-site	✓
9: family, dog, life, thanks, woman, work, friend, car, picture	Social well-being	✓
10: travel, cruise, experience, boomer, board, place, world, city, international	International travel experience	✓
11: art, sit, path, far, social, famous, garden, dog, street, member	Art and social experience	x
12: free, food, night, place, wine, eat, dinner, walking, kind, possible	Healthy diet and lifestyle	✓

Notes: ✓ = Verified; x = not Verified

Limitation

One limitation of the study is that using LDA algorithms, the topics tends to be too broad as it can be seen in the topics generated. This shortcoming of LDA algorithms make the interpretation and labeling of the topics more difficult. The number of topics generated in this study is done subjectively, a more heuristic approach with training data sets may be needed in the future to better determine the number of topics generated by the LDA model, which may lead to more detailed findings from the generated topics.

References

- Abdous, M. h., & He, W. (2011). Using text mining to uncover students' technology-related problems in live video streaming. *British Journal of Educational Technology*, 42(1), 40-49. doi:10.1111/j.1467-8535.2009.00980.x
- Ahamad, D., Mahmoud, D., & Akhtar, M. (2017). Strategy and implementation of web mining tools. *International Journal of Innovative Research in Advanced Engineering*, 4, 01-07.
- Banyai, M., & Glover, T. D. (2011). Evaluating Research Methods on Travel Blogs. *Journal of Travel Research*, 51(3), 267-277. doi:10.1177/0047287511410323
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *the Journal of machine Learning research*, 3, 993-1022.
- Blood, R. (2002). *We've got Blog: How Weblogs are changing our culture*: Basic Books.
- Bosangit, C., Dulnuan, J., & Mena, M. (2012). Using travel blogs to examine the postconsumption behavior of tourists. *Journal of Vacation Marketing*, 18(3), 207-219.
- Chen, C.-C., & Petrick, J. F. (2013). Health and Wellness Benefits of Travel Experiences: A Literature Review. *Journal of Travel Research*, 52(6), 709-719. doi:10.1177/0047287513496477
- Cox, C., Burgess, S., Sellitto, C., & Buultjens, J. (2009). The Role of User-Generated Content in Tourists' Travel Planning Behavior. *Journal of Hospitality Marketing & Management*, 18(8), 743-764. doi:10.1080/19368620903235753
- eMarketer, & Squarespace. (2016). Number of bloggers in the United States from 2014 to 2020 (in millions) [Graph]. Retrieved from <https://www.statista.com/statistics/187267/number-of-bloggers-in-usa/>
- Fan, W., Wallace, L., Rich, S., & Zhang, Z. (2006). Tapping the power of text mining. *Communications of the ACM*, 49(9), 76-82.
- Freespot. (2019). Top 30 Baby Boomer Travel Blogs And Websites To Follow in 2019. Retrieved from https://blog.feedspot.com/baby_boomer_travel_blogs/
- Fujii, K., Nanba, H., Takezawa, T., Ishino, A., Okumura, M., & Kurata, Y. (2016). *Travellers' behaviour analysis based on automatically identified attributes from travel blog entries*.

Paper presented at the Workshop on Artificial Intelligence for Tourism (AI4Tourism), PRICAI.

- Gelfeld, V. (2018). *2019 Boomer Travel Trends*. Retrieved from Washington, DC:
- Guetterman, T. C., Chang, T., DeJonckheere, M., Basu, T., Scruggs, E., & Vydiswaran, V. V. (2018). Augmenting qualitative text analysis with natural language processing: methodological study. *Journal of medical Internet research*, *20*(6), e231.
- Hsu, C. H., Cai, L. A., & Wong, K. K. (2007). A model of senior tourism motivations—Anecdotes from Beijing and Shanghai. *Tourism management*, *28*(5), 1262-1273.
- Jonas, A., Mansfeld, Y., Paz, S., & Potasman, I. (2011). Determinants of health risk perception among low-risk-taking tourists traveling to developing countries. *Journal of Travel Research*, *50*(1), 87-99.
- Jones-Diette, J. S., Dean, R. S., Cobb, M., & Brennan, M. L. (2019). Validation of text-mining and content analysis techniques using data collected from veterinary practice management software systems in the UK. *Preventive Veterinary Medicine*, *167*, 61-67. doi:<https://doi.org/10.1016/j.prevetmed.2019.02.015>
- Kazeminia, A., Del Chiappa, G., & Jafari, J. (2013). Seniors' Travel Constraints and Their Coping Strategies. *Journal of Travel Research*, *54*(1), 80-93. doi:10.1177/0047287513506290
- Kim, H. L., Uysal, M., & Sirgy, M. J. (2019). Seniors: Quality of Life and Travel/Tourism. In *Best Practices in Hospitality and Tourism Marketing and Management* (pp. 241-253): Springer.
- 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. *Technological Forecasting and Social Change*, *123*, 362-369. doi:<https://doi.org/10.1016/j.techfore.2017.01.001>
- Magnini, V. P., Crotts, J. C., & Zehrer, A. (2011). Understanding customer delight: An application of travel blog analysis. *Journal of Travel Research*, *50*(5), 535-545.
- Mangili, A., & Gendreau, M. A. (2005). Transmission of infectious diseases during commercial air travel. *The Lancet*, *365*(9463), 989-996.
- Mehrotra, S., & Roberts, S. (2018). *Identification and validation of themes from vehicle owner complaints and fatality reports using text analysis*. Paper presented at the 97th Annual Meeting of the Transportation Research Board, Washington, DC.
- Nagasundaram, M. (2008). E-Collaboration through blogging. In *Encyclopedia of E-Collaboration* (pp. 198-203): IGI Global.
- Pan, B., MacLaurin, T., & Crotts, J. C. (2007). Travel Blogs and the Implications for Destination Marketing. *Journal of Travel Research*, *46*(1), 35-45. doi:10.1177/0047287507302378
- Pühringer, S., & Taylor, A. (2008). A practitioner's report on blogs as a potential source of destination marketing intelligence. *Journal of Vacation Marketing*, *14*(2), 177-187.
- Saunders, L. E., Green, J. M., Petticrew, M. P., Steinbach, R., & Roberts, H. (2013). What are the health benefits of active travel? A systematic review of trials and cohort studies. *PloS one*, *8*(8).

- Schmallegger, D., & Carson, D. (2008). Blogs in tourism: Changing approaches to information exchange. *Journal of Vacation Marketing*, 14(2), 99-110.
- Schütze, H., Manning, C. D., & Raghavan, P. (2008). *Introduction to information retrieval* (Vol. 39): Cambridge University Press Cambridge.
- Silge, J., & Robinson, D. (2017). *Text mining with R: A tidy approach*: " O'Reilly Media, Inc."
- Wang, W., Feng, Y., & Dai, W. (2018). Topic analysis of online reviews for two competitive products using latent Dirichlet allocation. *Electronic Commerce Research and Applications*, 29, 142-156.
- Xiang, Z., Magnini, V. P., & Fesenmaier, D. R. (2015). Information technology and consumer behavior in travel and tourism: Insights from travel planning using the internet. *Journal of Retailing and Consumer Services*, 22, 244-249.
- Xu, H., Yuan, H., Ma, B., & Qian, Y. (2015). Where to go and what to play: Towards summarizing popular information from massive tourism blogs. *Journal of Information Science*, 41(6), 830-854. doi:10.1177/0165551515603323
- Yang, L., Wu, L., Liu, Y., & Kang, C. (2017). Quantifying tourist behavior patterns by travel motifs and geo-tagged photos from flickr. *ISPRS international journal of geo-information*, 6(11), 345.
- Yuan, H., Xu, H., Qian, Y., & Li, Y. (2016). Make your travel smarter: Summarizing urban tourism information from massive blog data. *International Journal of Information Management*, 36(6), 1306-1319.