Developing and Testing a Domain-Specific Lexical Dictionary for Travel Talk on Twitter (#ttot)[1]

Florian J. Zach[a], Scott A. Wallace[b], Iis P. Tussyadiah[c] and Priya Narayana S.[b]

[a] School of Hospitality Business Management
Washington State University, USA
florian.zach@wsu.edu

[b] School of Engineering and Computer Science
Washington State University, USA
{wallaces; p.narayanasubramani}@wsu.edu

[c] School of Hospitality and Tourism Management
University of Surrey, UK
i.tussyadiah@surrey.ac.uk

Abstract

The wealth of electronically generated communication combined with increased computing power and sophisticated algorithms provides the opportunity for destination managers to listen to travellers. Identification of sentiment with a domain-oriented lexicon is beneficial for natural language processing to analyse public opinion. Indeed, in the context of travel, sentiment analysis enables tourism decision makers to devise marketing and development strategies that address the information learned. This study presents a lexical dictionary approach for sentiment extraction and opinion mining of travel related messages posted using the Twitter microblogging service. In this study, we propose a human coded sentiment dictionary specific to the travel context. Terms were identified from a pool of more than 1.38 million travel related tweets collected over a nine-month period. Human coders assigned sentiment scores to these terms and the travelMT 1.0 dictionary was produced to enhance the existing labMT 1.0 dictionary. The quality of the travelMT 1.0 dictionary was tested against the original labMT 1.0 dictionary and human judges. We found that, with a larger number of travel terms in a tweet, the enhanced dictionary, travelMT 1.0, produces a more accurate sentiment score than the labMT 1.0 dictionary.

Keywords: lexical sentiment analysis, travel, Twitter, opinion mining.

1 Introduction

The analysis of public communication on social media and other platforms provides the opportunity to determine consumer sentiment towards brands (Ghiassi, Skinner, & Zimbra, 2013), to better understand traveller's experiences (Tussyadiah & Zach, 2017), and to facilitate predictions of the stock market (Bollen, Mao, & Zeng, 2011). Using publicly available communication data, especially opinion-rich data from blogs or reviews, is a new opportunity to understand people's opinion regarding products and services. The fragmented travel and tourism industry can particularly benefit from mining visitor opinions. That is, experiences created at a destination are typically a

---

[1] Citation: Zach, F.J., Wallace, S.A., Tussyadiah, I.P., & Narayana S., P. (2018). Developing and testing a domain-specific lexical dictionary for Travel Talk on Twitter (#ttot). In Stangl, B., & Pesonen, J. (Eds.), *Information & Communication Technologies in Tourism 2018* (pp. 528-539). Springer.

result of traveller interactions with products and services from multiple providers spanning from transportation to lodging, food and beverage and destination-related activities (sports, attractions etc.). While the individual providers might have access to proprietary data (e.g. individual ratings on hotel review websites or transaction and ad-tracking data), tourism managers that aim to market the destination as one seamless experience often do not have access to such data. Mining opinion about a destination and services thus enables tourism managers to understand what travellers like/dislike and to learn how they communicate this with others. As such, the sentiment embedded in public communication data can be analysed to guide destination development, marketing initiatives, and future investment.

The computational analysis of opinion rich data is conducted via sentiment analysis (Pang & Lee, 2008). There are two main approaches to automatically extract sentiment: sentiment classification and dictionary-based approach. Sentiment classification is a supervised machine learning approach that builds classifiers from labelled text data and extracted features (e.g. unigrams) (Taboada, Brooke, Tofiloski, Voll, & Stede, 2011). With lexical analysis, on the other hand, the sentiment of a text is calculated using a dictionary that associates key semantic elements, often words and phrases, with specific sentiment values (Turney, 2002). During analysis, words listed in the dictionary are extracted and their sentiment values from the dictionary are used to calculate a single score for the analysed text (e.g., the mean sentiment across all constituent words or phrases in the text). The dictionary-based approach is often preferred for its simplicity and, in many cases, it can perform better than the more complex machine learning models (Schmunk et al., 2014). As such, this study aims to generate a travel-centric dictionary for sentiment analysis of travel-related communication on Twitter. Specifically, this study builds upon the previously well-tested and applied labMT 1.0 dictionary (Dodds, Harris, Kloumann, Bliss, & Danforth, 2011).

2 Value of sentiment analysis for the travel industry

Tourists increasingly leave digital traces during all travel-related activities and, with the advent of social media, more and more conversations about tourism experiences take place online in the forms of consumer reviews, blogs and microblogs, as well as discussion forums. As a result, consumer-generated content associated with destination evaluations and tourist satisfaction or dissatisfaction with products and services becomes abundant and easily accessible. Collectively, consumer-generated content forms destination online reputation (Marchiori & Cantoni, 2011), serves as a highly relevant source of information for travellers and thus contributes to destination image formation, supporting travel planning and decision-making processes (Lexhagen et al., 2012). Conversely, for destination managers and travel providers, consumer-generated content provides a knowledge base to enhance service quality (Lexhagen et al., 2012) and to better appeal to relevant target markets (Leung et al., 2013).

However, the sheer number of consumer-generated content creates challenges due to the complex tasks of finding relevant information and monitoring the progress of relevant conversations online (Martínez-Cámara et al., 2012). Travellers scouring social media will benefit from a more concise representation of opinions toward a particular product or destination (Bosangit et al., 2009) and destination managers from being able to extract important features and link them to consumer perception and evaluation. However, the average human reader will have difficulty to accurately summarise the information and opinions contained in the various channels of social media today (Liu & Zhang, 2012). The task of understanding consumer-generated content is even more complex when opinions are not expressed explicitly. Therefore, it is necessary to develop systems that automatically search, retrieve, classify, and present point of views from a massive number of consumer-generated content. Sentiment analysis or opinion mining, "the computational study of people's opinions, appraisals, attitudes, and emotions toward entities, individuals, issues, events, topics, and their attributes" (Liu & Zhang, 2012, p. 415), has emerged to solve these complex challenges.

The interest in sentiment analysis among travel and tourism researchers has grown in recent years (e.g., Marrese-Taylor, Velásquez, & Bravo-Marquez, 2014; Schmunk et al., 2014; Ye, Zhang, & Law, 2009). Schuckert, Liu and Law (2015) reviewed publications on hospitality and tourism online reviews and found a cluster on sentiment analysis and opinion mining, where researchers demonstrated significant relationships between valence of online reviews and purchase intention (e.g., Capriello et al., 2013; Pekar & Ou, 2008). Researchers have utilized and compared different supervised machine learning algorithms (e.g., Support Vector Machine, Naïve Bayes, N-gram-based model; Ye, Zhang, & Law, 2009) as well as lexicon-based approach for sentiment classification of tourism-related content (Schmunk et al., 2014). Schmunk et al. (2014) found that, surprisingly, dictionary-based methods, which are easier to implement and do not necessitate in-depth knowledge of data mining, in some cases perform better than the more complex machine learning methods. However, as most previous research used general-purpose lexicons for sentiment analysis of travel-related products, researchers recognised the need for domain-specific lexicons with respect to product features to increase accuracy (Pekar & Ou, 2008).

3 Method

3.1 Travel-lexicon subset sentiment scores

There are two assumptions for calculating sentiment scores using lexical dictionaries: prior polarity (i.e., words have a sentiment value independent of context) and that the sentiment can be expressed as a numerical value (Osgood, Suci, & Tannenbaum, 1957). To build a travel dictionary, this study followed the same approach as the labMT 1.0 dictionary, which was built to evaluate happiness on Twitter (Dodds et al., 2011), although our study is smaller in scale. To build the labMT 1.0 dictionary, researchers collected three years of tweets (about 20 million per day) and identified a list of the 5,000 most frequent terms. That list was matched with the 5,000 most frequent terms from other sources (Google Books in English, music lyrics, and the New York Times) and resulted in a 10,222 unique list of most frequent terms in English. Using the Amazon Mechanical Turk (MTurk) service, each term received 50 human ratings on a scale from 1 (sad) to 9 (happy). As such, the Dodds et al. (2011) study was consistent with popular valence ratings of the Affective norms for English words (ANEW) study by Bradley and Lang (1999) and has proven useful for sentiment analysis.

Our study draws from more than 1.38 million travel tweets collected over a nine-month period from fall 2015 to summer 2016. All tweets contain the #ttot (travel talk on twitter) hashtag or are a response to tweets containing #ttot. Tweets from any source (human, organization, or bot) were retained. The most frequent 2,000 terms were extracted and manually reviewed to eliminate non-English terms, numbers, conversation queues from question and answer sessions (a1, a2, etc.), mentions (terms preceded with a @-symbol) of travel bloggers. Additionally, individual terms that are part of a geographic location (e.g. Costa) were merged to fully represent the location (e.g. Costa Rica). Terms that were listed both with and without a prefix (# for hashtags or @ for mentions) were retained Similarly, both singular and plural forms of terms were kept; for example, the terms hotel, hotels, #hotel and #hotels are included so long as each variant occurred within the top 2,000 terms. In an effort to develop a comprehensive dictionary, the terms were manually screened to identify terms that belong to sets (e.g. days of the week, months, US states, countries). If about one third of set terms were present, the remainder was added. Furthermore, the dictionary section of US travel guide books to Sweden, France, Brazil and Japan from different publishers were consulted to identify relevant travel terms that were subsequently added to the list (e.g. ferry, shower, border crossing, check point, first class, economy class, delay, stormy, cloudy). These efforts resulted in a list of 1,983 unique terms we call the travel-lexicon subset. Using MTurk, each term was rated by 38 to 43 individuals on the 1 (sad) to 9 (happy) scale. MTurkers were limited to being located in the USA and having performance rating of 99% or higher. Respondents were asked to evaluate the terms in the context of travel.

3.2 Tweet sentiment scores

To evaluate the usefulness of term sentiment scores identified in the travel-lexicon subset, a random set of 1,000 tweets was extracted from the overall pool of more than 1.38 million travel tweets. To qualify, tweets had to meet these criteria: English language (by checking both the language meta-information of the tweet and rejecting tweets that did not contain at least three words common in the English language), variety of users (maximum one tweet per user), range of sentiment (scores were calculated using labMT 1.0 and tweets with a range of labMT scores were selected), scorable (tweets had to contain at least one word each from the travel-lexicon subset and the labMT 1.0 dictionaries), original tweet (retweets were rejected), and unique content (each tweet had to produce a distinct term-frequency vector; meaning no two tweets could have the same frequency of terms).

The set was then scored three times. First, all 1,000 tweets were assigned sentiments using different sentiment dictionaries: labMT 1.0, and travelMT 1.0 – an expansion of labMT 1.0 by adding new terms unique to the travel-lexicon subset and updating existing scores for terms already listed in labMT 1.0 (see details in Section 4.4). Following Dodds et al. (2011), the sentiment for each tweet was calculated with a naïve algorithm that scaled up term sentiments to tweet sentiments by computing the mean of term sentiment scores found within a tweet. Formally, tweet sentiment (TS) was calculated as

$$TS = \frac{\sum_{i=1}^{N} ts_i f_i}{\sum_{i=1}^{N} f_i}$$

with ts_i representing term sentiment for any given dictionary and f_i representing frequency of the *i*th term. Differences between sentiment scores of the two dictionaries were used to identify a subset of 130 tweets with a balanced representation of tweets: similar tweets (difference score within +/- ¼ standard deviation from the mean difference score) or dissimilar tweets (difference score between 1 and 2 standard deviations higher or lower than the mean difference score). Second, these 130 tweets were also human scored from 1 (sad) to 9 (happy) by 44 to 46 individual MTurk coders. The same MTurk criteria as for term rating were used (location = USA, performance rating = 99% or higher).

4 Findings

4.1 Terms unique to travel-lexicon subset

Most of the 843 unique terms in the travel-lexicon subset are terms that are to a large extent travel-centric (see Table 1 for an excerpt). The mean sentiment across all unique terms is 6.03598.

Table 1. Top, median, and last ten unique travel-lexicon subset terms based on sentiment score (descending sentiment score).

Top 10	Score	Middle 10	Score	Last 10	Score
spectacular	8.46511	#Macedonia	6.16216	Libya	2.48780
day off	8.38095	Taipei	6.16129	South Sudan	2.45000
sunsets	8.30952	EUR	6.15625	wheelchair	2.41860
#paradise	8.30952	#Budapest	6.15385	Somalia	2.39024
waterfall	8.23810	24/7	6.14634	illegal substance	2.36585
vibrant	8.21951	#Hong Kong	6.14286	#danger	2.35000
breathtaking	8.18605	#volunteer	6.13514	North Korea	2.23256
beaches	8.17073	reservation	6.12500	Syria	2.20930
waterfalls	8.15000	Georgia (US state)	6.12195	missing luggage	1.66667
coastal	8.14634	#people	6.12195	missing person	1.14634

Notes: Across all 1,983 terms "relaxing" received the highest score (8.6) and "missing person" the lowest sentiment score.

4.2 Travel-lexicon subset terms in the labMT 1.0 dictionary

Sentiment scores of the 1,140 terms listed in both the travel-lexicon subset and labMT 1.0 were compared to identify terms with either close or very different scores. It was found that in terms of absolute score differences, 10 terms differ by 2 or more. These terms are big movers as they sway the sentiment considerably. The other differences are as follows: 174 terms have absolute differences from 1 to 2 (referred to as movers), 382 from .5 to 1, 279 from .25 to .5 and 295 differ by less than .25 (terms with a difference of less than 1 are non-movers). Among the big movers, three scored higher in a travel context; these are falls, hidden and retreat. Not surprisingly, terms such as school, work, jobs, lines and job score lower for travel. An interesting finding is that the terms kids and baby also score lower for travel (Table 2).

Table 3 shows an excerpt of ten movers (absolute score differences from 1 to 2) that have an absolute difference around 1.5. Among these ten terms four score higher in a travel context (breaks, sights, Italy and scenes). Of the six terms scoring lower are terms that already had a below average score in labMT 1.0 (quit, lonely and wrong).

Table 2. Big movers – travel-lexicon subset terms scored by labMT 1.0 with absolute difference of at least 2 (decending by absolute score difference).

Term	labMT 1.0	travel-lexicon subset	Difference (travel-lexicon subset - labMT 1.0)
school	6.26	3.38095	-2.87905
work	5.24	2.48718	-2.75282
kids	7.38	4.71429	-2.66571
falls	3.60	6.04762	2.44762
jobs	6.32	3.90244	-2.41756
lines	5.26	3.00000	-2.26000
job	5.96	3.78049	-2.17951
hidden	4.48	6.64286	2.16286
retreat	5.18	7.30233	2.12233
baby	7.28	5.23256	-2.04744

Table 3. Excerpt of movers - travel-lexicon subset terms scored by labMT 1.0 with absolute difference around 1.5 (decending by absolute score difference).

Term	labMT 1.0	travel-lexicon subset	Difference (travel-lexicon subset - labMT 1.0)
student	6.58	5.04878	-1.53122
quit	3.90	2.37500	-1.52500
Arkansas	5.40	3.88095	-1.51905
breaks	4.42	5.93023	1.51023
sights	6.00	7.50000	1.50000
lonely	2.86	1.37500	-1.48500
wrong	3.14	1.65854	-1.48146
united	7.32	5.85366	-1.46634
Italy	6.18	7.64286	1.46286
scenes	5.90	7.35714	1.45714

Table 4. Last ten non-movers – travel-lexicon subset terms scored by labMT 1.0 with absolute difference around 0 (ascending by absolute score difference).

Term	labMT 1.0	travel-lexicon subset	Difference (travel-lexicon subset - labMT 1.0)
dinner	7.04	7.39535	0.00465
Minnesota	5.24	5.24390	0.00390
calling	5.74	5.74359	0.00359
cook	6.64	6.64286	0.00286
Brazil	6.10	6.10256	0.00256
magazine	5.90	5.90244	0.00244

Term	labMT 1.0	travel-lexicon subset	Difference (travel-lexicon subset - labMT 1.0)
zoo	6.62	6.61905	-0.00095
Poland	5.88	5.88095	0.00095
Tennessee	5.82	5.82051	0.00051
discovered	7.00	7.00000	0.00000

Non-movers are terms with an absolute difference of less than 1. Table 4 lists the last ten terms with regards to absolute difference. Only one term ("discovered"), did not change in sentiments scores from labMT 1.0 to the travel context. Eight terms score marginally higher in the travel context. Interestingly, the labMT 1.0 scores for these last ten terms are all above 5; that is, they are all above the mean and thus more homogeneous than the top ten terms (see Table 2). Overall, the mean square error (MSE, calculated as mean of sum of square differences), for the 1,140 terms is 0.52631.

4.3 Comparing term variants

The analysis avoided lemmatizing terms; that means that terms were not reduced to their root form. As such, among the terms rated by human subjects were several terms that appeared in multiple variants; for example, subjects as singular and plural or with or without a prefix (# for hashtags or @ for mentions) or both or conjugations of verbs. One-way ANOVA was conducted on a few example terms to identify the means across all available variants were the same (null hypothesis). One-way ANOVA tests in the below sample Table 5 do not show significant differences between the means. It also shows that stemming works well in some cases (e.g. attraction example castle), but not for others (accommodation example hotel).

Table 5. Sample results of one-way ANOVA.

Term	N	Mean	Std. Dev.	Term	N	Mean	Std. Dev.
Attraction examples				**Accommodation example**			
castle	41	7.2683	1.65868	hotel	43	7.4651	1.48563
castles	42	7.5000	1.56564	hotels	42	6.7619	1.99768
#castle	39	7.2564	1.48178	#hotel	40	6.5000	1.46760
#castles	41	7.1951	1.74956	#hotels	43	6.7442	1.77406
F(3,159)=.285, p=.836				F(3,164)=2.531, p=.059			
				Transportation example			
holiday	41	7.8537	1.35205	cruise	40	7.2500	1.59727
holidays	43	7.8605	1.61218	cruising	41	6.9268	1.64909
#holiday	41	7.9512	1.41335	#cruise	40	6.4250	2.39537
#holidays	42	7.5000	1.86430	#cruising	41	7.2195	1.68069
F(3,163)=.668, p=.573				F(3,158)=.1.702, p=.169			
				Country example			
deal	41	7.4634	1.53456	brazil	39	6.1026	1.90284
deals	41	7.5854	1.41378	#brazil	41	6.1707	1.74503
#deal	42	7.4762	1.45230	F(1,78)=.028, p=.868			
#deals	41	7.4146	1.84358	**US state example**			
#traveldeals	42	7.1905	1.61151	florida	42	6.6190	2.29477

Term	N	Mean	Std. Dev.	Term	N	Mean	Std. Dev.
F(4,202)=.355, p=.840				#florida	38	6.2895	2.34703
				F(1,78)=.403, p=.528			

The analysis above suggests that while the travel-lexicon subset included multiple variants of the same root term, many of these variants did not have sentiment scores that were statistically distinguishable from one another. This, in turn, suggests that it may not be appropriate to distinguish all term variants when scoring sentiment.

4.4 Travel tweet sentiment scores

The effectiveness of a sentiment analysis technique can be evaluated by comparing an algorithmically computed sentiment score on travel related tweets to scores provided by human analysis. For this purpose, the 130 travel related tweets described in Section 3.3 were used. Table 6 below shows the mean squared error (MSE) of sentiment provided by labMT 1.0, and travelMT 1.0, which is the combination of labMT 1.0 plus the addition of the travel-lexicon subset. travelMT 1.0 adds 843 new travel-related terms to the dictionary and provides new scores for the 1,140 travel-related terms that were already a part of labMT 1.0. Based on the analysis in Section 4.3, travelMT 1.0 merges scores for term variants with different prefixes (e.g., #hotel, and hotel) to form a single term variant (e.g., hotel) whose semantic value is the average of all constituent terms. To maintain consistency with labMT 1.0's dictionary, stemming and lemmatization *is not used* to merge term variants with the same root (e.g., hotels and hotel remain as two distinct terms in travelMT 1.0 as they do in labMT 1.0)

Table 6: Deviation from Human-scored sentiment over 130 tweets

	labMT 1.0	travelMT 1.0
MSE	1.034	1.085

Across the entire set of 130 test tweets, the difference in mean squared error between labMT 1.0 and travelMT 1.0 compared against Human-scored sentiment value is minor (0.051) with travelMT 1.0 performing somewhat worse than labMT 1.0. However, the

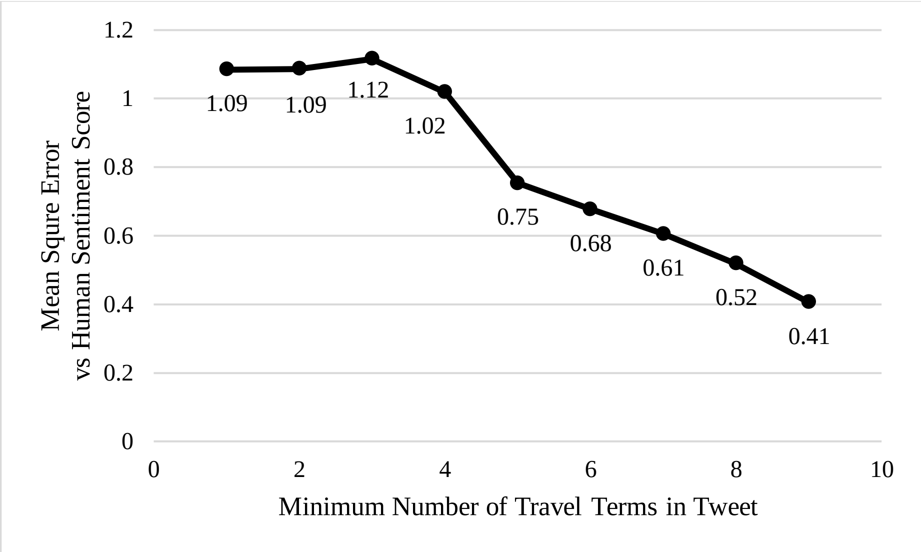

Fig. 1. Mean Square Error of travelMT 1.0 Prediction Error Decreases with More Travel Content

MSE improves when the number of travel-related terms in a tweet increases. Figure 1 plots the MSE between travelMT 1.0's sentiment prediction and the Human-scored sentiment value as a function of the number of travel-related terms are in the tweet. It is clear that travelMT 1.0 does improve sentiment scores *for tweets with sufficient travel-related content*. Figure 1 shows this relationship.

Table 7 illustrates representative tweets with varying travel content and indicates Human-scored sentiment value along with predicted sentiment values provided by labMT 1.0 and travelMT 1.0.

Table 7. Representative tweets with varying travel content and sentiment scores.

[# Travel Terms] Text	Human Scored Sentiment	labMT 1.0 [Error2]	travelMT 1.0 [Error2]
[1] Bodo library Norway. Queens College Oxford. Stourhead House Wiltshire	5.28	6.54 [1.57]	6.56 [1.63]
[4] Lovely having you @fit_travels - please come back any time you need to relax! #crownlanta #kohlanta #travel	6.85	5.84 [1.01]	5.81 [1.07]
[5] @TK_INDIA Don't lie. You are the worst airline company and worst customer service of the world.	3.74	4.78 [1.08]	4.61 [0.76]

[6] #bucketlist #Batanes the northern most #island of the #PH Said to have slow paced living and views of grandeur #ttot	6.30	5.31 [0.98]	5.68 [0.38]
[8] The tea garden in Munnar you can't miss it... #munnar #teagarden #kerala #photography #ttot #nature #landscape	6.11	5.38 [0.53]	5.59 [0.27]

5 Conclusion and Recommendation

This study makes two major contributions. First, we developed and tested a domain-specific lexical dictionary for travel-related online conversations (travelMT 1.0). Specifically, this study enhanced an existing dictionary by adding travel terms and updating scores to the travel context. Second, the travelMT 1.0 dictionary was successfully applied to rate travel-related tweets that contain multiple travel terms. While the dictionary underperformed slightly for tweets with only few travel terms, further analysis revealed that as the number of travel terms increases so does travelMT 1.0 performance. Further analysis revealed that to improve travelMT 1.0 scoring performance for Twitter, more terms can be added and more labMT 1.0 terms should be re-scored in a travel context. Another approach to achieve better scores for travel conversations is to develop a dictionary using supervised machine learning algorithms. Another shortcoming is that Twitter demographics do not necessarily match traveller demographics. To overcome this limitation, the dictionary should be tested in a more travel specific context, such as hotel reviews.

Finally, tourism managers can use the travelMT 1.0 dictionary to perform sentiment analyses on social media communication relevant to products and services provided, enabling them to learn about visitor sentiment towards their destination. Managers can extract which aspects of a visit were liked/disliked and can use this information to adjust their social media communication on the short term and make necessary investments to capitalize on positive or counteract negative experiences.

References

Abbey-Lambertz, K. (2016).We've been trying to make cities walkable for years. 'Pokémon Go' did it overnight. *The Huffington Post*. http://www.huffingtonpost.com/entry/pokemon-go-cities-walking-public-space_us_578e8581e4b07c722ebc993e

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science, 2*(1), 1-8. doi:10.1016/j.jocs.2010.12.007Bosangit, C., McCabe, S. & Hibbert, S. (2009). What is told in travel blogs? Exploring travel blogs for consumer narrative analysis. In Höpken, W., Gretzel, U. & Law, R. (Eds.), *Information and Communication Technologies in Tourism 2009* (pp.61-71). Springer.

Bosangit, C., McCabe, S., & Hibbert, S. (2009). What is Told in Travel Blogs? Exploring Travel Blogs for Consumer Narrative Analysis. In W. Höpken, U. Gretzel & R. Law (Eds.). *Information and Communication Technologies in Tourism 2009*. Wien New York: Springer, pp.61–71.

Bradley, M. M., & Lang, P. J. (1999). *Affective norms for English words (ANEW): Insutruction manual and affective ratings.* Technical Report C-1: The Center for Research in Psychophysiology, University of Florida.

Capriello, A., Mason, P.R., Davis, B., & Crotts, J.C. (2013). Farm tourism experiences in travel reviews: A cross-comparison of three alternative methods for data analysis. *Journal of Business Research*, 66(6), 778-785.

Dodds, P. S., Harris, K. D., Kloumann, I. M., Bliss, C. A., & Danforth, C. M. (2011). Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter. *PLoS ONE, 6*(12), e26752. doi:10.1371/journal.pone.0026752

Ghiassi, M., Skinner, J., & Zimbra, D. (2013). Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with Applications, 40*(16), 6266-6282. doi:10.1016/j.eswa.2013.05.057

Leung, D., Law, R., van Hoof, H., & Buhalis, D.. (2013). Social media in tourism and hospitality: A literature review. *Journal of Travel & Tourism Marketing*, 30(1-2), 3-22.

Lexhagen, M., Kuttainen, C., Fuchs, M., & Höpken, W. (2012): Destination Talk in Social Media: A Content Analysis for Innovation. In: Christou, E., Chionis, D., Gursory, D. & Sigala, M. (eds.). *Advances in Hospitality and Tourism Marketing & Management*, Corfu.

Liu, B., & Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In Aggarwal, C.C., Zhai, C.X. (Eds.), *Mining Text Data* (pp. 415-463). Springer.

Marchiori, E., & Cantoni, L. (2011). The online reputation construct: Does it matter for the tourism domain? A literatire review on destinationss' online reputation. *Information Technology & Tourism*, 13(3), 139-159.

Marrese-Taylor, E., Velasquez, J.D., & Bravo-Martinez, F. (2014). A novel deterministic approach for aspect-based opinion miing in tourism product reviews. *Expert Systems with Applications*, 41, 7764-7775.

Martínez-Cámara, E., Martin-Valdivia, M.T., Urena-Lopez, L.A., & Montejo-Raez, A. (2012). Sentiment analysis in Twitter. *Natural Language Processing*, 20(1), 1-28.

Osgood, C., Suci, G., & Tannenbaum, P. (1957). *The measurement of meaning*. Urbana, IL: University of Illinois Press.

Pang, B., & Lee, L. (2008). *Opinion Mining and Sentiment Analysis* (Vol. 2): Now Publishers.

Pekar, V., & Ou, S. (2008). Discovery of subjective evaluations of product features in hotel reviews. *Journal of Vacation Marketing*, 14(2), 145-155.

Schuckert, M., Liu, X., and Law, R. (2015). Hospitality and tourism online reviews: Recent trends and future directions. *Journal of Travel & Tourism Marketing*, 32(5), 608-621.

Schmunk, S., Höpken, W., Fuchs, M., & Lexhagen, M. (2014). Sentiment analysis - extracting decision-relevant knowledge from UGC. In Xiang, Z., & Tussyadiah, I. (Eds.), *Information and Communication Technologies in Tourism 2014* (pp. 253-265). Springer, Cham.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics, 37*(2), 267-307. doi:10.1162/COLI_a_00049

Turney, P. D. (2002). *Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews*. Paper presented at the Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, Philadelphia, Pennsylvania.

Tussyadiah, I. P., & Zach, F. (2017). Identifying salient attributes of peer-to-peer accommodation experience. *Journal of Travel & Tourism Marketing, 34*(5), 636-652. doi:10.1080/10548408.2016.1209153

Ye, Q., Zhang, Z., & Law, R. (2009). Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert Systems with Applications*, 36(3), 6527-6535.

Zhang, H., Fu, X., Cai, L.A., & Lu. L. (2014). Destination image and tourist loyalty: A meta-analysis. *Tourism Management*, 40, 213-223.