Topics and Sentiments in Online Place Reviews, an Innovative Way of Understanding the Perception of a City without Asking

Bart Neuts, Egbert van der Zee, Simon Scheider, Enkhbold Nyamsuren, Thérèse Steenberghen

(Dr. Bart Neuts, KU Leuven, Celestijnenlaan 200E, 3001 Heverlee, bart.neuts@kuleuven.be)
(Dr. Egbert van der Zee, Utrecht University, Princetonlaan 8a, 3584CB Utrecht, The Netherlands, e.l.vanderzee@uu.nl)
(Dr. Simon Scheider, Utrecht University, Princetonlaan 8a, 3584CB Utrecht, The Netherlands, s.scheider@uu.nl)
(Dr. Enkhbold Nyamsuren, Utrecht University, Princetonlaan 8a, 3584CB Utrecht, The Netherlands, e.nyamsuren@uu.nl)
(Prof. Dr. Thérèse Steenberghen, KU Leuven, Celestijnenlaan 200E, 3001 Heverlee, therese.steenberghen@kuleuven.be)

1 ABSTRACT

User-generated content provides rich and easily accessible data for tourism destination managers, especially when combined with a sentiment analysis to uncover perceptions and attitudes. These reviews are often primarily useful in a business/attraction-context and scaling up their relevance for destination management is problematic. Furthermore, the reliability of such online sources can be questioned, thereby impeding its application for research and practice. By combining data of a traditional in-situ survey in five main cultural heritage attraction in Antwerp (Belgium) with scraped data of these same attractions from the TripAdvisor website, this paper attempts to shed a light on the added value and reliability of a big data sentiment analysis. The sentiment analysis combines two lexicons as well as Latent Dirichlet Allocation. The results show promise in that, even though the characteristics between the in-situ sample and the scraped sample are quite different, the sentiments and themes are largely overlapping while the Net Promotor Score as calculated via the TripAdvisor reviews is close to the measured Net Promotor Score through the visitor survey. Still, certain limitations remain within the big data sentiment analysis approach, leading to the conclusion that both methods can be highly compatible in order to efficiently generate deeper, more complete results.

Keywords: sentiment analysis, latent dirichlet allocation, natural language processing, reliability testing

2 INTRODUCTION

Social media are gaining importance as a channel to share information on a diverse set of experiences (Munar & Jacobsen, 2014; Whiting & Williams, 2013). Ranging from photos and tweets to customer feedback or entire blogs, social media offer a wealth of information. Entrepreneurs use it as timely and direct feedback channels, while researchers explore the potential application as a data source for empirical research (see e.g. Garay Tamajón & Cànoves Valiente, 2017; Stella & Mavragani, 2015; Wong & Qi, 2017). Among other techniques, applying sentiment analysis to make sense of large sets of unstructured texts present on social media is for example used to uncover political preferences of social media users (Ceron et al., 2014), predict stock prices (Nguyen & Shirai, 2015) or measure customer satisfaction (Alaei et al., 2019). While the number of start-ups and academic papers applying sentiment analysis of social media is skyrocketing, a number of research gaps still exists. A prime example is that since the utility of sentiment analysis application to social media to improve service quality of hotels or other business has been proven (Duan et al., 2016; Chang et al., 2019), linking business/attraction-level reviews with a higher-level perception of place and visitor behaviour could assist destination management organizations and city planners (van Weerdenburg et al., 2019). To this purpose, web-scraped user-generated content needs to be analysed with respect to sentiments and topics in order to evaluate overarching themes and patterns which would otherwise be difficult to detect.

A second research gap relates to the reliability of user-generated content for attitudinal and perception research. Reliability issues could potentially originate from non-authentic reviews (e.g. Balagué et al., 2016) but can also be caused by non-representativeness of online reviewers as compared to the actual population of site/destination visitors (Xiang et al., 2017). User-generated content is potentially skewed towards younger users, limited by availability of language groups, and might attract reviews at the polar ends of satisfaction scales (Presi et al., 2014). The lack of insight into the representativeness, and thus into the usefulness of social media can be seen as a major impediment to its applicability for both research and practice.

This paper presents a proof of concept of what type of information can be obtained through sentiment analysis and topic mining, in comparison with tradional survey techniques, particularly focusing on individual's value structures and attitudes towards specific locations in the city of Antwerp, Belgium. We are interested in the question to what extent place descriptions in online reviews actually reflect the diversity of

REAL CORP 2020 Proceedings/Tagungsband
15-18 September 2020 – https://www.corp.at

ISBN 978-3-9504173-8-8 (CD), 978-3-9504173-9-5 (print)
Editors: Manfred SCHRENK, Vasily V. POPOVICH, Peter ZEILE, Pietro ELISEI,
Clemens BEYER, Judith RYSER, Christa REICHER, Canan ÇELIK

893

topics and sentiments that can be found in surveys. By combining the outcome of semantic analysis of scraped reviews on a selection of urban cultural heritage attractions from TripAdvisor with surveys collected in-situ, the results can uncover similarities and differences between both methods and assess the reliability of user-generated content as an alternative to traditional survey methods. Furthermore, a topical analysis of the TripAdvisor reviews on 5 urban visitor attractions might uncover city-level themes and therefore elevate the scope of individual businesses. Such information could be used to improve city marketing and planning practices.

3   URBAN TOURISM AND DESTINATION COMPETITIVENESS

Studies on urban tourism and the city as a place of recreation are relatively new, with the research topic only really maturing since the 1980s. Prior to this, and originating from a spatial modelling history forwarded by the likes of Christaller (1963), Miossec (1976), and Yokeno (1968), tourism and recreation were seen as functions of the urban periphery. An additional problem has been the difficulty to distinguish recreational visitors from other users in a multifunctional urban entity where facilties are largely co-consumed by a multitude of user types (Ashworth & Page, 2011). Changes in the economic fabric of cities, and the role of tourism as a potential catalyst for a service-oriented urban development, inspired a surge of research during the 1990s and 2000s.

While academic interest in the field of urban tourism is therefore relatively new, the activity itself has a much longer history, with urban regions being well-established destinations due to their function as political and economic centers and transportation hubs, even before tourism was acknowledged as a recreational activity (Urry, 1990). The importance of cities as a tourism destination has grown exponentially, mirroring the continuous rise of a new leisure society (Pearce, 2001). In 2018 the top 20 international tourist cities alone accounted for roughly 18.0% of global international tourist arrivals (mastercard, 2019) – thereby even taking abstraction of the multitude of domestic tourists and excursionists being attracted to cities for leisure purposes. The same source estimates international tourist expenditure for the top 20 urban tourism destinations at US$258.99 billion (mastercard, 2019). As such, there is a clear economic incentive for cities to compete on the international tourist market (van der Borg et al., 1996; Judd & Fainstein, 1999; Zukin, 1995).

Within the inter-urban global competition, cultural heritage is used as a main source to stand out (Ashworth & Page, 2011). Particularly for the leisure market, culture and heritage are among the top visitor motivations (Richards, 2018). For long term success it is essential that the marketing message reflects the reality of the experience and a positive referal is generated (Govers & Go, 2004; Martín-Santana, 2017). It is therefore common practice for destinations to perform visitor surveys in order to collect a wide range of variables on visitor characteristics, transport methods, information sources used, motivations, attractions visited, tourist experiences, and satisfaction and loyalty (Lewalter et al., 2015; Pearce & Moscardo, 1985). While insightful and providing details that cannot be learned from pure arrival data collected by national statistical bureaus, a limitation of these visitor surveys is the expense related to the necessity of an on-site face-to-face methodology. Next to this, surveys are generally consisting of predetermined, closed questions which do not allow exploration of enot included topics (Alaei et al., 2019). Therefore, and also resulting from the ever increasing availability of online big data, there is an increasing interest and opportunity for destination management organizations to study online user-generated content as a potential alternative to uncover tourist motivations, behaviour, satisfaction, and spread (Alaei et al., 2019; Oriade & Robinson, 2018; Taecharungroj & Mathayomchan, 2019; van der Zee et al., 2020).

4   METHODOLOGY

4.1 On-site Visitor Surveys in 5 Tourist Attractions

During the period 2014-2019, Visit Flanders, the destination management organization of Flanders (Belgium), developed a subsidy-programme for tourism projects and attractions with leveraging potential for the wider sector and destination. Such projects were primarily focused on international visitors and mainly – although not uniquely – in the theme of cultural heritage. Since accountability is becoming increasingly important, a return-of-investment evaluation of publicly financed projects was warrented. Therefore, as a requirement, recipients of subsidies were required to conduct visitor surveys at the attraction in order to

collect data on, among other things, economic return, project scale, visitor satisfaction, brand effects, and international potential.

For the sake of this study, five projects were selected from the overarching thematic programme 'Antwerp Baroque': Museum aan de Stroom, Onze-Lieve-Vrouwe Kathedraal, Plantin Moretus, Rubenshuis, and Sint Carolus Boromeus. Visitor surveys took place on-site, using tablets for answer registration and being interviewer completed. The main survey ran between 1 June 2018 until 6 January 2019. Questions related to visitor profile, visitor experience, and destination. Table 1 provides a short overview of the main questions and measurement levels.

Category	Variable	Question	Measurement level
Personal characteristics	Gender	What is your gender?	Categorical
	Age	When were you born?	Ratio
	Place of residence	Where do you live?	Categorical
Trip characteristics	Information	Via which information sources have you learned of [SITE]	Categorical
	Experience	How often do you visit a museum or exhibition?	Ordinal
	Visitor type	Are you staying overnight in Antwerp or in another area in Flanders?	Categorical
	Travel company	How many people are in your travel company?	Ratio
	Children	Are there children in your travel company?	Categorical
Visitor experience	Site recommendation	Would you recommend [SITE] to friends, family and relatives?	Ratio
	Motivation	How important was 'Antwerp Baroque' for your visit to Antwerp?	Ordinal
Destination	City recommendation	Would you recommend Antwerp as a cultural destination?	Ratio
	Other attractions	Which other attractions have you visited in Antwerp?	Categorical

Table 1. Overview of main visitor survey questions

At the five combined attractions, a total of 2,474 surveys were collected, 45.6% identified as male visitors and 54.1% respondents were female. In terms of age, the majority of visitors belonged to higher age groups: 20.8% were above 64, 20.7% between 55 and 64, 18.2% from 45 to 54, 14.4% between 35 and 44, 14.6% between 25 and 34, and the remaining 11.2% being from 18 to 24. Locals were best represented in the sample, with 36.2% living in Belgium. About one in ten (12%) were travelling as a family. Visitors from the Netherlands were the second largest group at 13.8%, followed by Germany (10.1%), France (7.5%), the United Kingdom (5.3%), the United States (4.5%), and Spain (4.4%). The remaining 303 responses (12.3%) were collected from a wide range of nationalities.

The primary information sources used to plan the visit are undefined other sources (30.4%) – which an for instance relate to organized group tours – followed by the attraction website (20.6%) and recommendations by friends and family (19.1%). Review sites such as TripAdvisor only informed 3.3% of the sample respondents. The majority of the sample (50.5%) were motivated cultural tourists, visiting cultural sites 5 times a year or more. Over half (54%) did not visit more than one attraction, and if multiple attractions were visited, these were most likely a combination of Onze-Lieve-Vrouwe Kathedraal, Rubenshuis, Sint Carolus Boromeus and/or Sint-Pauluskerk. The Net Promotor Score (i.e. difference between promotors with a satisfaction score of 9 or 10 and detractors with a satisfaction score of 0 to 6) in the sample was +45.

4.2 Natural Language Processing on Scraped Review and Survey Data

Visitor sentiments of the 5 Antwerp tourist attractions were scraped from TripAdvisor on 19 december 2019. In order to simplify the Natural Language Processing (NLP) only reviews in English were collected. Scraping used the RSelenium – for fetching the page – and rvest – for extracting page components – libraries in R 3.4.0. Through the use of Document Object Model (DOM-) parsing, the dynamic contents of the TripAdvisor pages could be retrieved.

A total of 2,438 reviews were retrieved, 70 for Museum aan de Stroom, 1,339 for Onze-Lieve-Vrouwe Kathedraal, 35 for Plantin Moretus, 1,004 for Rubenshuis, and 35 for Carolos Boromeus. It turned out that (a) the number of scraped reviews is almost similar to the number of on-site surveys collected, and (b) by far the biggest contribution to review data comes from Onze-Lieve-Vrouwe Kathedraal and Rubenshuis. In terms of pure sample size, one might therefore wonder what the added value of web scraping is in comparison to traditional survey methods. One advantage is the comparatively low time and effort required for automatized scraping. In contrast, on-site reviews are time-consuming, demanding for both interviewer and interviewee, and costly. A second advantage is the opportunity to collect historic visitor data. In our sample, the earliest review was given in 2010, with 64 reviews written in 2011, 139 reviews in 2012, 193

reviews in 2013, and 245 reviews in 2014. The majority of reviews (72.1%) originated in the last five years, with 411 reviews in 2015, 377 reviews in 2016, 372 reviews in 2017, 331 reviews in 2018 and 299 reviews in 2019. The final 51 reviews missed information on the date of experience.

In terms of travel company, 14.1% of reviewers declared themselves as solo travellers, with 37.7% travelling as a couple, 11.8% being part of a family trip, and 17.2% travelled with a group of friends. Only 3.6% of the scraped sample were marked as business travel while for 15.6% of the sample, the composition of the travel company was not known. In terms of country of residence, the singular scraping of English-language reviews is clearly reflected in the numbers, with – of the 2,226 reviews with location data – Belgium accounting for 328 reviews, the Netherlands for 125 reviews, France for just 29 observations, German tourists writing 44 comments, and Spain accounting for 20 reviews. By far the largest data is generated by tourists from The United Kingdom (with 618 reviews), and the United States (with 477 reviews), therefore providing an imbalance between actual visiting nationalities, and collected sample via scraping. If we recalculate the 5-level TripAdvisor score on a 10-point scale the Net Promotor Score can be calculated as +53.2, being quite close to the Net Promotor Score of the on-site survey.

An increasingly popular way of analysing large quantities of unstructured, qualitative data is through NLP (Alaei et al., 2019). NLP is often applied to ascribe sentiments to microblogs, such as tweets and reviews in order to analyse how people feel, but also to uncover clusters of discussed topics in order to make sense of what people are writing about.

Sentiment analysis can be defined as extracting "a sentiment expressed in a document toward a certain aspect based on the subjectivity and the linguistic characteristics of the words within an unstructured text" (Kirilenko et al., 2018 p 1013). In this paper, we analyse both the review dataset as well as the open-ended survey questions answered in English using two different, unsupervised approaches to NLP. From 556 answers to the open-ended questions, a total of 163 entries in English were used for the analysis. Minimum, mean, and maximum word lengths of the survey entries are 1, 6, and 61 words respectively. The corresponding word lengths for the reviews are 9, 58, and 1019 words respectively. Overall, there are significant differences between the two datasets both in terms of the number of entries and entry size.

For the first NLP approach, we used two different predefined, 'off-the-shelf' sentiment lexicons and tested how the survey and review datasets match the lexicons. While using off-the-shelf sentiment analysis approaches have some downsides, e.g. the fact that they were not created for the purpose of the study and are thus possibly less suitable for uncovering topic-specific sentiments, using pre-existing (Alaei et al., 2019), verified lexicons saves a large amount of time and resources (Kirilenko et al., 2018). In order to cross-validate the sentiment analysis, the data was both matched to the AFINN lexicon (Nielsen, 2011) as well as to the NRC lexicon (Mohammad et al., 2013).

The AFINN lexicon consists of 2477 words which are ascribed a score ranging from -5 (derogative, words such as "bastard") to 5 (very positive, words such as "breathtaking" or "superb"). The AFINN lexicon is biased toward negative words that constitute 65% of the lexicon. The NRC emtion lexicon is a dataset of 6468 English words, which have been ascribed to one of 8 emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, and trust) and optionally either a positive or negative sentiment resulting in 13.901 entries. While the AFINN lexicon was created by a single person based on manually examining and scoring tweets, the NRC lexicon was created through crowd-sourcing (Mohammad & Turner, 2010). Both lexicons have been verified by peers and have been widely applied to uncover sentiments from short texts such as Tweets (see e.g. Mohammad et al., 2013), online reviews (Kiritchenko et al., 2014) or open-ended survey questions (Kirilenko et al., 2018). The combination between the two lexicons allows to not only to examine the level of positive and negative sentiments, but also to see which emotions can be ascribed to the texts in the datasets, allowing to better understand what causes positive or negative sentiments. For the analysis, we used the tidytext library in R v3.6.2 that by default includes both lexicons.

For the second NLP approach, the review and survey data have been studied for the presence of re-occuring topics applying a machine learning technique known as Latent Dirichlet Allocation (Blei et al., 2003). This unsupervised bag-of-words method examines the frequency distribution of words over texts and automatically extracts a predefined number of latent topics in the form of probability vectors over the corpus dictionary (Rosetti et al., 2016). The word probabilities indicate the likelihood of words co-occuring under the same topic. Furthermore, for each topic, the method generates the probability of occurrence in a text

document. In this way, LDA is able to represent the text corpus as a mixture of topics, where the document-topic probability estimates the topic mixture of a given text, and the topic-word probability estimates the mixture of words that are used to talk about a topic. The unsupervised LDA models can be used both for analysing texts according to topic dimensions, as well as to predict topic occurrences in new texts. LDA topic models have previously been applied to give further explanation on why tourist ratings in TripAdvisor reviews (Rosetti et al., 2016), to use various online sources in order to extract place activities for locations within a city (van Weerdenburg et al., 2019) and to derive controllable dimensions for managing hotel-guest interactions from online reviews (Guo et al., 2017). In this paper, we apply an unsupervised LDA topic model to uncover which topics are present in review and open-ended survey data on a selected set of heritage attractions in Antwerp. This approach allows us not only to uncover how visitors feel about their visis (which is done by a sentiment analysis), but also what they say about the touristic attractions when asked directly (in the open-ended surveys) and when they voluntarily choose to share their opion online (in the review dataset). By comparing the probability of texts over the different topics which occur in both datasets, an estimation can be given on whether the same topics are discussed.

In order to train the LDA model, we first cleaned both corpora by removing punctuation, whitespace and english stopwords, and then turned them into document-term matrices using the term-frequency inverse document frequency (tf-idf) measure, removing all words which are less frequent than 0.1. Then we estimated different LDA models on the review texts for topic numbers from 2 to 15. We picked a model with 14 topics, because it showed the highest log-likelihood. We then applied this model to estimate corresponding topic probabilities for the survey texts. The probabilities were then compared against each other, and we also used them to pick example texts from the online reviews.

4.3 Results and discussion

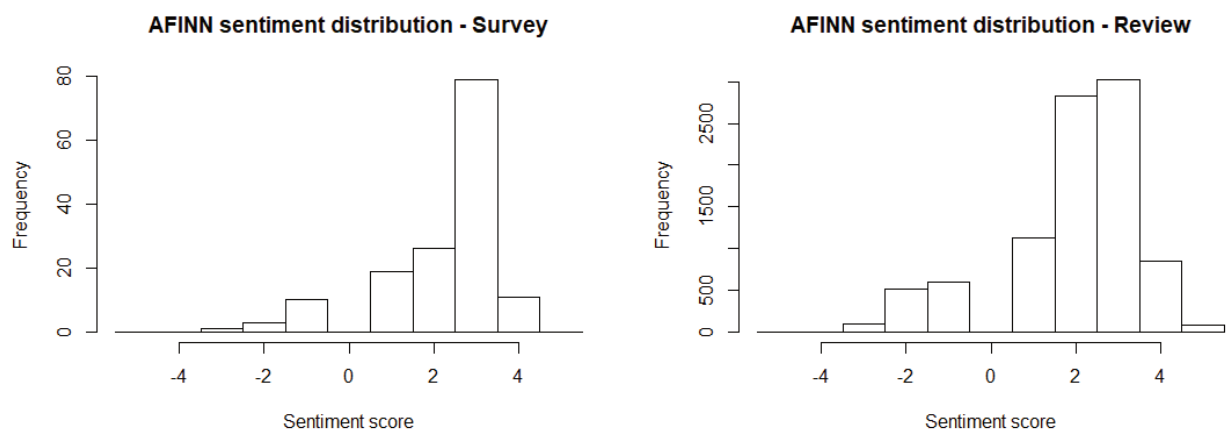
Figure 1. Distribution of sentiment scores in survey and review texts applying the AFINN lexicon
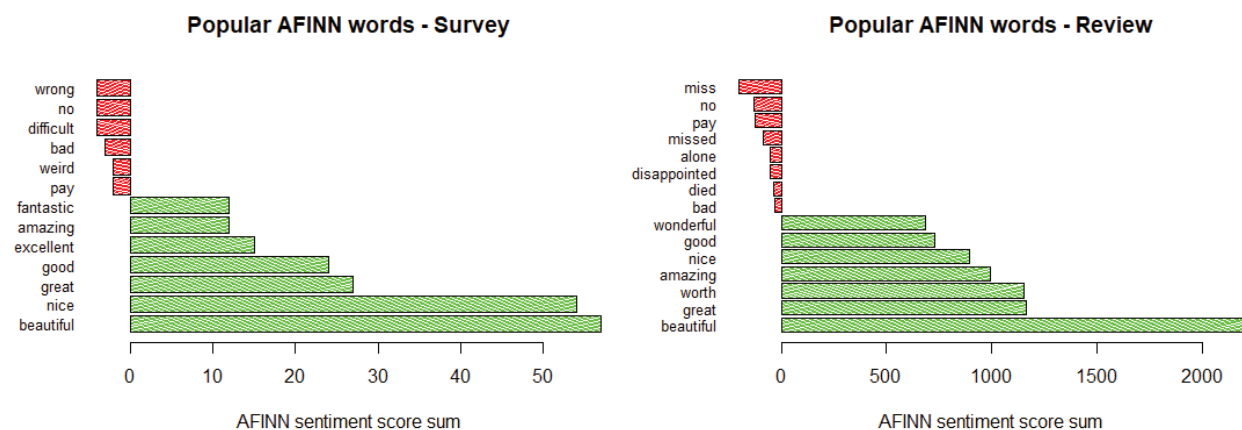

Figure 2. Frequently occuring words and related positive and negative sentiments in the survey and review texts applying the AFINN lexicon

REAL CORP 2020 Proceedings/Tagungsband
15-18 September 2020 – https://www.corp.at

ISBN 978-3-9504173-8-8 (CD), 978-3-9504173-9-5 (print)
Editors: Manfred SCHRENK, Vasily V. POPOVICH, Peter ZEILE, Pietro ELISEI,
Clemens BEYER, Judith RYSER, Christa REICHER, Canan ÇELIK

897

When comparing the survey and review datasets applying the AFINN lexicon, similar patterns appear. The positive sentiments strongly outnumber negative sentiments in both reviews and surveys. This result is significant considering the negative bias of the AFINN lexicon. Among the seven highest scoring positive words, beautiful, nice, great, good, and amazing are common to the review and survey datasets. Among the seven lowest scoring negative words, pay, bad, and no occur in both datasets. This result suggests similarities between the two datasets at the lexicon level as well as the sentiment level. The survey data, although less numerous in the number of texts as well as in average word length, contains a higher proportion of words that match a score of +3 or higher. The most frequent words, "beautiful", "great" and "nice" arescored according to the AFINN lexiconwith a +3. In a majority of the cases, a word such as 'nice' relates to the entire attraction (e.g. "nice place") or a part of the attraction ("nice garden" / "rooftop"). In some cases however, "nice" relates to a suggestion by the reviewer or survey respondent: e.g. "information in German would have been nice". The present bag-of-words approach is not able to filter the words for the context in which they are used. In the review texts, more words are present which are scored +2. Examples are the word 'worth', which is often used in the context of "worth a visit". Words with negative scores, such as 'miss' or 'missed' (score -2) are sometimes used in a negative context ("we missed part of the exhibition due to unclear signage or limited opening hours"), but also sometimes in a more positive way ("not to miss!").
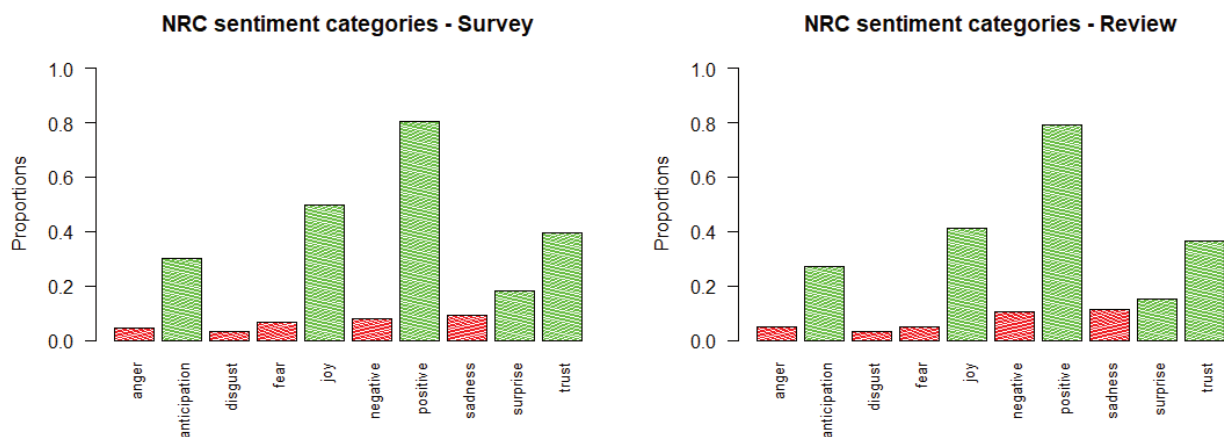

Figure 3. Distribution of emotions in survey and review texts applying the NRC lexicon

The survey and reviews were then analysed with the NRC lexicon which uncovers 7 emotions (Figure 3) and related words (Figure 4). In both Figure 3 and Figure 4, an interesting resemblance can be found between the emotions present in the survey and in the review datasets. Both positive and negative emotions are present in similar proportions. In both sets, positive emotions are related to 'joy', 'trust' and 'anticipation', which correspond to the presence of words such as 'beautiful'. The emotional state of 'joy' was linked to words such as 'information', while 'trust' corresponded with words such as 'guide'. Words such as 'church' or 'art' related to multiple emotions: 'joy', 'anticipation' and 'trust'.

Negative emotions are rather rare in the dataset. In the surveys, some texts refer to the availability of toilets. Some of the words ascribed to negative emotions appear to have a less negative connotation when cross-checked manually: references to the word 'cross' in the reviews relate to descriptions of paintings within some of the churches and museums, which can depict scenes related to suffering and martyrdom and reviewers use words matching negative emotions to describe them, whilst the same review describes these paintings as 'beautiful' as well. In general, negative comments relate to a lack of information, entrance fees and a lack of available toilets.

In the 14 topic LDA model, latent topics are described by the probability of words that describe the topic. Manuel analysis of the 10 most probable words therefore allows us to label and interpret these topics (Fig. 5). In the online review dataset, some topics relate to tickets and admission fees and whether these are worth to spend in relation to what the attraction offers (topic 1) or related to its striking appearance (topic 6). Topic 8 also mentions tickets, but this topic relates more to practical information on tickets (topic 8). Topic 14 relates to practical information considering visiting times and ongoing renovations. Other topics relate to the quality of the presented works of art or exhibitions (topic 2), artifacts (topic 9), the buildings themselves (topic 4), gothic architecture (topic 11), the direct surroundings (topic 7), views and vistas (topic 10), history

of daily live (of Ruebens, topic 12) and some other remember us not too miss certain places (topic 13 and 3). Topic 5 relates to the provided touristic information and (audio) guides. The fact that these topics can be interpreted well in the context is an indication that the word frequencies actually capture a diversity of themes running across the texts, as opposed to mere text artefacts.
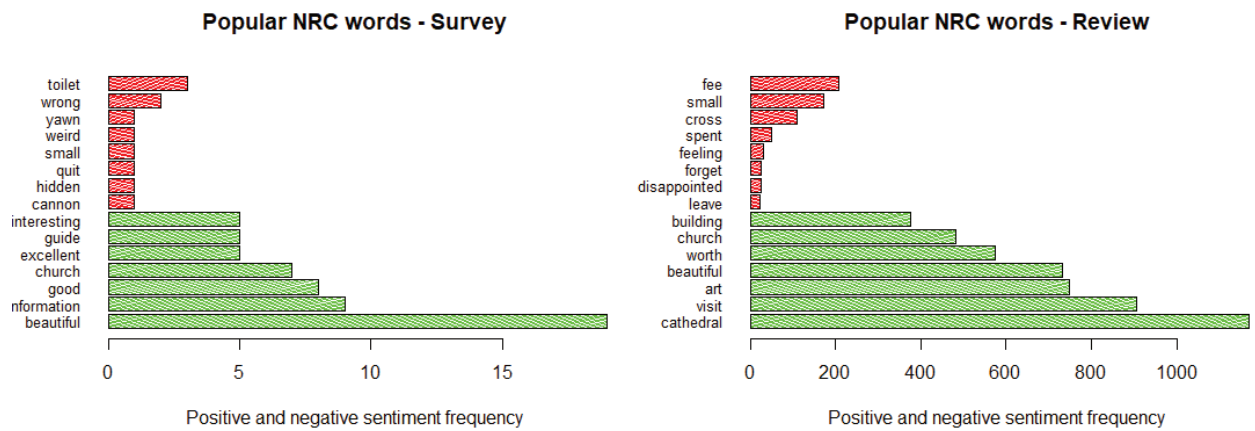

Figure 4. Frequently occuring words and related positive and negative sentiments in the survey and review texts applying the NRC lexicon
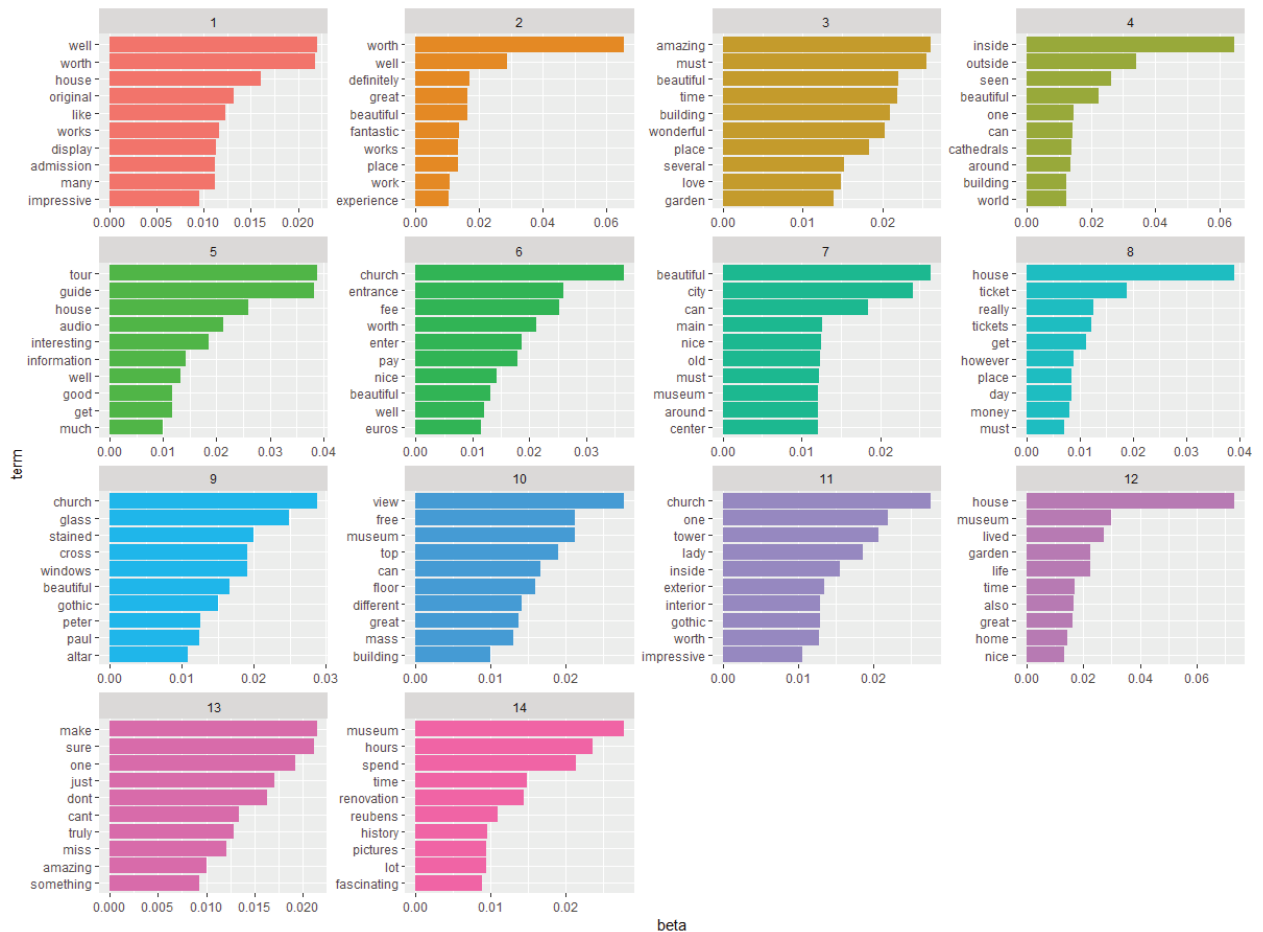

Figure 5. 14 topics derived from unsupervised topic LDA model

Furthermore, we also computed the average probability of each topic over all text documents, for both reviews and survey texts. This provides us with a way to directly compare the presence or dominance of the topics in the two corpora. In Fig.6, Topic 5, 12 and 14 stand out, while the other topics have quite similar levels of presence Ruben both datasets. Topic 5, which relates to provided information, (audio) guides and its added value, and topic 14, relating to opening hours, time to spend in the attraction and renovations, obviously have a higher probability of being discussed in reviews. This may be due to online reviews often

focusing on practical information about the usability of the touristic infrastructure. Correspondingly, topics 4 and 12, which both relate to describing the in- and exterior of attractions, have a higher probability to be discussed in review texts. Topic 5 has a the highest probable occurrence on average, followed by topic 12 and 6. This indicates that discussing the quality of provided information and audio guides (topic 5) is a relatively important theme, as well as discussing whether the aesthetics of a building makes up for charging an entrance fee (topic 6), or to what extent the attraction helps to give a vivid display of history (topic 12). In general, our analysis indicates that the topic distributions are very similar across both corpora. The small differences in the average need to be seen in the context of a high topic variance across all text documents, which lowers the significance of the differences in the mean. In the future, we plan to test these differences more systematically.
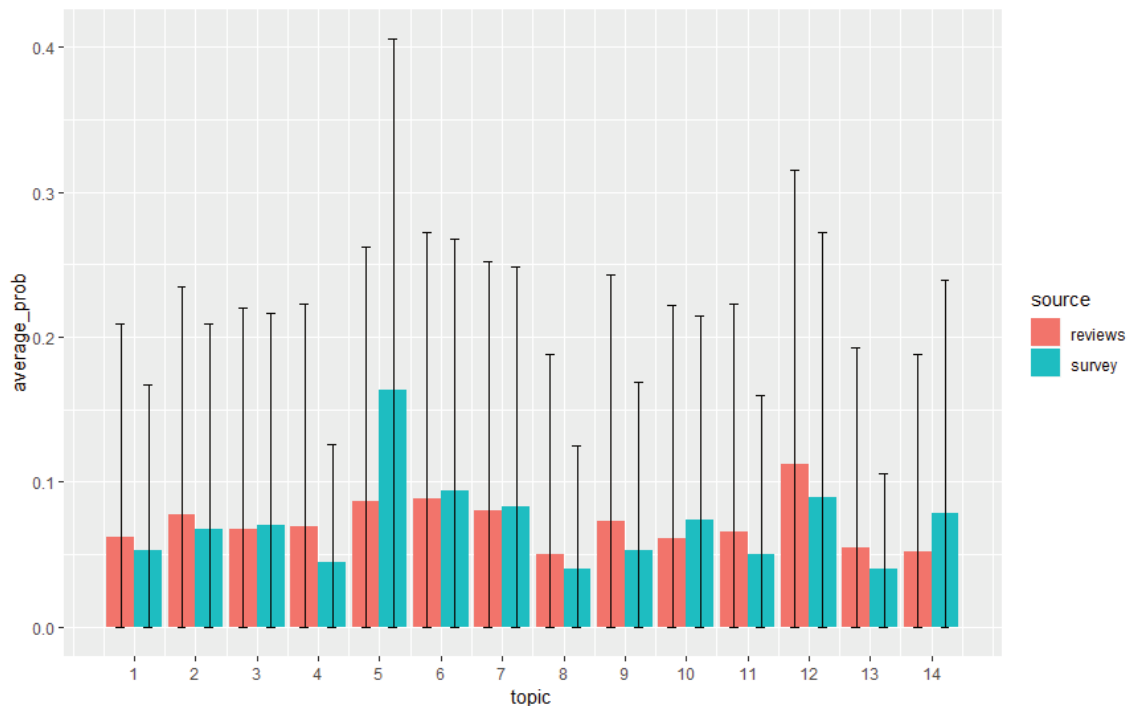

Figure 6. Probability of presence of LDA topics in review and survey dataset

5 CONCLUSIONS

This proof of concept paper on the opportunities provided by big data sentiment analysis as an alternative to on-site visitor surveys set out to answer two research questions: (a) to what extent web-scraped user-generated content could provide both internal (attraction-specific) and external (destination-wide) topics that might assist destination marketing and planning, and (b) to test the reliability of topics and sentiments of online reviews by comparing them with the results from more representative in-situ visitor surveys.

The answer to the second question is the most straightforward, the overlap in themes and emotions between the open-ended questions of in-situ surveys and online reviews was significant for both the AFINN and NRC lexicons. This is particularly interesting, since the make-up of survey respondents and online reviewers was very dissimilar. We might therefore hypothesize that visitors to tourism attractions have a rather fixed set of elements that are deemed important for evaluating a visit.

The first research question cannot be answered definitely. While the results of the LDA analysis do show promise in also uncovering extra-mural topics (e.g. Topics 4 and 7), both survey and TripAdvisor reviews remain predominantly – and quite logically – focused on the attraction being visited. In this sense, the LDA offers an interesting algorithm for gaining a deeper understanding of correlating topics, using the sentiment analysis as a stepping stone for a city-level reputational study would require additional sources, not in the least a combination of geo-localized data. A potential opportunity for further research would then be to run the sentiment analysis on an attraction-specific basis for multiple attractions across a city which can then, when combined, create a more in-depth analysis of the strengths and weaknesses of the broader service sector.

As a general summary, we might conclude that rather than seeing big data sentiment analysis as an alternative to traditional surveys, both methods can complement each other. Given the strong correlations between the open questions in the surveys and the online reviews, a higher efficiency might therefore be achieved by limiting on-site surveys to closed ended questions on visitor profiles, information sources used, and combined visits, elements which are difficult to uncover from online reviews. On the other hand, particularly the unstructured LDA supports a richer analysis of experiences, which is difficult to achieve via a traditional survey method.

6   REFERENCES

Alaei, A. R., Becken, S., & Stantic, B. (2019). Sentiment analysis in tourism: capitalizing on big data. Journal of Travel Research, 58(2), 175-191.

Ashworth, G., & Page, S.J. (2011). Urban tourism research: Recent progress and current paradoxes. Tourism Management, 32, 1-15.

Balagué, C., Martin-Fuentes, E., Gómez, M. J. (2016). Reliability of authenticated versus nonauthenticated hotel reviews: TripAdvisor and Booking.com case. Cuadernos de Turismo, 38, 513-516.

Blei, D. M. (2012). Probabilistic topic models. Communications of the ACM, 55(4), 77-84.

Blei, D. M. Ng, A. Y., & Jordan, I. (2003). Latent dirichlet allocation, Journal of machine Learning research 3 (2003) 993–1022.

Ceron, A., Curini, L., Iacus, S. M., & Porro, G. (2014). Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France. New media & society, 16(2), 340-358.

Chang, Y. C., Ku, C. H., & Chen, C. H. (2019). Social media analytics: Extracting and visualizing Hilton hotel ratings and reviews from TripAdvisor. International Journal of Information Management, 48, 263-279.

Christaller, W. (1963). Some Considerations of Tourism Location in Europe: The Peripheral Regions – Underdeveloped Countries – Recreation Areas. Papers of the Regional Science Association, 12, 95-105.

Duan, W., Yu, Y., Cao, Q., & Levy, S. (2016). Exploring the impact of social media on hotel service performance: A sentimental analysis approach. Cornell Hospitality Quarterly, 57(3), 282-296.

Garay Tamajón, L., & Cànoves Valiente, G. (2017). Barcelone seen through the eyes of TripAdvisor: actors, typologies and components of destination image in social media platforms. Current Issues in Tourism, 20(1), 33-37.

Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. Tourism Management, 59, 467-483.

Govers, R., & Go, F. M. (2004). Cultural identities constructed, imagined and experienced: A 3-gap tourism destination image model. Tourism (13327461), 52(2).

Judd, D. R., & Fainstein, S. S. (Eds.). (1999). The tourist city. New Haven: Yale University Press.

Kirilenko, A. P., Stepchenkova, S. O., Kim, H., & Li, X. (2018). Automated sentiment analysis in tourism: Comparison of approaches. Journal of Travel Research, 57(8), 1012-1025.

Kiritchenko, S., Zhu, X., Cherry, C., & Mohammad, S. (2014, August). NRC-Canada-2014: Detecting aspects and sentiment in customer reviews. In Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014) (pp. 437-442).

Lewalter, D., Phelan, S., Geyer, C., Specht, I., Grüninger, R., & Schnotz, W. (2015). Investigating visitor profiles as a valuable addition to museum research. International Journal of Science Education, Part B, 5(4), 357-374.

Martín-Santana, J. D., Beerli-Palacio, A., & Nazzareno, P. A. (2017). Antecedents and consequences of destination image gap. Annals of Tourism Research, 62, 13-25.

Mastercard. (2019). Global destination cities index 2019. Retrieved from https://newsroom.mastercard.com/wp-content/uploads/2019/09/GDCI-Global-Report-FINAL-1.pdf

Miossec, J.M. (1976). Un modèle de l'espace touristique. L'Espace Géographique, 6, 41-48.

Mohammad, S., Kiritchenko, S., & Zhu, X. (2013, June). NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets. In Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013) (pp. 321-327).

Mohammad, S. M., & Turney, P. D. (2010). Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text (pp. 26-34). Association for Computational Linguistics.

Munar, A. M., & Jacobsen, J. K. S. (2014). Motivations for sharing tourism experiences through social media. Tourism management, 43, 46-54.

Nguyen, T. H., & Shirai, K. (2015). Topic modeling based sentiment analysis on social media for stock market prediction. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers) (pp. 1354-1364).

Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages. Pp 93-98.

Oriade, A., & Robinson, P. (2018). Prosuming tourist information: Asking questions on TripAdvisor. International Journal of Tourism Research, 21(1), 134-143.

Pearce, D. G. (2001). An integrative framework for urban tourism research. Annals of tourism research, 28(4), 926-946.

Pearce, P. L., & Moscardo, G. (1985). Visitor evaluation: An appraisal of goals and techniques. Evaluation Review, 9(3), 281-306.

Presi, C., Saridakis, C., & Hartmans, S. (2014). User-generated content behaviour of the dissatisfied service customer. European Journal of Marketing, 48(9/10), 1600-1625.

Putri, I. R., & Kusumaningrum, R. (2017). Latent Dirichlet allocation (LDA) for sentiment analysis toward tourism review in Indonesia. In Journal of Physics: Conference Series (Vol. 801, No. 1, p. 012073). IOP Publishing.

Richards, G. (2018). Cultural tourism: A review of recent research and trends. Journal of Hospitality and Tourism Management, 36, 12-21.

Rossetti, M., Stella, F., & Zanker, M. (2016). Analyzing user reviews in tourism with topic models. Information Technology & Tourism, 16(1), 5-21.

Stella, K., & Mavragani, E. (2015). Assessing destination image: An online marketing approach and the case of TripAdvisor. Journal of Destination Marketing & Management, 4(3), 187-193.

Taecharungroj, V., & Mathayomchan, B. (2019). Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand. Tourism Management, 75, 550-568.

Urry, J. (1990). The tourist gaze. London: Sage.

Van der Borg, J., Costa, P., & Gotti, G. (1996). Tourism in European heritage cities. Annals of tourism research, 23(2), 306-321.

van der Zee, E., Bertocchi, D., & Vanneste D. (2020). Distribution of tourists within urban heritage destinations: a hot spot/cold spot analysis of TripAdvisor data as support for destination management. Current Issue in Tourism, 23(2), 175-196.

van Weerdenburg, D., Scheider, S., Adams, B., Spierings, B., & van der Zee, E. (2019). Where to go and what to do: Extracting leisure activity potentials from Web data on urban space. Computers, Environment and Urban Systems, 73, 143-156.

Whiting, A., & Williams, D. (2013). Why people use social media: a uses and gratifications approach. Qualitative Market Research: An International Journal.

Wong, C. U. I., & Qi, S. (2017). Tracking the evolution of a destination's image by text-mining online reviews – the case of Macau. Tourism Management Perspectives, 23, 19-29.

Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. Tourism Management, 58, 51-65.

Yokeno, N. (1968). La localisation de l' industrie touristique: application de l' analyse de Thunen-Weber. Cahiers de Tourism Série C, 9, 1-18.