Digital Narratives of Place

Learning About Neighborhood Sense of Place and Travel Through Online Responses

Ashok Sekar, Roger B. Chen, Adrian Cruzat, and Meiyappan Nagappan

As the market penetration of mobile information and communication technologies continues to grow, visitor feedback, such as online reviews of locations or sites visited, will continue to grow in parallel at finer temporal and geographic scales. This growth in data opens the opportunity for travel demand analysts to assess location attractiveness on the basis of online reviews and subsequently inform destination choice models. In geography and urban planning, the construct of sense of place (SOP) has emerged as an indicator for visitor association or connection with a place or site. An opportunity exists for examining SOP through the lens of text mining (i.e., extracting information from online text reviews and forming digital narratives of place). Several websites devoted to sharing feedback on experiences and overall perceptions exist, including Yelp and TripAdvisor. With text-mining methods, previously unidentified SOP-related topics and issues may emerge from online reviews and serve as a basis for subsequent analysis. The results from this study indicate that these emerging topics or terms require more contextual information and interpretation. As a stand-alone method, text mining is insufficient for identifying SOP topics, given the complexity of dimensions that characterize SOP. In addition, the results suggest that timing and seasonality play an important role in visitors' evaluation of a site; these factors have received less attention in the literature. With respect to text mining as a methodology to gain insights into SOP and supplement existing travel analysis, several barriers exist, including interpretation of topics from topic models. Nonetheless, these approaches are promising and require more research to guide practical implementation for inferring SOP from online text reviews and integration with existing travel analysis approaches.

Increasingly, transportation and urban planners have focused on fostering and building livable communities that benefit community well-being along several dimensions, including health and socialization. Livable communities critically require sense of place (SOP), which characterizes how humans interact with their natural and built environments and with each other, collectively. Locations with a strong SOP can facilitate lasting connections between visitors and the location. In addition, SOP has gradually entered several organizational decision-making levels, from local municipalities and neighborhoods (*1, 2*) to international discussions on ecology, the environment, and sustainability (*3*). SOP has also gained momentum in many other research fields, ranging from anthropology to environmental psychology. Applications include the planning and design of urban spaces (*4, 5*) and natural resource management (*6*). Furthermore, the United Nations Environment Programme has identified SOP as an essential feature of sustainable environments, including aspects of the surrounding ecosystem (*3*).

SOP has both human and physical dimensions (*7*). The human dimensions have been researched extensively and are often considered core to SOP (*5, 7, 8*). These human dimensions find their basis in attitude theory, which defines three distinct factors: affective, cognitive, and conative. Subsequently, researchers have characterized SOP along these three dimensions (*9*). Place attachment, the affective component, is defined as the positive bond developed between a person and his or her environment (*10*). Place dependence, the cognitive component, measures the perceived strength of association between a person and a place (*11*). Place identity, the conative component, represents the individual's identity in relation to the physical environment (*12, 13*). Other studies have identified additional influential aspects, such as place satisfaction, and social and architectural or aesthetic settings. Place satisfaction is the summary judgment of the perceived quality of a place or environment (*14*). The aesthetic and social settings are more loosely defined. Aesthetics includes views on architecture, the beauty of the place, the balance of decorative and functional attributes, artistic value, and peaceful and relaxing atmosphere. Social includes topics such as social atmosphere; level of crowdedness; amount of activity; safety; and level of friendliness to people (generally), kids, and family (*5, 15, 16*).

From an implementation and practitioner standpoint, urban design and natural resource management contexts have shown the strongest interest in SOP. Within urban design, SOP is considered a guiding principle for designing public spaces and built environments, to shape social contexts and foster social connections. From the perspective of the natural environment, SOP can also provide a framework for encouraging or strengthening commitment to and environmental stewardship of a given place, such as national parks, which is necessary for growth and maintenance (*17*).

Given that SOP explores the perceptual and psychological relationships between people and places, researchers are beginning to explore the applicability of SOP in travel behavior. First, SOP advances behavior models by adding a psychological element to choice process, which is usually modeled on the basis of economic realism. Researchers have explored the influence of some or all aspects of SOP as an explanatory variable for travel choice modeling. Zandvliet et al. (2006) studied place identity and its relation to destination choices in the Netherlands (*18*). A series of research papers from the University of California, Santa Barbara, explore many travel

A. Sekar and R. B. Chen, Golisano Institute for Sustainability, Rochester Institute of Technology, 111 Lomb Memorial Drive, Rochester, NY 14623-5608. A. Cruzat and M. Nagappan, B. Thomas Golisano College of Computing and Information Sciences, Rochester Institute of Technology, 20 Lomb Memorial Drive, Rochester, NY 14623-5604. Corresponding author: R. B. Chen, rbchen@gmail.com.

Transportation Research Record: Journal of the Transportation Research Board, No. 2666, 2017, pp. 10–18.
http://dx.doi.org/10.3141/2666-02

behavior facets of visitors (arrival time, mode, frequency, sequence of activities, companionship, and long-distance travel) and SOP of two malls in Santa Barbara, California (*5, 15, 19*).

Quantitative approaches toward SOP measurement are typically multidimensional and examine the strength that each SOP dimension associates with a particular location. Intercept surveys containing Likert scale attitudinal statements are typically used to measure these dimensions (*20*). Responses can estimate and measure the strength of each statement response toward each dimension. Factor analysis and structural equation modeling are common methodological approaches for relating SOP with other observed exogenous variables, such as trip frequency (*5, 8, 9, 19*). Researchers have also evaluated SOP with qualitative methods, such as visitor interviews and engaging community members with face-to-face conversation and photos of the location (*21, 22*). Despite its applicability in many areas, few guidelines or codes exist for designing SOP and evaluating its strength or presence. Approaches that provide more systematic evaluations about a location that also relate to the attitudes and behavior of people visiting a place may be helpful for both practitioners and researchers.

OPPORTUNITIES IN DATA MINING

The recent explosion in mobile information and communication technologies allows place and site visitors to share experiences and online text feedback or reviews over a more granular temporal and geographic scale that can inform SOP (*23–25*). An opportunity exists to examine SOP through the lens of data mining (i.e., extracting information and data online and forming digital narratives of place). In particular, text mining, which falls under the umbrella of data mining, is promising from the standpoint of systematically analyzing text on the basis of the usage and association, or clustering, of words and subsequent interpretation.

Text mining uncovers strong trends, topics, or both within textual data, such as online consumer text reviews. With respect to assessing a location's SOP, the application of collected online reviews is virtually nonexistent in the research and practitioner literature. Several applications of qualitative approaches exist, such as face-to-face interviews and case-specific interpretation by analysts. Dias et al. analyzed online reviews of vacation rentals in Portugal through a qualitative approach and identified broad themes that described the surrounding landscape and leisure activities affiliated with a place and recommendations for rental owners and visitors (*24*). In another study, Oz and Temizel qualitatively analyzed reviews from Turkey on the website Foursquare to identify parts of speech indicating place attachment (*26*). A qualitative approach is infeasible in the age of mobile information and communication technologies, with millions of visitors sharing massive quantities of text reviews on their experiences. Text mining serves as a feasible approach for analyzing this large volume of online text data, potentially revealing topics of concern related to SOP.

Intercept surveys conventionally used for SOP studies require money and time to survey visitors at sites of interest. Text mining can passively collect and analyze reviews across many geographic levels. In addition, text mining may reveal location-specific issues or topics related to the attractiveness of locations that are missed in intercept surveys. For example, text mining of reviews for a neighborhood may reveal a strong attractiveness to the local food served at food establishments, but a conventional SOP intercept survey that broadly addresses SOP may miss this issue.

In the field of computer science, approaches that uncover topics within text are called topic models. Topic models can potentially elicit themes, which in the case of this study are SOP dimensions (e.g., attachment or satisfaction), from online text reviews by visitors. Successful application of topic models in other fields includes inferring topics from academic journal websites and Wikipedia and finding patterns in genetic data (*27*). An output from topic models is a list of words or terms representing a topic. The list of words is formed on the basis of the frequency of occurrence in the corpus. A corpus is a collection of documents, which in this study is a collection of online text reviews. The analyst needs to make a qualitative judgment on the meaning of the topics identified in the topic models. For example, Blei implemented a topic model on 17,000 articles from the academic journal *Science* (*27*). One of the popular topics identified was "computer," on the basis of the following words output by the model: computer, models, information, data, computers, system, network, systems, model, and parallel.

Unlike topic models for academic journals and news articles, which have seen successful applications (*28, 29*), interpreting SOP topics from online reviews is challenging. Online reviews contain informal language; identifying SOP dimension from a list of words requires a deeper understanding of the words and their context. For example, an identified topic related to food can contain the words food, wings, love, beer, chicken. However, relating this topic to a SOP dimension is not straightforward and requires additional context. For example, this topic could be associated with satisfaction, in terms of food options available, but this association is not easily determined solely on the basis of the identified topic.

STUDY OBJECTIVES

Given the improvements in information and communication technology access, an opportunity exists to reexamine the concept of SOP from new forms of online data, especially narrative reviews and text feedback from site or place visitors. Although responses from conventional attitudinal surveys and qualitative approaches continue to inform practitioners of place making, analyzing online visitor responses may permit reconstructing a digital narrative of place and thereby reveal new dimensions of SOP through a new lens. Furthermore, understanding SOP can help inform travel demand models that seek to understand the destination choices of travelers. This work is part of a broader study to evaluate the SOP of neighborhoods in Rochester, New York, using different approaches, such as visitor intercept survey, data mining online reviews, and evaluation of architectural and design considerations.

The main objective of this paper is to analyze visitor feedback collected from online websites devoted to soliciting and posting text reviews. The two websites used were Yelp and TripAdvisor, although the analysis could be applied to any website with text reviews that are publicly accessible. Reviews for places and locations within neighborhoods, such as food establishments and parks, are analyzed through text mining to reveal new topics or dimensions of SOP that may contribute to or reinforce previously established SOP dimensions. Within this overarching objective, this study has the following broad goals:

1. Investigate SOP through the lens of text mining online text reviews for topics.
2. On the basis of these topics, identify areas of improvement for existing intercept survey tools, including topics or issues previously unidentified about SOP.

To accomplish these goals, the methodological approach in this study consists of three main components:

1. Collecting and processing online reviews from Yelp and TripAdvisor.
2. Performing a topic model to identify themes most prevalent across the reviews, with a particular focus on topics relating to SOP.
3. Comparing the identified topics with those found in the literature and existing sources on SOP, in addition to those used in an intercept survey in the broader study.

ONLINE DATA COLLECTION AND PROCESSING

This section briefly presents the approach taken to collect and process online reviews of neighborhoods from websites with visitor feedback for evaluating SOP. Broadly, the process involved generating a list of web addresses or uniform resource locators (URLs), each containing reviews about places in the neighborhoods of interest. A Java-based web scraper was developed to extract reviews and associated metadata, such as visitor ratings. Finally, the reviews were processed for further analysis.

Defining Neighborhoods of Interest

The neighborhoods chosen for the analysis are (*a*) College Town, (*b*) East End, and (*c*) Rochester Public Market. They were selected on the basis of their architectural relevance and popularity in Rochester. In addition, each neighborhood offers a contrast from the other two. College Town is a mixed-use development subneighborhood in the neighborhood of Upper Mount Hope near the University of Rochester. College Town is characterized by shopping, dining, working, hospitality, and upscale living, all within walking distance. The location attracts mostly college students from the University of Rochester, although visitors to the medical center are also frequent. The East End neighborhood is located in downtown Rochester, marked between East Avenue, Alexander Street, and Main Street. It is characterized by vibrant nightlife and cultural attractions. Rochester Public Market is a local farmers market that offers fresh produce, ethnic delicacies, specialty items, general merchandise, and more. An array of local businesses, including cafes, food stands, coffee shops, florists, specialty food purveyors, and breweries, can be found on the market grounds and in the surrounding Market District.

List of Attractions and Locations

Neighborhoods, by themselves, do not have an online presence where visitors provide reviews about the neighborhood specifically (e.g., travel sites such as Yelp and TripAdvisor). However, attractions in the neighborhoods, such as restaurants, hotels, markets, public parks, and other businesses, may have an online presence. Therefore, the authors propose to collect reviews from all the attractions in the neighborhoods with the assumption that the combined reviews contain information about the SOP of the neighborhood. The authors collected the list of attractions in College Town and the East End from their business association websites (*30, 31*). The name of a business or attraction from the business association's website was manually looked up by the authors in Yelp and TripAdvisor and the URLs were collected. Not all locations had a web presence. Yelp was a more popular platform than TripAdvisor, with more businesses listed, and had more user reviews than TripAdvisor. Because the website utility varies across reviewers, the analysis performed here was website dependent.

Through data mining, the authors collected 4,167 unique reviews across all three neighborhoods and both websites. For College Town, 769 online reviews were collected, with 236 from TripAdvisor and 533 from Yelp. For the East End, 3,180 reviews were collected, with 1,348 from TripAdvisor and 1,832 from Yelp. For Rochester Public Market, 218 reviews were collected, with 77 from TripAdvisor and 141 from Yelp. Rochester Public Market had only one attraction, the market itself. The East End had 96 attractions and College Town had 23 attractions, across both websites. The data used for analysis were extracted in March 2016 but span 1 year prior.

Data Collection

The data collection process involved scanning through the content in each website URL using the web scraper and extracting all the necessary information. Given the small scope (120 locations, ~4,000 reviews) as compared with studies that collect millions of data observations, a rudimentary web scraper was sufficient. Figure 1 captures the major tasks of the web scraper. The web scraper processes

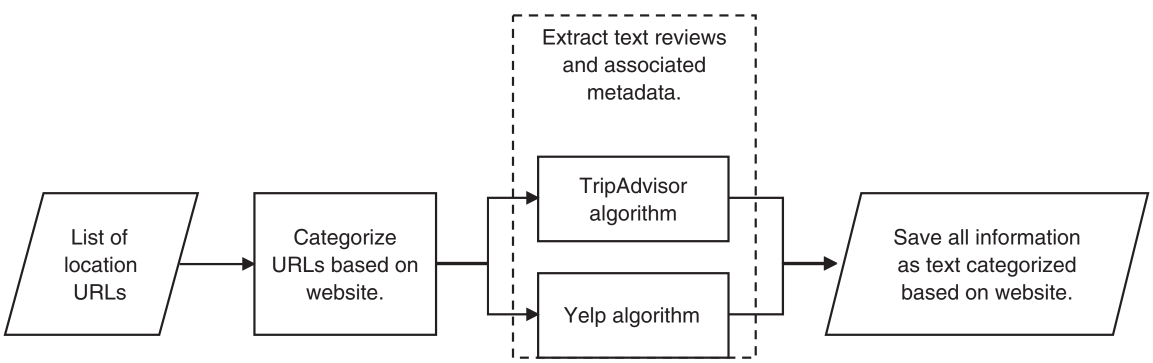

FIGURE 1 Data extraction process.

one URL (one business location) at a time and loops until the end of the list. (The scrapper is free at https://github.com/CruzatAdrian/OnlineReviewScrapper.) The second task segments the URLs into Yelp or TripAdvisor, which is necessary because each website has unique structure and information. In the third step, the relevant information is extracted. The scrapper functions by going through the Hypertext Markup Language (HTML) version of an URL, finding specific parts of the body text to be extracted (e.g., name of the reviewer, text review or user rating). Each URL can have multiple pages. When the scraper reaches the end of a page, it automatically navigates to the next page, and the process repeats for each URL. The implementation consists of functions that would look for specific HTML tags (using regular expression, or regex) for each field of the review (e.g., <UserName>) and then clean and save the data.

Data Processing

Unprocessed raw online reviews contain informal language, including spelling mistakes, atypical abbreviations, and slang. Therefore, it is necessary to process them for further analysis. Data processing removes unnecessary words or characters from the reviews, followed by some natural language processing steps such as stemming and stop word removal. Unnecessary words or characters in the reviews include HTML tags and special characters.

Stemming

The processes of stemming a word consist of eliminating any suffixes it may have to get its stem or root (e.g., transforming the words "consisting" and "consisted" into "consist"). It is a well-researched process that is complex but necessary in obtaining accurate results. Stemming allows for words that share a common sentiment to not compete but instead be treated as the same word. For example, the frequency of the words "enjoys" and "enjoyment," which share the root "enjoy," will be counted separately if not for stemming. For this project, the authors decided to use a third-party library, Snowball (*32, 33*), to handle the stemming. This tool generates stems that are more often consistent with the English language. Further details about stemming are available elsewhere (*32*). An important disadvantage of stemming is related to words with multiple meanings. Parking is a topic that is relevant to this analysis. The root form of parking is park which could also mean public parks or gardens used for recreation and also industrial parks. Therefore, care must be taken when the stemmed form of words is interpreted.

Stop Word Removal

Stop word removal is another important processing step. During this step, stop words, the words that have no meaning on their own or lack contextual significance, are removed from the corpus. The purpose of this step is to eliminate words from the corpus that could interfere or cloud the analysis. For topic modeling specifically, this step prevents terms such as "the," "you," or "another" from occurring in a topic. Stop words provide no insight on the possible topics discussed on the reviews. For this study, the authors used Terrier's stop word list as the primary source of stop words (*34*).

METHODOLOGICAL APPROACH

The processed reviews were analyzed by using two statistical techniques: (*a*) term frequency–inverse document frequency (TF-IDF) and (*b*) latent Dirichlet allocation (LDA), a topic modeling technique. TF-IDF provides a measure of the importance of the various terms in the corpus, whereas LDA is used to identify many clusters of words that co-occur in the corpus. The popular words and the topics identified are then interpreted qualitatively through the SOP framework. This section describes the statistical techniques and the interpretation approach thoroughly.

Term Frequency–Inverse Document Frequency

TF-IDF is a metric used to identify the importance of a term in a collection of documents or corpus, calculated as the product of term frequency and inverse document frequency. Term frequency is defined as the raw frequency of a term in a document, and inverse document frequency is the natural log of the ratio of number of documents in a corpus and number of documents in that corpus that contain a term (t). IDF ranks the importance of a term in the corpus. Equation 1 determines TF-IDF.

$$\text{TF-IDF} = f_{t,d} \cdot \ln\left(\frac{N}{n_t}\right) \tag{1}$$

where

$f_{t,d}$ = raw frequency of a term (t) in a document (d),
N = total number of documents, and
n_t = number of documents that contain the given term (t).

A Java application was also created for calculating the TF-IDF scores. The input to the application includes the text reviews and a list of stop words. The outputs include (*a*) the frequency of terms and TF-IDF values and (*b*) a corpus to be used for the topic modeling discussed in the next section. Each line in the corpus is considered a document and is made up of three space-separated fields: (*a*) the ID of the document, (*b*) the language of the document (hard coded as English for this project), and (*c*) stemmed terms (excluding stop words) from a given review.

Topic Modeling Using LDA

LDA (*35*) is a common statistical topic modeling approach that automatically indexes, searches, and clusters terms to form unstructured and unlabeled topics (*36*). In the data set, the documents are the text files that contain all the reviews about a neighborhood, and topics within the documents are generated through LDA. LDA accomplishes these tasks by first discovering a set of topics within the documents and then representing each document as a mixture of topics. In LDA, topics are explicitly created through a generative process, using machine learning algorithms to deduce the probability of terms present in each topic and the probability of a topic found in each document through an iterative process. More detail on the LDA process is available in the literature (*34, 35*).

The authors used MALLET, a popular third-party tool for implementing LDA (*36*). MALLET generates two outputs. The first output is a text file containing *k* topics, each with *t* terms. The number of topics *k* and the number of terms *t* in a topic depend on the specification. In this study, *k* ranged from 10 to 50 and *t* ranged from 5 to 20. The second output specifies the topic composition of the document.

Qualitative Interpretation of Results

The outputs from TF-IDF and LDA are a list of terms and a list of topics, respectively. A qualitative approach was used in this study to infer SOP dimensions from the terms and topics. To improve reproducibility of this work, the authors propose the following framework. First, a list of key terms and sentences that represent each SOP dimension was constructed using the previous SOP surveys found in the literature. Second, the authors associated the terms and topics generated through TF-IDF and LDA to SOP dimensions using the list generated in the previous step.

To generate this list, the authors relied on an intercept survey that was conducted in the same Rochester neighborhoods as part of the larger study beyond this paper. The intercept survey contained attitudinal statements for measuring SOP, taken from the literature. These statements provide linguistic cues for interpreting the online reviews. Table 1 provides the list of statements and example words for each SOP dimension.

The example words in Table 1 allow for a heuristic relationship building between the output from TF-IDF and LDA and SOP dimensions identified in the literature. For example, intercept survey statements on aesthetics contain words associated with the architecture or visually appealing design of the location. Statements on attachment relate to the connectedness of the visitors with the place. Therefore, terms in the survey that describe emotion (e.g., happy, sad) are categorized as attachment. Dependence is the comparison of the location in question with any other similar location concerning how the location meets the needs of the consumer. Identity captures the intended behavior of the person. Satisfaction refers to the visitors' content with the services and products offered. Therefore, any term that is related to the goods and services provided in a place could be categorized as satisfaction. Finally, social captures the friendliness of the location to family, friends, and kids and the nature of the people in the location. This list of words in Table 1 serves only as a guideline for inferring references to SOP dimensions from the outputs from TF-IDF and LDA.

As can be seen from the example words for identity, a single term is insufficient for inferring the SOP dimension. In those cases, a topic is interpreted with all the terms together. SOP is inferred from the terms or constructed phrases. Given more resources, an automated process is possible for associating statements from the survey, SOP dimensions, and representative terms.

TABLE 1 List of Terms for Each SOP Topic

SOP Dimension	Statements from Intercept Survey	Representative Terms or Phrases
Satisfaction	I am satisfied with the food options I am satisfied with the products offered I am satisfied with the parking space I am satisfied with the proximity of the parking space I am satisfied with the level of services I am satisfied with the entertainment options I am satisfied with the amount of people	Food, products, amenities, entertainment, people, parking, car, bicycle, walking, access, public, transit, satisfied, satisfy
Attachment	I feel a strong connection with the place It is a place that makes me feel relaxed I will be disappointed if it did not exist It makes me feel happy	Connection, disappointment, happy, attach
Dependence	It meets my need better than any location It has more diversity than any other place It has stores that lack specific items I want	Needs, diversity, missing, depend
Identity	It reflects the type of person I am It makes me feel too self-conscious It makes me feel comfortable since I identify with place It says very little about me It makes me feel I can be myself Is a good reflection of my identity I only come here when I have a specific reason in mind	Type of person, self-conscious, I can be myself, little about me, identify
Aesthetics	It has a visually appealing architecture It has a peaceful and relaxing atmosphere Has a good balance of decorative features and businesses It is a beautiful place Has artistic value	Beautiful, artistic, appealing, visually, architecture
Social/cultural	Has a definite social atmosphere Is a great family-friendly place to be Is a great kid-friendly place to be Has generally friendly people around Reflects the culture of Rochester Involves a risk of unpleasant encounters when traveling to it Is always overcrowded Has too much going on in it Makes me afraid to walk around	Social, family, friendly, culture, afraid, unpleasant, socialize

RESULTS AND SYNTHESIS

The corpus collected for each neighborhood and website was analyzed individually.

TF-IDF Results

First, the TF-IDF results are presented and discussed. Recall that TF-IDF provides the importance of each term in a list of terms from each corpus, with importance measured with the TF-IDF score (Equation 1). Table 2 lists the top 20 terms on the basis of the TF-IDF score for each neighborhood and website. A term is defined as a stemmed form of a word, as discussed in the section on data processing.

In Table 2, observations about terms frequently used in reviews are apparent. First, the most common terms across all neighborhoods and websites are "place," "time," and "food." In addition, between College Town and the East End, "beer" and "order" are common. The following is a list of terms for each location that were found on both websites:

- Rochester Public Market: vendor, park, local, produc, price, find, and shop.
- East End: beer, bar, order, little, friend, night, and service.
- College Town: beer, order, like, great, service, tabl, and nice.

In common terms between pairs of neighborhoods but specific to a website, the East End and College Town show more overlap in terms with each other than with Rochester Public Market. For Yelp, the terms "like," "time," "place," and "food" were common across all three neighborhoods; between College Town and the East End only, the following terms were also common: "beer," "order," "bar," "drink," "fri," and "service." For TripAdvisor, the terms "place," "time," and "food" were common across all three neighborhoods; between College Town and the East End these terms were also common: "room," "restaur," "beer," "service," "hotel," "order," "like," and "tabl." Overall, the results suggest a closer association between the College Town and East End neighborhoods relative to Rochester Public Market.

The terms provide insights into SOP. Terms such as "food," "drink," "service," "bar," and "hotel" are interpreted to represent location products or services. Accordingly, these are reasoned to be indicative of the satisfaction dimension of SOP because these terms are used to measure satisfaction or dissatisfaction in online reviews. However, these terms are notably missing from Table 1. This result suggests that although these are important terms of satisfaction found in online reviews, they were missing from the intercept survey, which was based on the literature. In Table 1, which is based on the intercept survey, the category of satisfaction did address satisfaction of bars, drinks, or hotels, which the TF-IDF of online reviews indicates as important.

Another set of prevalent terms contains words related to seasonality or time, such as "friday," "saturday," "morning," "night," and "summer." Interpretation from the analyst through reading reviews to gain context indicates that these terms may refer to the time of visit or some form of seasonality in activities. These terms do not represent any of the six SOP dimensions, according to the association in Table 1. However, seasonality could potentially be a new SOP

TABLE 2 Top 20 Terms from TF-IDF Number

Rochester Public Market				East End				College Town			
Yelp		TripAdvisor		Yelp		TripAdvisor		Yelp		TripAdvisor	
Vendor	92.46	Vendor	50.36	Beer	1,148.85	Room	734.46	Burger	483.93	Hotel	168.34
Park	73.57	Food	44.55	Bar	1,080.94	Great	679.23	Beer	352.44	Room	124.67
Love	70.72	Find	41.22	Order	1,044.34	Food	664.77	Order	305.13	Locat	120.26
Local	69.18	Saturday	40.13	Like	1,037.3	Place	637.58	Bar	304.25	Great	119.5
Peopl	69.11	Truck	38.71	Place	1,020.64	Restaur	606.63	Place	272.89	Beer	118.06
Like	68.92	Visit	36.66	Great	1,002.09	Beer	591.76	Fri	260.93	Nice	114.49
Produc	68.82	Best	36.43	Food	997.39	Time	559.34	Like	259.93	Stay	111.82
Time	67.06	Shop	36.41	Drink	993.02	Service	533.84	Food	257.61	Order	107.41
Price	66.93	Place	36.06	Time	944.48	Nice	533.41	Great	248.76	Restaur	107.23
Will	65.94	Great	35.99	Littl	870.56	Hotel	530.07	Time	245.96	New	105.68
Find	65.94	Price	35.43	Friend	858.86	Order	525.4	Back	226.94	Food	105.2
Shop	65.11	Produc	34.63	Tri	849.55	Friend	519.76	Friend	226.69	Like	98.22
Best	64.62	Day	34.16	Wing	834.09	Bar	503.23	Service	225.29	Custard	97.92
Place	64.3	Fresh	32.96	Night	826.8	Like	493.68	Cooki	221.66	Place	96.22
Chees	64.09	Local	32.66	Nice	804.97	Littl	490.93	Tabl	218.27	Tabl	92.58
Week	63.08	Farmer	32.25	Back	794.03	Night	489.02	Came	217.84	Staff	92.12
Buy	63.07	Offer	31.82	Sauc	778.38	Stay	488.19	Wait	215.16	Time	91.13
Food	62.41	Park	31.3	Service	769.92	Tabl	481.4	Drink	215.11	Service	89.6
Dont	60.28	Time	30.47	Menu	758.15	Dinner	479.33	Tri	212.52	Menu	87.8
Stand	59.9	Good	30.46	Fri	756.56	Love	466.87	Nice	210.86	Clean	86.84

dimension or an expansion of a previous dimension. For instance, a visitor could identify with a place only during a certain time or season annually.

A critical challenge with TF-IDF is that many terms, such as "friend," "like," "great," are associated with more than one SOP dimension. For example, the term "like" can convey dependence or satisfaction, depending on its usage and context. The statement "I like the food here" denotes satisfaction, whereas the statement "The quality of products here is like no other in Rochester" denotes dependence. Therefore, clean and clear-cut interpretations may require qualitatively looking at each review that contains the terms. Furthermore, one popular location with a significantly higher number of reviews compared with a second location in the neighborhood may bias the results. In such cases, the top terms from TF-IDF would likely represent that one location. However, the authors do not consider this bias as a shortcoming because one of the objectives is to identify aspects of a neighborhood that are of importance to respondents (visitors), and overrepresentation means the location is popular and therefore important for analysis. If these aspects differ from the context of an intercept survey, a case can be made for their inclusion.

LDA Topic Modeling Results

The challenge of interpretation present in TF-IDF may be resolved by using topic modeling. Using the LDA algorithm presented and discussed previously, 10 topics, each containing five terms, were generated. For the two websites and the three neighborhoods, 60 topics were generated. The topics provided more context and allowed the authors to establish a relationship with SOP. Because of difficulties in interpretation, topics with more than five terms were avoided. Table 3 shows a partial list of topics generated for Rochester Public Market, along with the authors' interpretation of the topics. The complete list is available at https://goo.gl/ufCvDs. Although other interpretations are certainly possible, a robust approach would be to collect the interpretation from many individuals.

Each topic generated can refer to multiple SOP dimensions or to none. For example, the first topic generated from the reviews for Rochester Public Market, "love food summer wegman truck," is interpreted by the authors as "I love food trucks during the summer" and "(Rochester Public Market) is better than Wegmans." This interpretation suggests LDA-generated topics contain two dimensions of SOP: satisfaction and dependence. The love for food trucks is interpreted as expressing satisfaction. Comparing Wegmans, a supermarket in the Rochester area, with Rochester Public Market indicates dependence. As a second contrasting example, consider the LDA-generated topic "food don't sure weekend like," which is difficult to clearly interpret. To assess the reference to SOP, an interpretation would be needed. Interpretations that cannot be associated with any SOP dimensions also exist. For example, the authors' interpretation of the topic "time park select home place" is a statement about the location without any SOP-associated dimensions.

Finally, looking at the collection of topics relative to SOP dimensions found in the literature, locations were characterized by a homogenous set of dimensions. For example, topics identified from East End and College Town were mainly about satisfaction of various products and services offered in the neighborhood. This lack of granularity of topics likely results from the motivation for many of the online reviews to provide feedback on a service or product. This motivation can also explain why almost none of the topics generated covered SOP dimensions such as aesthetics and social. To gain a broader set of topics, a future study might collect reviews from a broader set of businesses and services in the neighborhood for the analysis.

One theme missing from the current SOP intercept survey is time or seasonality consideration. Visitors typically travel to a site or neighborhood year-round. Some aspects of SOP that may work well in one season may not hold in other seasons. The food trucks at Rochester Public Market are an excellent example. Conceivably, these might help foster SOP, but because they occur only in the summer, there is an underlying seasonality dimension. In addition, factors explaining SOP might also vary by time of day. Evenings and weekends restrict the social atmosphere associated with the East End neighborhood. Thus, considering these timing and seasonality issues is important, as suggested by the text mining. Parking is the only transportation-related topic identified through this exercise, and reviews about parking always represent satisfaction.

Overall, topic modeling allows for easier interpretation and association of topics with dimensions, thereby making it a useful tool for SOP researchers using online reviews. However, there are caveats. Online reviews cannot completely replace intercept surveys because only certain SOP dimensions, such as satisfaction, are represented

TABLE 3 Sample of Topics, Interpretations, and Association with SOP Dimensions Generated Through LDA for Rochester Public Market

			Sense of Place Dimensions						
Website	Topics from RPM	Interpretation	AES	ATT	DEP	IDT	SAT	SOC	Missing and Interested Themes
Yelp	Love food summer Wegman truck	I love food truck during the summer. And RPM is better than Wegmans.			X		X		Seasonality or time
Yelp	Park great place fresh price	Great place to find fresh food at great prices, but parking is an issue.		X			X		Parking
Yelp	Saturday morn love week fresh	Go there on Saturday morning to get fresh produce for the week.			X				Seasonality or time
Yelp	Food dont sure weekend like	Cannot interpret							Seasonality or time
TripAdvisor	Time park select home place	Parking, time, place, and home							Parking, seasonality, or time

NOTE: RPM = Rochester Public Market; AES = aesthetics; ATT = attachment; DEP = dependence; IDT = identity; SAT = satisfaction; SOC = social.

well in the online reviews. Given that SOP is a multidimensional concept, information about all the dimensions is necessary to measure SOP.

The methodology presented here cannot measure the positive or negative aspect of any SOP dimension. For example, the term "love food" can be associated with satisfaction, but can also represent dissatisfaction when interpreted as "do not love." To overcome this shortcoming, reviews must be converted from a negative form to a more direct, affirmative form (e.g., from "I do not love the food here" to "I hate the food here").

CONCLUSION

As the adoption of mobile information and communication technologies increases, the potential for insight and volume of feedback provided by visitors to locations will grow continuously, in particular text feedback through the form of online reviews. Along with this growth, analyzing these collections of reviews or the corpus with respect to location attractiveness requires new methods, especially for integrating the outcomes with existing travel demand analysis approaches.

This study examined SOP through the lens of text mining. Specifically, topic modeling approaches, TF-IDF and LDA, were applied to online text reviews to identify the main topics of concern to visitors. These identified topics were compared with the key dimensions of SOP found through the literature and implemented in intercept surveys used to infer the degree of SOP associated with a location.

The exploratory analysis revealed that as a stand-alone analysis, text mining, using topic modeling, requires supplemental domain expertise to interpret the outcomes appropriately. For a sentiment analysis, the task of interpreting favorable versus unfavorable reviews is easier relative to the context of interpreting SOP dimensions, which require context. The study results indicate that seasonality and timing of activities are particularly important to visitors. In addition, "food" as a means of describing the attractiveness of a place is currently inadequately captured in SOP dimensions defined through the literature and carried over into intercept surveys.

Future work includes automating the interpretation process described in this paper, possibly using Amazon's Mechanical Turk. Future work also includes a more explicit consideration of integrating the identified topics from the topic models into existing travel demand models and discrete choice models.

ACKNOWLEDGMENT

The authors thank the University Transportation Research Center at the City College of New York for funding a portion of this research work.

REFERENCES

1. Soini, K., H. Vaarala, and E. Pouta. Residents' Sense of Place and Landscape Perceptions at the Rural–Urban Interface. *Landscape and Urban Planning,* Vol. 104, No. 1, 2012, pp. 124–134. https://doi.org/10.1016/j.landurbplan.2011.10.002.
2. Tester, G., E. Ruel, A. Anderson, D.C. Reitzes, and D. Oakley. Sense of Place Among Atlanta Public Housing Residents. *Journal of Urban Health: Bulletin of the New York Academy of Medicine,* Vol. 88, No. 3, 2011, pp. 436–453. https://doi.org/10.1007/s11524-011-9579-0.
3. Newman, P., and I. Jennings. *Cities as Sustainable Ecosystems: Principles and Practices.* Island Press, Washington, D.C., 2012.
4. Billig, M. Sense of Place in the Neighborhood, in Locations of Urban Revitalization. *GeoJournal,* Vol. 64, No. 2, 2005, pp. 117–130. https://doi.org/10.1007/s10708-005-4094-z.
5. Deutsch, K., S.Y. Yoon, and K. Goulias. Modeling Travel Behavior and Sense of Place Using a Structural Equation Model. *Journal of Transport Geography,* Vol. 28, 2013, pp. 155–163. https://doi.org/10.1016/j.jtrangeo.2012.12.001.
6. Brown, G., and C. Raymond. The Relationship Between Place Attachment and Landscape Values: Toward Mapping Place Attachment. *Applied Geography* (Sevenoaks, England), Vol. 27, No. 2, 2007, pp. 89–111. https://doi.org/10.1016/j.apgeog.2006.11.002.
7. Stedman, R.C. Is It Really Just a Social Construction? The Contribution of the Physical Environment to Sense of Place. *Society and Natural Resources,* Vol. 16, No. 8, 2003, pp. 671–685. https://doi.org/10.1080/08941920309189.
8. Tapsuwan, S., Z. Leviston, and D. Tucker. Community Values and Attitudes Towards Land Use on the Gnangara Groundwater System: A Sense of Place Study in Perth, Western Australia. *Landscape and Urban Planning,* Vol. 100, No. 1–2, 2011, pp. 24–34. https://doi.org/10.1016/j.landurbplan.2010.09.006.
9. Jorgensen, B.S., and R.C. Stedman. Sense of Place as an Attitude: Lakeshore Owners Attitudes Toward Their Properties. *Journal of Environmental Psychology,* Vol. 21, No. 3, 2001, pp. 233–248. https://doi.org/10.1006/jevp.2001.0226.
10. Low, S.M., and I. Altman. Place Attachment. In *Place Attachment* (I. Altman and S.M. Low, eds.), Plenum Press, New York, 1992, pp. 1–12.
11. Stokols, D., and S.A. Shumaker. The Psychological Context of Residential Mobility and Well-Being. *Journal of Social Issues,* Vol. 38, No. 3, 1982, pp. 149–171. https://doi.org/10.1111/j.1540-4560.1982.tb01776.x.
12. Proshansky, H.M., A.K. Fabian, and R. Kaminoff. Place-Identity: Physical World Socialization of the Self. *Journal of Environmental Psychology,* Vol. 3, No. 1, 1983, pp. 57–83. https://doi.org/10.1016/S0272-4944(83)80021-8.
13. Proshansky, H.M. The City and Self-Identity. *Environment and Behavior,* Vol. 10, No. 2, 1978, pp. 147–169. https://doi.org/10.1177/0013916578102002.
14. Mesch, G.S., and O. Manor. Social Ties, Environmental Perception, and Local Attachment. *Environment and Behavior,* Vol. 30, No. 4, 1998, pp. 504–519. https://doi.org/10.1177/001391659803000405.
15. Deutsch, K., and K. Goulias. Exploring Sense-of-Place Attitudes as Indicators of Travel Behavior. *Transportation Research Record: Journal of the Transportation Research Board,* No. 2157, 2010, pp. 95–102. https://doi.org/10.3141/2157-12.
16. Deutsch, K.E., and K.G. Goulias. *Understanding Places Using a Mixed Method Approach.* University of California Transportation Center, Berkeley, 2011.
17. Williams, D.R., and S.I. Stewart. Sense of Place: An Elusive Concept That Is Finding a Home in Ecosystem Management. *Journal of Forestry,* Vol. 96, No. 5, 1998, pp. 18–23.
18. Zandvliet, R., M. Dijst, and L. Bertolini. Destination Choice and the Identity of Places: A Disaggregated Analysis for Different Types of Visitor Population Environment in the Netherlands. *Journal of Transport Geography,* Vol. 14, No. 6, 2006, pp. 451–462. https://doi.org/10.1016/j.jtrangeo.2005.10.009.
19. Lee, J.H., A.W. Davis, and K.G. Goulias. Exploratory Analysis of Relationships Among Long-Distance Travel, Sense of Place, and Subjective Well-Being of College Students. Presented at 94th Annual Meeting of the Transportation Research Board, Washington, D.C., 2015.
20. Stedman, R.C. Sense of Place and Forest Science: Toward a Program of Quantitative Research. *Forest Science,* Vol. 49, No. 6, 2003, pp. 822–829.
21. Kyle, G., and G. Chick. The Social Construction of a Sense of Place. *Leisure Sciences,* Vol. 29, No. 3, 2007, pp. 209–225. https://doi.org/10.1080/01490400701257922.
22. Stedman, R., T. Beckley, S. Wallace, and M. Ambard. A Picture and 1000 Words: Using Resident-Employed Photography to Understand Attachment to High Amenity Places. *Journal of Leisure Research,* Vol. 36, No. 4, 2004, pp. 580–606.

23. Humphreys, L., and T. Liao. Foursquare and the Parochialization of Public Space. *First Monday,* Vol. 18, No. 11, 2013. https://doi.org/10.5210/fm.v18i11.4966.
24. Dias, J.A., F. Perdigão Ribeiro, and A. Correia. Online Reviews of Short-Term Visits: Exploring Sense of Place. *International Journal of Culture, Tourism and Hospitality Research,* Vol. 7, No. 4, 2013, pp. 364–374. https://doi.org/10.1108/IJCTHR-02-2012-0006.
25. Schwartz, R. *Online Place Attachment: Exploring Technological Ties to Physical Places. Mobility and Locative Media: Mobile Communication in Hybrid Spaces.* Routledge, New York, 2015, pp. 85–95.
26. Oz, B.K., and T.T. Temizel. On Inference of Sense of Place from Geo-Social Networks. Presented at 9th International AAAI Conference on Web and Social Media, Oxford, England, 2015.
27. Blei, D.M. Probabilistic Topic Models. *Communications of the ACM,* Vol. 55, No. 4, 2012, pp. 77–84. https://doi.org/10.1145/2133806.2133826.
28. Zhao, W.X., J. Jiang, J. Weng, J. He, E.-P. Lim, H. Yan, and X. Li. Comparing Twitter and Traditional Media Using Topic Models. In *Advances in Information Retrieval: European Conference on Information Retrieval 2011* (P. Clough et al., eds.), *Lecture Notes in Computer Science,* Vol. 6611, Springer, Berlin, 2011, pp. 338–349. https://doi.org/10.1007/978-3-642-20161-5_34.
29. Wang, C., and D.M. Blei. Collaborative Topic Modeling for Recommending Scientific Articles. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* Association for Computing Machinery, New York, 2011, pp. 448–456. https://doi.org/10.1145/2020408.2020480.
30. College Town Rochester. http://www.collegetownrochester.com/. Accessed June 7, 2016.
31. Rochester's East End. http://rochesterseastend.com/. Accessed May 13, 2016.
32. Porter, M.F. *Snowball: A Language for Stemming Algorithms.* 2001.
33. Snowball. http://snowballstem.org/. Accessed June 18, 2016.
34. Ganesan, K. Text-Mining-Resources. https://bitbucket.org/kganes2/text-mining-resources/downloads. Accessed June 19, 2016.
35. Blei, D.M., A.Y. Ng, and M.I. Jordan. Latent Dirichlet Allocation. *Journal of Machine Learning Research,* Vol. 3, Jan. 2003, pp. 993–1022.
36. Blei, D.M., and J.D. Lafferty. Topic Models. In *Text Mining: Classification, Clustering, and Applications,* CRC Press, Boca Raton, Fla., 2009, pp. 71–94.

The Standing Committee on Effects of Information and Communication Technologies on Travel Choices peer-reviewed this paper.