Hotel Rating Prediction System Based on Time Factors:
Using Reviews and Sentiment Analysis

Pei-Hua Lee, China Medical University Hospital, Taiwan

Yu-Kai Sun, National Taitung Junior College, Taiwan

Yin-Pei Ke, National Chung Cheng University, Taiwan

Pei-Ju Lee, National Chung Hsing University, Taiwan*

ABSTRACT

While the internet provides abundant information, it often leads to information overload of users when purchasing goods. Tripadvisor.com, despite having a date sorting function, struggles to effectively filter relevant comments to users and neglects that consumer preferences may change over time. Therefore, this study aims to develop a website with visual charts showing changes in sentiment over time in reviews. The goal is to determine if this website improves user efficiency compared to the original website, reducing search time and aiding decision-making. The chart generation process involves four stages: collecting and preprocessing comments, constructing a hotel feature dictionary, classifying sentences and computing sentiment scores, and embedding charts on the website. 36 Tripadvisor.com users participate in experiments to evaluate the impact of old and new interfaces on answer quantity and search time. The NASA.tlx scale is used to assess the mental load experienced with both interfaces.

KEYWORDS

Feature Classification, Hotel Preference, Sentiment Analysis, Social Review, Visualization

INTRODUCTION

User-generated content (UGC) is vital for consumers, retailers, and managers, as customer opinions impact retail business significantly (Giachanou & Crestani, 2016). Online reviews have become essential for consumers when assessing the quality of products such as hotels and restaurants before making purchase decisions (Archak et al., 2011; Zhu & Zhang, 2010). Consumers rely on these reviews to learn from others' experiences and evaluate product quality (Forman et al., 2008; Kim et al., 2006; Mudambi & Schuff, 2010).

Reviews play a significant role in consumers' decision-making processes. Consumers rely on the polarity of reviews to assess product or service quality, aiding informed purchasing decisions (Pai et al., 2013). In a survey, 86% of respondents stated that reviews significantly influence their purchase

DOI: 10.4018/JOEUC.342129 *Corresponding Author

decisions (PowerReviews, n.d.). Review sentiment is considered the second most crucial factor in evaluating consumer reviews (Paget, n.d.). By analyzing online reviews, merchants can understand consumer preferences, improve product quality, and cater to consumer needs (Tang et al., 2014).

However, the abundance of information in reviews can be overwhelming for consumers. To address information overload, three approaches are proposed: visualization, which involves summarizing important information in a graphical format; document summarization, which entails presenting key details; and review ratings (Chang et al., 2017; Lee et al., 2016; Falschlunger et al., 2016; Banerjee & Chua, 2016). Visualization techniques, such as graphic visualization, help reduce information overload and enhance decision-making by presenting complex data in a clear manner (Daniel et al., 2010; Falschlunger et al., 2016). Visualizations capitalize on the *picture advantage effect*, as people find pictures easier to understand than words or numbers (Paivio & Csapo, 1973; Kelleher & Wagener, 2011).

In addition, consumer preferences can change over time, making it crucial to explore review text and incorporate a time axis for calculating rating scores (Chang et al., 2017). Few studies have examined temporal factors and sentiment analysis, primarily focusing on product characteristics (Li et al., 2015). Charts can efficiently convey product positioning, aiding consumers in their decision-making process and helping companies develop new products (Lee et al., 2016). Timeline-based charts and charts in different orientations provide a quick overview of current hotel experience trends (Chang et al., 2017).

This study aims to develop a website with visual charts showing changes in sentiment over time in reviews. The goal is to determine if this website improves user efficiency compared to the original website, reducing search time and aiding decision-making. In particular, we try to answer the following hypotheses:

1. The accuracy of using the new interface will be higher than the accuracy of using the old interface.
2. The response time using the new interface will be less than the response time using the old interface.
3. The time to find the correct answer and fill in the user answer varies for different facing interfaces.
4. The order in which the interfaces were tested differed in the time it took for users to find the correct answer and user answer.

Historical data from TripAdvisor.com for 10 hotels was selected, and a dictionary of hotel characteristics was constructed. Review sentences were categorized based on these characteristics, and sentiment scores were calculated using the MPQA Corpus subjective dictionary. The study recruited 28 participants to evaluate the new interface's accuracy, response time, and user experience compared to the original website.

LITERATURE REVIEW

Data Visualization Studies

Visual analytics offer an effective way to analyze data and support complex decision-making and data search (Sacha et al., 2014). Several studies have demonstrated that visualization aids recognition of complex information and is easier to understand than textual content (Kelleher & Wagener, 2011). However, the type of visualization used depends on the purpose; for example, long bar charts are suitable for comparing values, while line charts are useful for identifying trends in data (Benbasat & Tan, 1990).

Time-Series Visualization of Online Reviews

Visualization techniques can be used to observe data changes over time in hotel ratings and sentiment orientation. Chang et al. (2017) used Tableau to present information visually and made use of different types of visual analysis, such as timeline analysis, location analysis, and pie charts. Consumer reviews provide key information for other consumers and operators, and charts can show the positioning of a product or service very directly and thus help people to understand the performance of the product, as well as provide users with the information they need to make decisions. Graphing can be used to analyze consumer reviews on social networks (for example, forums or Twitter).

This study used a time-series analysis in line graph format to calculate traveler rating scores for orientation to segmentation and to compare each month's ratings. Li et al. (2015) incorporated time as a factor to present reports on trends in hotel characteristics by year.

Theme Classification Techniques

Hu and Liu (2004) considered product features as nouns and noun phrases, used part-of-speech tagging to capture nouns and noun phrases in sentences, and applied association rules to treat frequently-occurring nouns as product features. Lee et al. (2016) considered the co-occurrence of words and constructed a virtual file using the concept of distributional similarity to retrieve nouns above a threshold value, calculate the correlation between two words using the Jaccard coefficient, remove irrelevant words, and calculate word frequency to group similar product features. After word frequency is calculated, similar product features are grouped into the same product feature set. In some studies, above-threshold word frequencies were considered as product features (Li et al., 2015; Zhan et al., 2009).

The *orientation* of a review refers to its theme, and Bagheri et al. (2013) concluded that that orientation analysis is critical: if we do not know the theme of a review, then the applicability of sentences or opinions within the review will be limited. Rhee and Yang (2015) compared six hotel review themes in their orientation analysis, namely location, sleep quality, service, value, cleanliness, and room, and they found that each orientation has a different level of importance to different consumers, so it is important to understand review information in terms of each orientation.

Latent Dirichlet allocation (LDA), a topic modeling approach, applies a probabilistic model to find semantic topics in a text collection (Blei et al., 2003). For text classification, Sinoara et al. (2019) show that LDA is more effective for low-dimensional than for high-dimensional space classification. Word2Vec is an artificial neural network prediction model employing the continuous Skip-gram algorithm as well as continuous bag-of-words (CBOW) algorithm framework, which represents single words in a vocabulary as multidimensional vectors and uses a large amount of unlabeled data for training (Mikolov et al., 2013). The Skip-gram input word tags determine the surrounding words, while the CBOW input surrounding words, predict the word tags, and capture thematic similarities between words.

Sentiment Analysis–Related Research

Sentiment analysis, also known as opinion mining, is used in natural language processing to process and analyze text. For polarity, Hu and Liu (2004) used WordNet to capture opinion words (for example, "great" or "amazing"), and they used those words to decide the opinion direction of each sentence, classifying the sentences as expressing positive or negative sentiment. Many studies have been conducted to determine the emotional polarity of features by performing opinion orientation identification following feature capture (Bai et al., 2005; Liu, 2010; Lloret et al., 2015; Palakvangsa-Na-Ayudhya et al., 2011; Shah et al., 2016; Tang et al., 2014; Zhan et al., 2009).

A typical review comment contains both subjectivity and objectivity. Subjectivity reflects the consumer's feelings and sensations after using a product or experiencing a service, while objectivity usually reflects aspects such as product specifications and price. A review that has subjective and objective content will be more informative and beneficial to consumers (Ghose & Ipeirotis, 2011; Liu et al., 2013; O'Mahony & Smyth, 2010; Zhan et al., 2009)

The more commonly used open-source sentiment dictionaries in academia include: MPQA Subjectivity Cues Lexicon, which provides 8222 subjective words, labeling each word with its lexical nature, affective polarity (positive, negative, neutral), and intensity (strong, weak) (Riloff & Wiebe, 2003; Wilson et al., 2005); SentiWordNet, a sentiment dictionary developed for WordNet, which indicates whether the polarity of each sentiment word is positive, negative, or neutral (Baccianella et al., 2010); and General Inquirer (GI), which has 1914 positive words and 2293 negative words and labels each word with its emotional polarity, intensity, and lexicality (Stone & Hunt, 1963).

Review Helpfulness

Review helpfulness refers to the number of user votes that express positive feedback, and it represents the subjective responses of consumers after reading a review (Cao et al., 2011; Ghose & Ipeirotis, 2011; Martin & Pu, 2014). Many e-commerce sites ask visitors "Was this review helpful?" after each review to obtain user feedback, and they often present the final results on the webpage via statements such as "30 out of 40 people found this review helpful" for other users' reference. Helpful reviews are read by other consumers in greater numbers and increase the efficiency of the review-reading process, since consumers perceive helpful reviews as highly reliable (Cao et al., 2011). Hwang et al. (2014) analyzed and predicted the usefulness of reviews by considering three categorical characteristics: review content, review sentiment, and review quality. They found that all three characteristics were important predictors and were considered to have the greatest impact on review usefulness. Previous studies have found that reviewer information or reputation influences consumers' final purchase decisions (Forman et al., 2008), so this study identifies reviewer information as an important predictor of review usefulness.

RESEARCH METHODOLOGY

This study proposes a method to visualize the emotional content of comments by incorporating time as a factor, considering the hotel features, and using an emotion dictionary to calculate emotion scores and map their trends for hotel features. The structure of this research method is shown in Figure 1.The first step addresses the comment collection and pre-processing, including spell-checking, root reduction, and word marking; the second step is the comment feature selection stage, in which the hotel features are selected and classified for the comment sentences; in the third step, the hotel sentiment score calculation stage, the sentences are first classified by year and then have their sentiment scores calculated for each year; the final step is to generate graphs of hotel sentiment.

Review Collection and Pre-Processing of Data

We used a Python crawler to collect data from online reviews of hotels provided by TripAdvisor.com. We collected 16,367 reviews of 10 hotels to use as the data set for this study. As shown in Figure 2, the information contained in a review includes (a) the review title; (b) the text of the review; (c) the review date; (d) the rating score; (e) the number of "helpful" votes received by the review, and so on.

When users write comments, they often pay little attention to correct spelling; for example, "[t] he quality used on everything is the best" contains a mis-keyed spelling. Therefore, this study uses Google Spell Check to check the spelling of each collected comment and corrects any spelling errors before pre-processing the text.

This study used the Stanford CoreNLP tool developed by the Stanford Natural Language Processing team to perform word processing (Manning et al., 2014). Pre–word processing is divided into three steps: word segmentation, stemming, and parts-of-speech (POS) tagging. Word segmentation is the process of separating each comment with a period(.), exclamation mark (!), or question mark (?). Stemming, or root restoration, restores the words to their original form by grammatical rules; the lexical or POS tagging identifies the lexical nature of each word as noun (NN), adjective (JJ), adverb (RB), and so on. In the example in Table 1, a comment was divided into two sentences by the above method, and the final result was obtained after sentence breaking, root reduction, and lexical tagging.

Figure 1. Research Framework

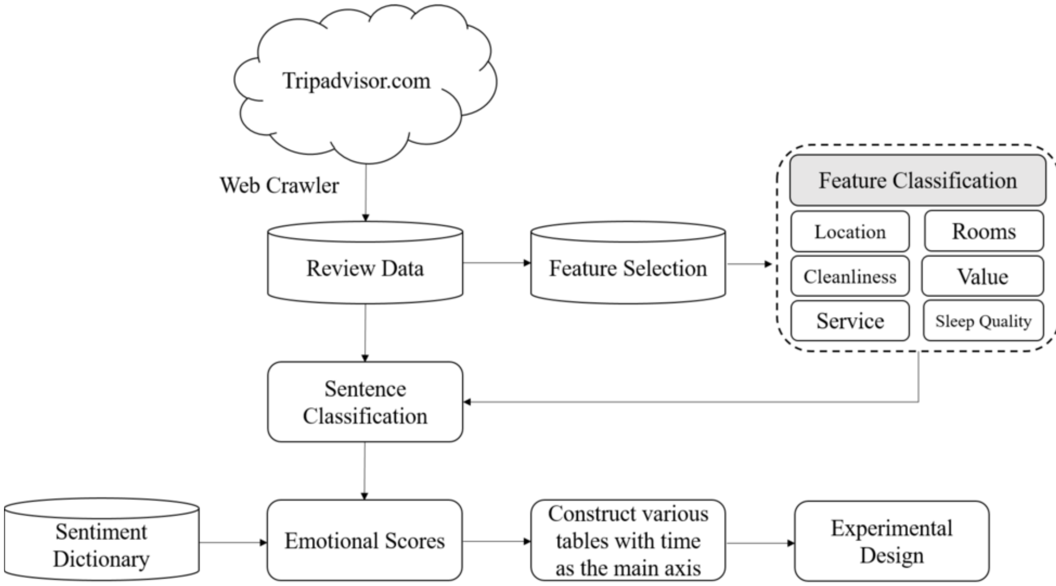


Figure 2. TripAdvisor.com Review Example

Selection of Characteristics

After the TripAdvisor.com reviews were processed, the nouns in the documents were characterized and categorized. We then used Word2Vec to train the text to find semantic similarities among the words, classifying them into six types of sentences as shown in Figure 3.

Table 1. Example of Comment Text Pre-Processing

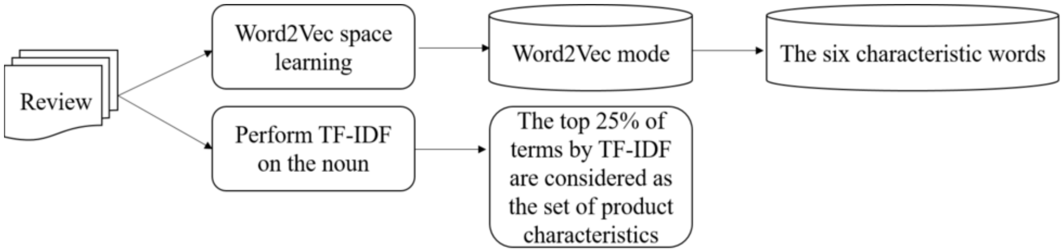
Reviews	We didn't expect this level of pampering, the quality of the food. Always present, yet very unobtrusive service staff.				
	ID	Word	Lemma	Position	POS
Sentence #1					
	1	We	we	0-2	PRP
	2	did	do	3-6	VBD
	3	n't	not	6-9	RB
	4	expect	expect	10-16	VB
	5	this	this	17-21	DT
	6	level	level	22-27	NN
	7	of	of	28-30	IN
	8	pampering	pampering	31-40	NN
	9	,	,	40-41	,
	10	the	the	42-45	DT
	11	quality	quality	46-53	NN
	12	of	of	54-56	IN
	13	the	the	57-60	DT
	14	food	food	61-65	NN
	15	.	.	65-66	.
Sentence #2					
	1	Always	always	67-73	RB
	2	present	present	74-81	JJ
	3	,	,	81-82	,
	4	yet	yet	83-86	RB
	5	very	ver	87-91	RB
	6	unobtrusive	unobtrusive	92-103	JJ
	7	service	service	104-111	NN
	8	staff	staff	112-117	NN
	9	.	.	117-118	.

Note. Source: Compiled by this study.

Selected Hotel Features

Term frequency (TF) refers to the frequency with which a word appears in a particular review, while inverse-document frequency (IDF) refers to the total number of reviews in which the word appeared. Most previous studies have constructed feature words by considering nouns and noun phrases. In this study, the nouns were retrieved from all reviews and the frequencies calculated by Equation 1 to find the most representative nouns, taking into account the studies of Li et al. (2015) and Zhan et al. (2009).

Figure 3. Constructing a Dictionary



The TF-IDF of a term is defined by the following equation:

$$tfidf\left(t_j\right) = \sum_{i=1}^{|R|} tf\left(r_i, t_j\right) \times \log \frac{|R|}{df\left(t_j\right)} tf\left(r_i, t_j\right) \tag{1}$$

where $tf\left(r_i, t_j\right)$ is the frequency of word t_j in comment r_i, $df\left(t_j\right)$ is the number of comments in which word t_j appears, and |R| is the number of useful comments. The top 25% of terms by TF-IDF are considered as the set of product characteristics, totaling 12,762 words in this study.

Classification of Hotel Characteristics

This study refers to TripAdvisor.com's six categories for hotel ratings, namely location, sleep quality, service, value, cleanliness, and room, which are defined as types of hotel features (Rhee & Yang, 2015). We used Word2Vec to measure the co-occurrence between feature words and features.

Word2Vce (Mikolov et al., 2013) is a toolkit developed by Google for obtaining word vectors. It makes use of two different learning algorithms: CBOW, which aims to predict words given surrounding words, and Skip-gram, which predicts a set of words when a word is known. In this study, Word2Vec's word similarity measure is used to group similar words in a category; for instance, breakfast, toast, and milk are considered similar.

After pre-processing the comment contents with words, we used Word2Vec for word training to generate the word vector. In this study, the dictionary defined the nouns associated with six characteristic words—location, sleep quality, service, value, cleaning, and room—and explored the relationship between the top 25% of nouns and the six characteristic words. We used cosine similarity as defined as Equation 2 to compare the similarity of two words; if the same word appears in multiple classifications, a higher similarity is used. Letting x_i and y_j be the noun for the feature word and the 25% representative noun respectively, the final representative category for location has a total of 3907 words; sleep quality has a total of 2380 words; room has 714 words; service has 3831 words; value has 620 words; and cleaning has 1310 words. After word categorization was completed, Tripadvisor.com users were asked to evaluate whether or not the categorization was good; 15 words were randomly selected from each of the six categories for a total of 90 words, and users were asked to evaluate whether the categorization was correct or not. As a result, 69 words were deemed correctly categorized, or a proportion of 69/90 = 0.76 or 76%.

$$\cos\left(x_i, y_j\right) = \frac{x_i \times y_j}{\left|x_i\right| \times \left|y_j\right|} \tag{2}$$

Table 2. Example of Review Categorization

Reviews	Raw text
R_1	I had reservations for my wedding. Everything about this place is wonderful. The staff are friendly, the complimentary breakfast, parking and WiFi are awesome. The bed is comfortable. So much so that we tried to get the manufacture. We extended our stay one more night. My husband and I would def stay there again.
R_2	Melissa creates amazing cocktails and serves up a warm experience at the bar of this lovely hotel. The staff overall, from Ralph to Mike ensure that every guest has a comfortable and warm stay. Whenever I travel for business, Melissa welcomes me to unwind and even catch up on work at times while relaxing near the bar and balancing the evening with engaging conversation and genuine hospitality!
R_3	My husband and I travel to NY about eight times a year. I am so pleased that the Bronx now has a Residence Inn. It's brand new and the studio suite was large, equipped with a full kitchen, and there's a bar in the lobby. It was a cold rainy night and we were glad to be able to walk inside through the Atrium to Applebee's. The Atrium connects to Montefiore Medical Center, there is no charge for parking, the neighborhood is safe, the subway a short walk away, and the staff will fulfill your every need in a New York minute. This hotel required fewer points than a Manhattan Marriott, so I was able to shave the price of my stay down under $150. We both play guitar, and appreciate the extra room. Truly relaxing. The hotel will also shuttle you the short distance to the lovely rustic City Island. Lovely part of New York City.

Sample Sentences From Hotel Reviews

The following three examples of review sentences for the Residence Inn, a hotel in New York City, are categorized as shown in Table 2.

All the nouns from these three reviews were extracted as characteristic words, and a single noun was then extracted for each word, as shown in Table 3.

In this study, all terms from reviews of the Casablanca Hotel by Library Hotel Collection in New York City were selected and ranked by calculating the frequency of the terms using Equation 1, and the representative terms were identified as the characteristic terms for this hotel. The top 10 terms in the hotel dictionary, as constructed using these terms, are listed in Table 4.

The sentences were classified into six categories by comparing them with the feature dictionary, as shown in Table 5.

Emotion Analysis

Based on the classification of hotel characteristics described in the Selection of Characteristics section, a collection of six hotel characteristic sentences was generated. Each set was then classified into sentiment categories by month, sentiment scores were calculated for each sentence, and they were classified into positive and negative sentiment scores. Then, the sentiment scores for the processed and classified sentences were used to construct the graph.

Table 3. Review Terms

Reviews	Retrieved Terms
R_1	Reservations, wedding, Everything, staff, breakfast, parking, WiFi, bed, manufacture, stay, night, husband
R_2	Cocktail, experience, bar, hotel, staff, guest, stay, business, work, time, evening, conversation, hospitality
R_3	Husband, time, year, brand, studio, suite, kitchen, bar, lobby, night, charge, parking, neighborhood, subway, point, price, stay, guitar, room, hotel, shuttle, distance, part

Note. Source: Compiled by this study.

Table 4. Example of Feature Selection

Characteristics	Example of Dictionary Content
Rooms	room, bed, floor, bathroom, window, space, view, bedroom, kitchen, TV
Cleanliness	lounge, décor, standard, plenty, comfort, facility, bath, toilet, cleanliness, clean
Services	wine, café, staff, service, hotel staff, reception, smile, Wi-Fi, hospitality, member
Sleep Quality	sleep, neighbor, night, noise, atmosphere, dream, pillow, door, air, refreshment
Location	location, area, city, street, station, train, distance, bus, airport, place
Value	visit, restaurant, experience, star, value, deal, quality, cost, breakfast, price

Note. Source: Compiled by this study.

Incorporating Time and Sentence Fraction Sentiment Calculation

The sentences were categorized by month and the sentiment scores were then calculated based on the sentences in each hotel feature set. In this study, Opinionfinder was used to classify words into five polarities: Strong Positive, Positive, Neutral, Negative, and Strong Negative (Riloff & Wiebe, 2003). The Opinionfinder tool identifies the subjective strengths of words in these five categories; it provides unsupervised sentiment detection using the MPQA Corpus subjective word dictionary, which contains 8221 words, and labels each word with its lexicality, root reduction, and sentiment polarity. The sentiment score of the sentence is calculated using the following formula:

Table 5. Example of Sentence Classification by Hotel Feature

Hotel Features	Comment Sentences
Location	Everything about this place is wonderful.
	The Atrium connects to Montefiore Medical Center, there is no charge for parking, the neighborhood is safe, the subway a short walk away.
	The hotel will also shuttle you the short distance to the lovely rustic City Island.
Sleep Quality	Bed was comfortable so getting a good nights sleep was not an issue.
Rooms	The bed is comfortable.
	It's brand new and the studio suite was large, equipped with a full kitchen, and there's a bar in the lobby.
	The room, the amenities of the hotel all made for a great stay that only made a lovely wedding into a great weekend.
	The room we booked was a Studio, King which was a very comfortable and well arranged.
	Rooms need blackout curtains that are fitted to the windows. We were trying to sleep during the day which was next to impossible with the noise and light.
Service	The staff are friendly.
	Melissa creates amazing cocktails and serves up a warm experience at the bar of this lovely hotel.
	The staff overall, from Ralph to Mike ensure that every guest has a comfortable and warm stay.
Value	This hotel required fewer points than a Manhattan Marriott, so I was able to shave the price of my stay down under $150.
Cleanliness	The cleanliness of the hotel was impeccable and the complimentary breakfast was delicious!

Note. Source: Compiled by this study.

$$Sentiment_i = \left(str\_pos_i \times 2 + weak\_pos_i\right) - \left(str\_neg_i \times 2 + weak\_neg_i\right) \quad (3)$$

where $\mathrm{str\_pos}_i$, $\mathrm{weak\_pos}_i$, $\mathrm{weak\_neg}_i$, and $\mathrm{str\_neg}_i$ are the numbers of words with strong subjective positive sentiment, weak subjective positive sentiment, weak subjective negative sentiment, and strong subjective negative sentiment in sentence *i*, respectively. Finally, the sentiment score of each sentence is calculated as the sum of the word scores, and sentences are classified as $Positive_i$ and $Negative_i$, with $Sentiment_i > 0$ being positive and $Sentiment_i < 0$ being negative.

Table 6 shows the calculated polarities for the three reviews of the Residence Inn in New York City. First, the sentiment score and sentiment classification were calculated for the review sentences over the six characteristics. The sentiment scores were calculated using Equation 3, and sentence polarity was determined as described above. For example, in the sentence "Everything about this place is wonderful," the word "wonderful" is marked as a strong subjective positive emotion, so the emotion score is 2×1 = 2 and the emotion polarity is determined as positive.

Calculating Sentiment Scores

In this study, the scores of review sentences in six feature categories in each season are calculated. $Season - feature\,score_i$ is the sentiment score of feature category *i* in each season, $Sum\,of\,Positive_{si}$ is the sum of positive scores of feature sentence category *si* in each season, $Sum\,of\,Negative_{si}$ is the sum of negative scores of feature sentence category *si* in each season, $Sum\,of\,Sentence_i$ is the number of sentences of characteristic category *i* in each quarter. March–May is taken as the first quarter, June–August as the second quarter, September–November as the third quarter, and December–February as the fourth quarter. The following is the formula for calculation:

$$Season - feature\,score_i = \frac{Sum\,of\,Positive_{si} + Sum\,of\,Negative_{si}}{Sum\,of\,Sentence_i} \quad (4)$$

Table 7 shows the review sentences for the Residence Inn in New York City, with positive sentences for the location category as the sum of sentiment scores for the third quarter as 8 and negative sentences as 0. The number of sentences for the third quarter, 3, was then used to calculate the season–feature score as (8+0)/3 = 8/3 for the location category in the third quarter of 2017.

Create a Chart

As outlined in the Calculating Sentiment Scores section, review data can be compiled for each hotel as shown in Table 8, and the calculated scores can be used to construct visual charts such as line graphs, dashboards, or pie charts using Microsoft's Power BI tool, with each hotel having its own visual chart. Power BI is an interactive visualization tool that provides a wide variety of visual charts and report styles and that can be connected to huge amounts of data in the cloud or internally, from sources such as Hadoop, Spark, and so on. Because it can connect to data from different sources, Power BI can provide in-depth analysis for a range of situations.

Using the Power BI tool, we can import external data to construct a line graph and observe the change of sentiment score over the seasons through the graph; graphs can be linked to each other, as in Figure 4. The content of the data will be changed into 2015 Seasons service content. Finally, this study uses html to create the website to construct 10 hotel pages and then links them with Power BI software to embed the charts into the website. Figures 5 and 6 show one of the hotel pages. The interface contains the following information: (a) a line graph of sentiment scores by time; (b) a histogram of sentiment scores by face; (c) hotel stars; (d) useful review filters; (e) hotel review stars; (f) travel

Table 6. Sentence Affective Polarity Classification

Hotel Features	Comment Sentences	Emotional Scores	Sentence Emotion	Month	Year
Location	Everything about this place is wonderful.	2	Positive	11	2017
	The Atrium connects to Montefiore Medical Center, there is no charge for parking, the neighborhood is safe, the subway a short walk away.	2	Positive	11	2017
	The hotel will also shuttle you the short distance to the lovely rustic City Island.	4	Positive	11	2017
Sleep Quality	Bed was comfortable so getting a good nights sleep.	3	Positive	5	2017
Rooms	The bed is comfortable.	2	Positive	11	2017
	It's brand new and the studio suite was large, equipped with a full kitchen, and there's a bar in the lobby.	4	Positive	11	2017
	The room, the amenities of the hotel all made for a great stay that only made a lovely wedding into a great weekend.	6	Positive	12	2017
	The room we booked was a Studio, King which was a very comfortable and well arranged.	2	Positive	12	2017
	Rooms need blackout curtains that are fitted to the windows. We were trying to sleep during the day which was next to impossible with the noise and light.	−1	Negative	7	2017
Service	The staff are friendly.	2	Positive	11	2017
	Melissa creates amazing cocktails and serves up a warm experience at the bar of this lovely hotel.	5	Positive	11	2017
	The staff overall, from Ralph to Mike ensure that every guest has a comfortable and warm stay.	3	Positive	11	2017
Value	This hotel required fewer points than a Manhattan Marriott, so I was able to shave the price of my stay down under $150.	2	Positive	11	2017
Cleanliness	The cleanliness of the hotel was impeccable and the complimentary breakfast was delicious!	3	Positive	11	2017

Note. Source: Compiled by this study.

Table 7. Example of Calculating Sentiment Scores for Feature Categories

Hotel Features	Comment Sentences	Emotional Scores	Sentence Emotion	Month	Year
Location	Everything about this place is wonderful.	2	Positive	11	2017
	The Atrium connects to Montefiore Medical Center, there is no charge for parking, the neighborhood is safe, the subway a short walk away.	2	Positive	11	2017
	The hotel will also shuttle you the short distance to the lovely rustic City Island.	4	Positive	11	2017

Note. Source: Compiled by this study

Table 8. Quarterly Sentiment Scores for a Hotel Feature

Hotel Feature	Year, Season	Emotional Scores
Location	2017: Season 1	3
	2017: Season 2	2.2
	2017: Season 3	2.5
	2017: Season 4	3.1
Sleep Quality	2017: Season 1	0.4
	2017: Season 2	2.3
	2017: Season 3	4.2
	2017: Season 4	3.5
Rooms	2017: Season 1	5
	2017: Season 2	6.1
	2017: Season 3	4.5
	2017: Season 4	6.2
Service	2017: Season 1	6
	2017: Season 2	3.1
	2017: Season 3	4.1
	2017: Season 4	5.1
Value	2017: Season 1	5.5
	2017: Season 2	3.2
	2017: Season 3	4.2
	2017: Season 4	4.1
Cleanliness	2017: Season 1	2.1
	2017: Season 2	1.1
	2017: Season 3	0.8
	2017: Season 4	1.1

Note. Source: Compiled by this study.

type; (g) month filter; (h) a bar graph of sentiment scores by time; and (i) review information. In addition to embedding the charts into the website, the new interface also retains the reviews section of the old website, as consumers perceive useful reviews to be highly reliable (Cao et al., 2011), and this study adds a selection of useful reviews to the original section, allowing the desired reviews to be found based on time and hotel characteristics.

Experimental Design

This study investigates an interface that incorporates visual graphs (the new interface) to allow users to speed up comment browsing and sets the following hypotheses:

1. The accuracy of using the new interface will be higher than the accuracy of using the old interface
2. The response time using the new interface will be less than the response time using the old interface.
3. The time to find the correct answer and fill in the user answer varies for different facing interfaces.

Figure 4. Original Interface Schematic Diagram

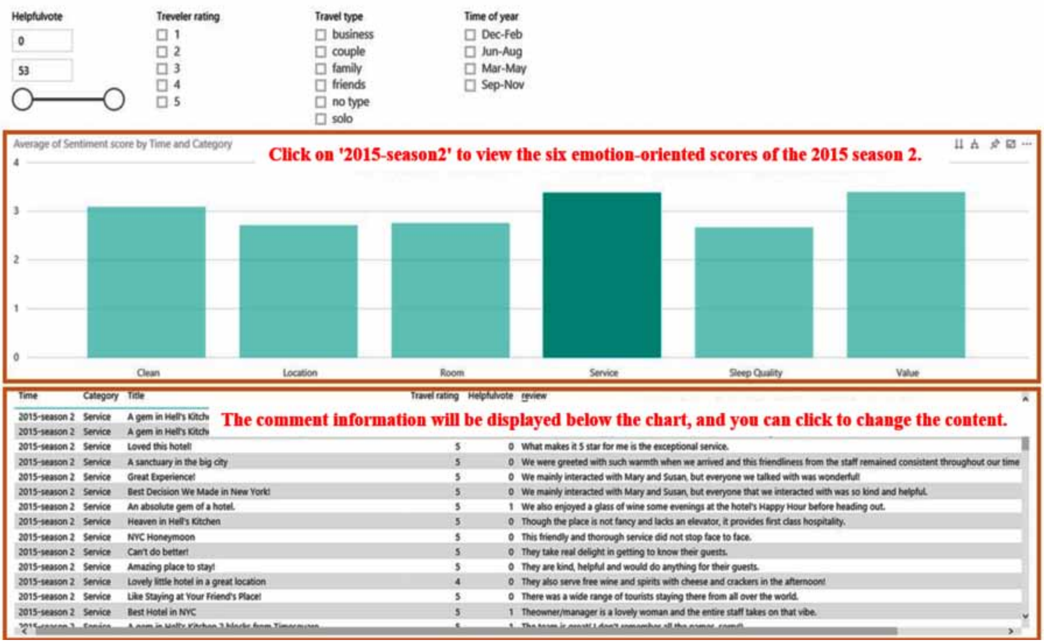
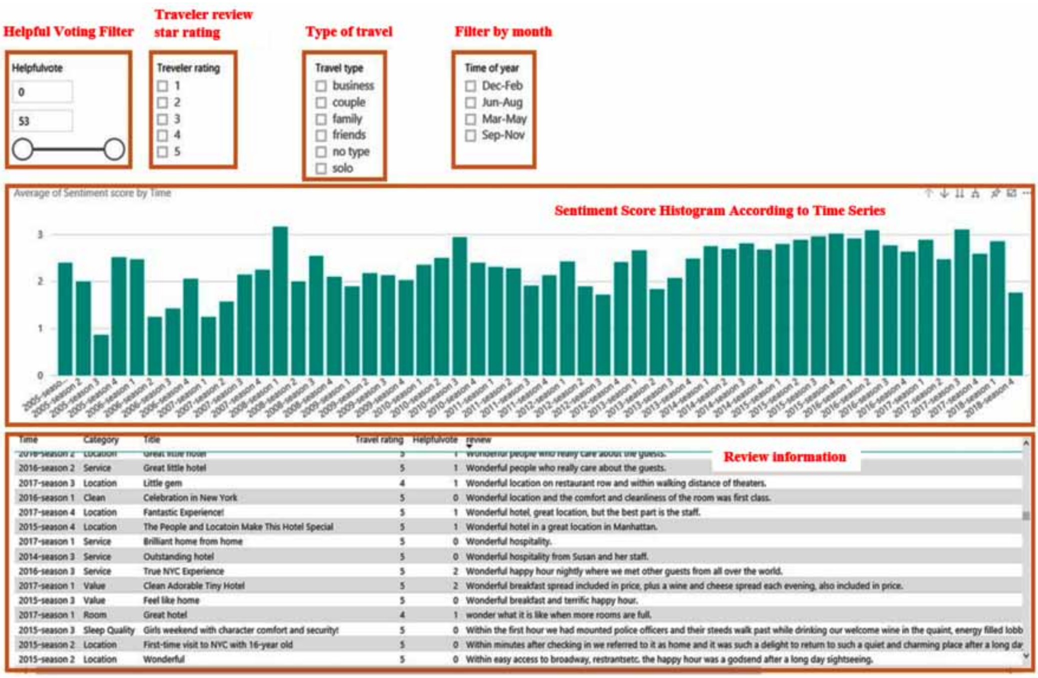


Figure 5. New Interface Schematic Diagram, First Section

Figure 6. New Interface Schematic Diagram, Second Section



4. The order in which the interfaces were tested differed in the time it took for users to find the correct answer and user answer.

The 28 participants were either university undergraduate or graduate students who had used the Tripadvisor.com platform, recruited via a social network. This experiment consisted of four different questionnaires; the testers each took one of the four questionnaires and designed it according to the interface (new/original) and orientation (value, room, service, and cleaning). These were New Interface Value Orientation/Original Interface Room Orientation, Original Interface Value Orientation/New Interface Room Orientation, New Interface Service Orientation/Old Interface Cleaning Orientation, and Original Interface Service Orientation/New Interface Service Orientation, as shown in Table 9. The content of each questionnaire (see Appendix 1) is divided into three parts:

Table 9. Experimental Design Table

Experiment: Questionnaire 1	New	Original	Seven subjects
	Value	Rooms	
Experiment: Questionnaire 2	Original	New	Seven subjects
	Value	Rooms	
Experiment: Questionnaire 3	New	Original	Seven subjects
	Service	Cleanliness	
Experiment: Questionnaire 4	Original	New	Seven subjects
	Service	Cleanliness	

Note. Source: Compiled by this study.

1. Part 1 is basic information. This includes gender and whether you have used Tripadvisor.com.
2. Part 2 is English proficiency. This includes whether or not you have obtained an English certificate and whether you agree that you speak English well (average, disagree, strongly disagree, agree, strongly agree).
3. Part 3 involves the determination of the correct answer using the new/original interface and how long it took to find the answer. For example, which hotel is trending (up or down) in guest satisfaction from 2014–2017 using the new interface? In this case, use the new interface to find out which hotel is trending (up or down) from 2014–2017 (first- and second-placed hotel), and then record the time to find the answer.

Mental load (problem solving ability, working memory requirements during reasoning or thinking) may affect user satisfaction and performance when completing complex tasks (Schmutz et al., 2009). This experiment therefore used a NASA tool, NASA-TLX, which assesses users' mental load based on the weighted average of six indicators as shown in Table 10, namely mental demand, physical demand, time demand, self-performance, effort, and frustration. The mental load rating is obtained by multiplying the weights of each indicator by scores for the six indicators and adding them together. Participants fill out two scales, one after using the new interface and one after using the original interface. Appendix 2 gives the content of the questionnaire.

EXPERIMENTAL RESULTS AND ANALYSIS

Dataset

The study was conducted over about three weeks from December 24 to January 10, 2018, during which two administrations were performed. The first was a pre-test in which six questionnaires were collected to test how long it took users to find the correct answer; the timing of the experiment and the definition of terminology were refined before the second administration. A total of 28 questionnaires were collected for the second administration.

The descriptive statistics of the dataset are described as following:

- The gender of the participants: 16 subjects were female and 12 were male.
- Whether the participants have used the Tripadvisor.com website: all subjects have used the Tripadvisor.com.

Table 10. NASA-TLX Scale Indicators

Indicators	Level	Description
Mental demand	Low/High	How much mental and perceptual activity (e.g., thinking, deciding, observing, etc.) is required? Are the tasks easy or demanding, simple or complex?
Physiological needs	Low/High	How much physical activity is required (e.g., mobility)? Are the tasks easy or demanding, easy or strenuous?
Time requirements	Low/High	How much time pressure does the pace of the task make you feel? Is the pace of the task slow or fast?
Self-performance	Good/ Bad	How satisfied are you with your performance in meeting the task objectives? How well do you feel you have achieved your task objectives?
Efforts	Low/High	How much effort is required to achieve the level required for this task?
Level of frustration	Low/High	What is the level of uncertainty, frustration, irritation, nervousness, etc. that you feel when the task is being carried out? How frustrated did you feel during the task?

Note. Source: https://humansystems.arc.nasa.gov/groups/TLX/

Table 11. Data Set for Questionnaire

No.	Column Name	Column Type	Column Description	Variable Number
1	Interface	String	New Interface / Old Interface	Independent variable
2	Aspect	String	Value / Rooms / Service / Cleanliness	Independent variable
3	New_ Logarithm	Number	The correct number of answers using the new interface	Dependent variable
4	New_ Time	Date	Use the new interface to fill in the answer time	Dependent variable
5	Old_ Logarithm	Number	Correct number of answers using the original interface	Dependent variable
6	Old_ Time	Date	Use the original interface to fill in the answer time	Dependent variable

Note. Source: Compiled by this study.

- The education level of the participants: 26 subjects were graduate students and 2 were undergraduates.
- Whether the participants have English certificates: 25 subjects have English certificates, and three do not.
- How well the participants think they speak English: two are self-rated as good, 21 as average and five as poor.

The collected questionnaires were organized according to the variables listed in Table 11 and Table 12.

Experimental Results and Evaluation

This study was designed to evaluate the effectiveness of using the new interface and whether it would cause mental overload for users. The experiment was divided into four parts: (a) to assess whether the new interface had an effect on the number of correct answers and time spent by the test subjects,

Table 12. NASA-TLX Data Set

No.	Column Name	Column Type	Column Description
7	New_MentalDemand	Number	Mental demand score for using the new interface
8	New_PhysicalDemand	Number	Physiological demand score for using the new interface
9	New_TemporalDemand	Number	Time requirement fraction for using the new interface
10	New_Performance	Number	Self-performance scores using the new interface
11	New_Effort	Number	Effort score for using the new interface
12	New_Frustration	Number	Frustration level score for using the new interface
13	Old_MentalDemand	Number	Mental demand score for using original interface
14	Old_PhysicalDemand	Number	Physiological demand score using original interface
15	Old_TemporalDemand	Number	Time requirement fraction for using original interface
16	Old_Performance	Number	Self-performance scores using the original interface
17	Old_Effort	Number	Effort score using original interface
18	Old_Frustration	Number	Frustration score using original interface

Note. Source: Compiled by this study.

using a paired t-test; (b) to assess whether differently-oriented interfaces might have different effects on the number of correct answers and time spent by the test subjects, using factorial ANOVA; (c) to assess whether the number of correct answers and time spent by the two interfaces were important to the test sequence, using factorial ANOVA; and (d) to assess whether the number of correct answers and time spent by the test subjects were important to the test sequence, again using factorial ANOVA. The importance of the time dimension in the test sequence was assessed using factorial ANOVA, and the NASA-TLX scale was used to measure the difference in mental load between users using the original and new interfaces, assessing the significance using a paired t-test.

In Experiment A, this study investigated whether the original and new interfaces affected the number of correct answers found and the time spent by the respondents. The number of New_answer/Old_answer pairs and the time spent using the new and old interfaces (New_time/Old_time) were evaluated using a paired t-test for the 28 questionnaires. The sample statistics showed that the mean number of pairs of correct answers in the new interface was higher than the mean number of answers in the original interface, and the time taken to find the answer in the new interface was 200.32 seconds lower than the time taken to find the answer in the original interface (361.61 seconds), as shown in Table 13. The results in Table 14 show that the *p*-value ($p = 0.009 < 0.05$) indicates that the new interface did affect the time taken by the respondents to find the correct answer and the time spent.

In Experiment B, the effect of different interfaces on the number of correct answers and filling time was investigated. Comparing the number of correct answers and the time taken to fill in the answers for the value, room, service, and cleaning orientations, Table 15 represents the subjected factors. The effect of using the original and new orientations on the number of correct answers was found to be significant ($p = 0.000 < 0.05$), as shown in Table 16, indicating that using different orientations affected the number of correct answers. However, the number of answers is not affected by the use of different interfaces ($p = 0.065 > 0.05$). From Table 17, we can see that the new interface has a significant effect on the filling-in time compared to the original ($p = 0.011 < 0.05$), meaning

Table 13. Experiment A Sample Statistics

Paired Sample Statistics					
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	New_answer	0.89	28	0.315	0.060
	Old_answer	0.39	28	0.497	0.094
Pair 2	New_time	200.32	28	201.350	38.051
	Old_time	361.61	28	253.558	47.918

Table 14. Experiment A: Paired Sample Assay

Paired Samples Test									
		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	New_answer – Old_answer	0.500	0.638	0.121	0.252	0.748	4.145	27	**0.000**
Pair 2	New_time – Old_time	−161.286	302.461	57.160	−278.568	−44.004	−2.822	27	**0.009**

Table 15. Inter-Subject Factors of Experiment B

Between-Subjects Factors			
		Value Label	N
Interface	1.00	new	28
	2.00	original	28
Aspect	1.00	value	14
	2.00	room	14
	3.00	service	14
	4.00	clean	14

Table 16. Experimental B: Factorial ANOVA Analysis

Tests of Between-Subjects Effects					
Dependent Variable:	Answer				
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	4.857[a]	7	0.694	4.163	0.001
Intercept	23.143	1	23.143	138.857	0.000
Interface	3.500	1	3.500	21.000	**0.000**
aspect	1.286	3	0.429	2.571	**0.065**
interface * aspect	0.071	3	0.024	0.143	**0.934**
Error	8.000	48	0.167		
Total	36.000	56			
Corrected Total	12.857	55			

a. R Squared = .378(Adjusted R Squared = .287)

Table 17. Experimental B: Factorial ANOVA Analysis

Tests of Between-Subjects Effects					
Dependent Variable:	Time				
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	700236.214[a]	7	100033.745	1.925	0.086
Intercept	4420692.071	1	4420692.071	85.066	0.000
interface	364183.143	1	364183.143	7.008	**0.011**
aspect	103976.357	3	34658.786	0.667	**0.576**
interface * aspect	232076.714	3	77358.905	1.489	**0.230**
Error	2494449.714	48	51967.702		
Total	7615378.000	56			
Corrected Total	3194685.929	55			

a. R Squared = .219(Adjusted R Squared = .105) that the use of different interfaces affects the filling-in time; the difference between aspects has no significant effect on the filling-in time ($p = 0.576 > 0.05$), meaning that we find no effect for the different aspects in Figure 7. From another point of view, Figure 8 shows that the four orientations do not affect the number of responses due to the original versus new orientations in Figure 9. Figure 10 shows that the four orientations do not affect the filling-in time due to the difference between the original and new interfaces. Next, a post hoc test was conducted using the Scheffe method to compare whether there is a significant difference between the two groups of the orientations; we observe from Tables 18 and 19 that the number of responses and the response time are not affected by the different orientations.

In Experiment C, we investigated whether the number of correct answers and the time spent on the two interfaces were important for the order of the test. Table 20 shows the tested factors; Table 21 shows that there is no significant interaction between order and interface ($p = 0.684 < 0.05$), while Table 22 shows that there is no significant interaction between order and interface ($p = 0.134 < 0.05$). There is no significant interaction between the two parallel lines in Figure11, which means that the original and new interfaces have no effect on the order of the test in Figure 12.

Experiment D explored the level of mental load on the users in using the original and new interface; the higher the score, the greater the mental load and the need to think hard to find the answer. To understand the level of frustration in using the original versus the new interface, we observe from Table 23 that the mean score of all variables in the new interface is significantly smaller than the mean score of all variables in the original interface; Table 24 shows that the mental and physical demands are significant ($p = 0.023 < 0.05$, $p = 0.011 < 0.05$); time demand ($p = 0.017 < 0.05$), self-performance ($p = 0.002 < 0.05$), effort ($p = 0.000 < 0.05$), and frustration ($p = 0.004 < 0.05$) are all show significant effects. The level of mental load using the new interface was thus significantly

Figure 7. Experiment B, Answers: Factorial ANOVA Line Graph

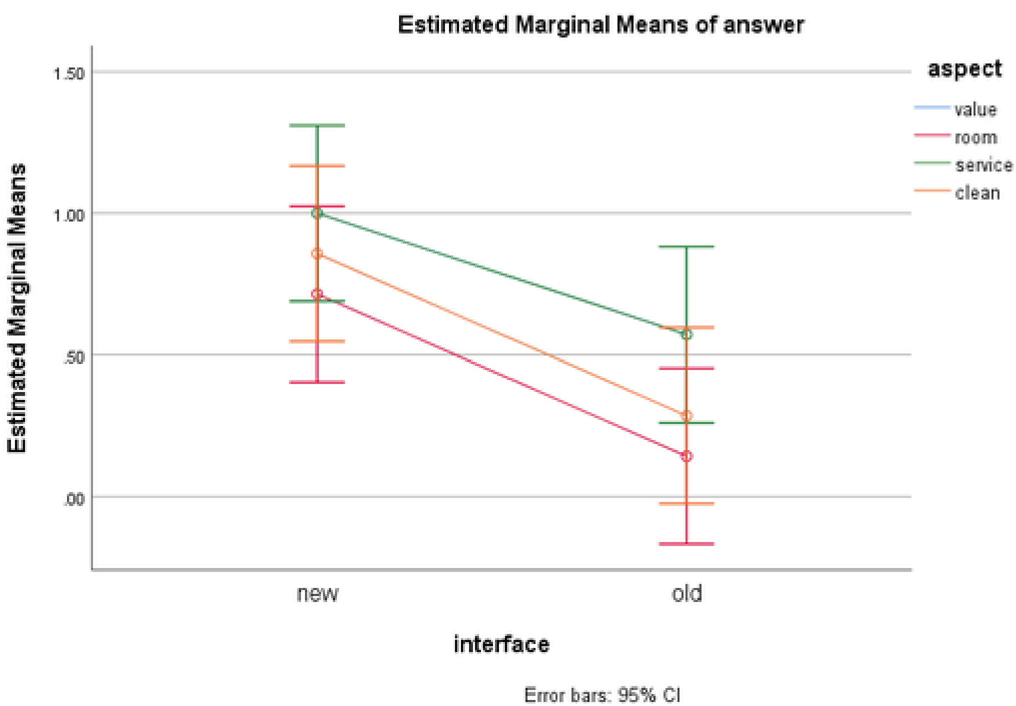


Figure 8. Experiment B, Answers: Factorial ANOVA Bar Graph

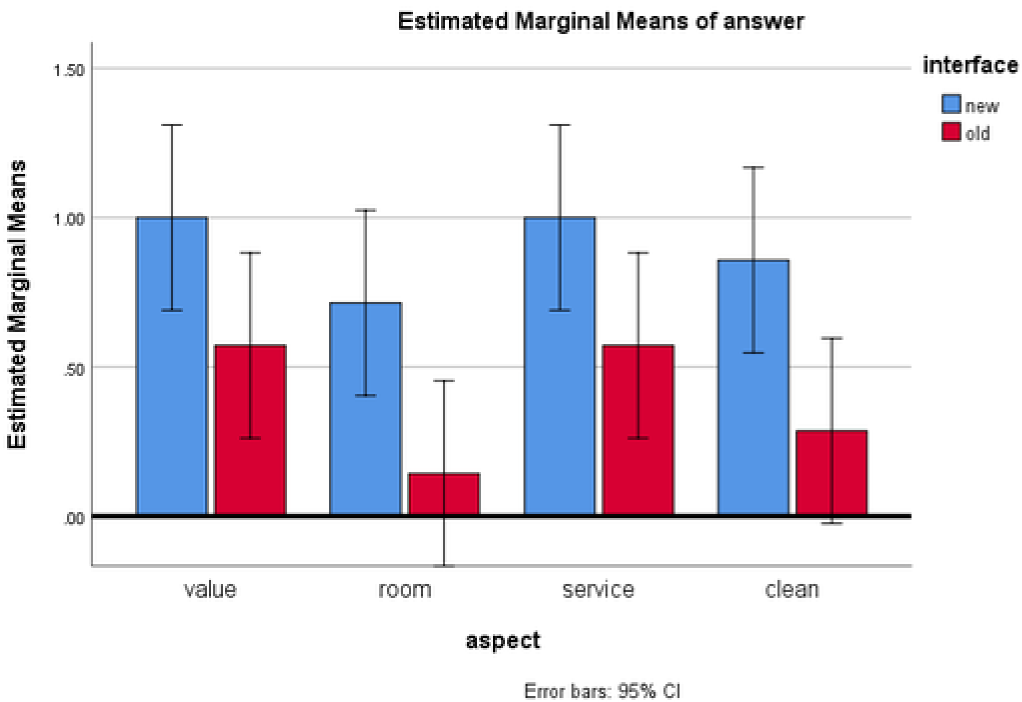


Figure 9. Experiment B, Time Variable: Factorial ANOVA Line Graph

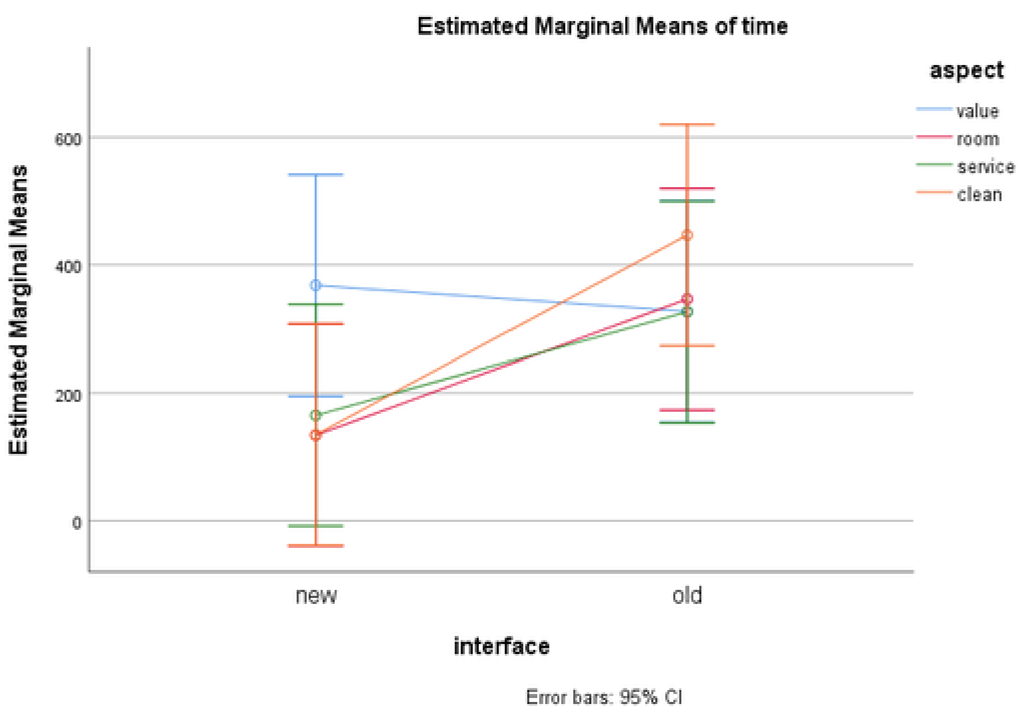


Figure 10. Experiment B, Time Variable: Factorial ANOVA Long Bar Graph

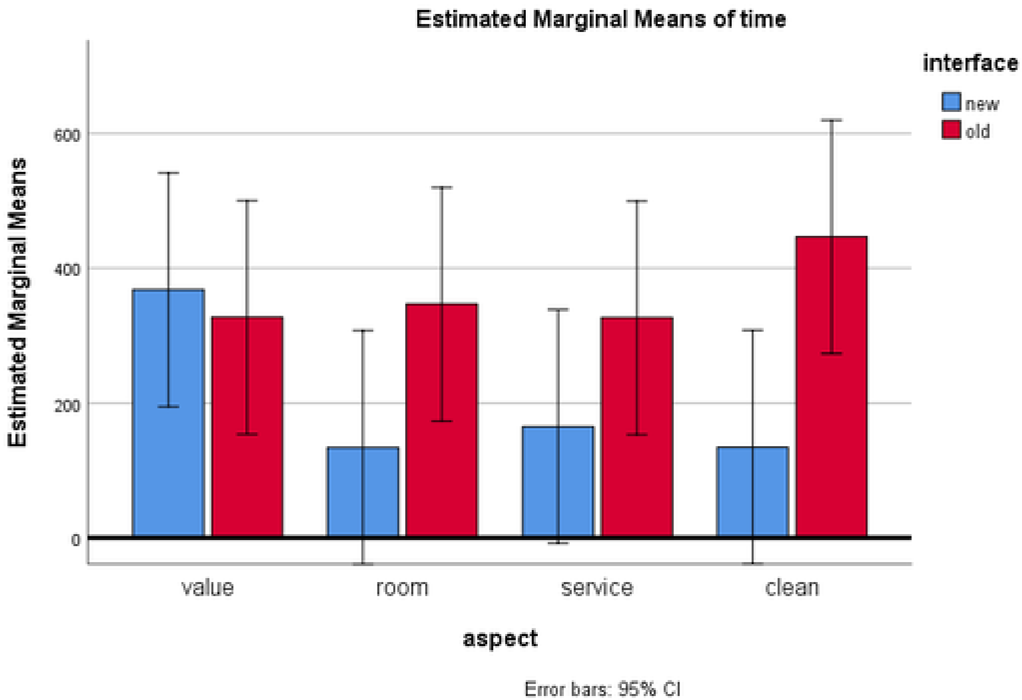


Table 18. Experiment B: Post Hoc Test

Multiple Comparisons						
Dependent Variable: Answer						
Scheffe						
(I)aspect	(J)aspect	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
value	room	.3571	.15430	**.162**	−.0899	.8042
	service	.0000	.15430	**1.000**	−.4471	.4471
	clean	.2143	.15430	**.591**	−.2328	.6613
room	value	−.3571	.15430	**.162**	−.8042	.0899
	service	−.3571	.15430	**.162**	−.8042	.0899
	clean	−.1429	.15430	**.835**	−.5899	.3042
service	value	.0000	.15430	**1.000**	−.4471	.4471
	room	.3571	.15430	**.162**	−.0899	.8042
	clean	.2143	.15430	**.591**	−.2328	.6613
clean	value	−.2143	.15430	**.591**	−.6613	.2328
	room	.1429	.15430	**.835**	−.3042	.5899
	service	−.2143	.15430	**.591**	−.6613	.2328

Based on observed means.
The error term is Mean Square (Error) = .167.

Table 19. Experiment B: Post Hoc Test

Multiple Comparisons						
Dependent Variable: Time						
Scheffe						
(I) aspect	(J) aspect	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
value	room	107.36	86.162	**.672**	−142.28	356.99
	service	101.79	86.162	**.708**	−147.85	351.42
	clean	57.00	86.162	**.932**	−192.64	306.64
room	value	−107.36	86.162	**.672**	−356.99	142.28
	service	−5.57	86.162	**1.000**	−255.21	244.06
	clean	−50.36	86.162	**.952**	−299.99	199.28
service	value	−101.79	86.162	**.708**	−351.42	147.85
	room	5.57	86.162	**1.000**	−244.06	255.21
	clean	−44.79	86.162	**.965**	−294.42	204.85
clean	value	−57.00	86.162	**.932**	−306.64	192.64
	room	50.36	86.162	**.952**	−199.28	299.99
	service	44.79	86.162	**.965**	−204.85	294.42

Based on observed means.
The error term is Mean Square (Error) = 51967.702.

Table 20. Inter-Subject Factors of Experiment C

Between-Subjects Factors			
		Value Label	N
Order	1.00	first	14
	2.00	second	14
interface	1	new	14
	2	original	14

lower than that of the original interface, so the new interface could reduce fatigue while achieving the desired goal.

CONCLUSION AND RECOMMENDATIONS

This study proposes a four-stage approach to constructing charts: collecting reviews and pre-processing the data; constructing a dictionary of six hotel characteristics; categorizing sentences and calculating sentiment scores; and finally constructing the chart to be embedded on the website. The experimental results showed that using the new interface had a positive impact on the number of answers found and the time it took to find them, and that using the new interface was less mentally taxing than using the original one.

Table 21. Experiment C: Factorial ANOVA Analysis

Tests of Between-Subjects Effects					
Dependent Variable:	**Answer**				
Source	**Type III Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**
Corrected Model	2.679[a]	3	0.893	5.357	0.006
Intercept	10.321	1	10.321	61.929	0.000
order	0.893	1	0.893	5.357	**0.030**
interface	1.750	1	1.750	10.500	**0.003**
Order * interface	0.036	1	0.036	0.214	**0.648**
Error	4.000	24	0.167		
Total	17.000	28			
Corrected Total	6.679	27			

a. R Squared = .401(Adjusted R Squared = .326)

Table 22. Experiment C: Factorial ANOVA Analysis

Tests of Between-Subjects Effects					
Dependent Variable:	**time**				
Source	**Type III Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**
Corrected Model	244715.536[a]	3	81571.845	1.740	0.186
Intercept	2417268.893	1	2417268.893	51.571	0.000
order	80678.893	1	80678.893	1.721	**0.202**
interface	51514.321	1	51514.321	1.099	**0.305**
Order * interface	112522.321	1	112522.321	2.401	**0.134**
Error	1124948.571	24	46872.857		
Total	3786933.000	28			
Corrected Total	1369664.107	27			

a. R Squared = .179(Adjusted R Squared = .076)

The new interface retains the blocks of the original interface and has a graphical display so that users can click on it to find the relevant comments and therefore find the answers faster, while the original interface only has date sorting, so it is necessary to browse the comments one by one to find the answers.

There are limitations to this study, and future research can help us to deepen our understanding of these areas. This study uses free comments in an online community. There are no certain grammatical rules and the comment content often contains popular vocabulary on the internet. In these situations, the words may not be recognized. This study uses the hotel characteristics (that is, location, sleep quality, service, value, cleanliness, room) defined by TripAdvisor.com. However, the hotel characteristics in which tourists are interested may not be limited to the above six categories, or they may vary by season, country, and race. Therefore, future research can expand the topic of hotel characteristics to include data from different countries. In addition, the research structure and model of this study can also be applied to other online review platforms, such as IMDB or yelp.com,

Figure 11. Experiment C, Answer: Factorial ANOVA Line Graph

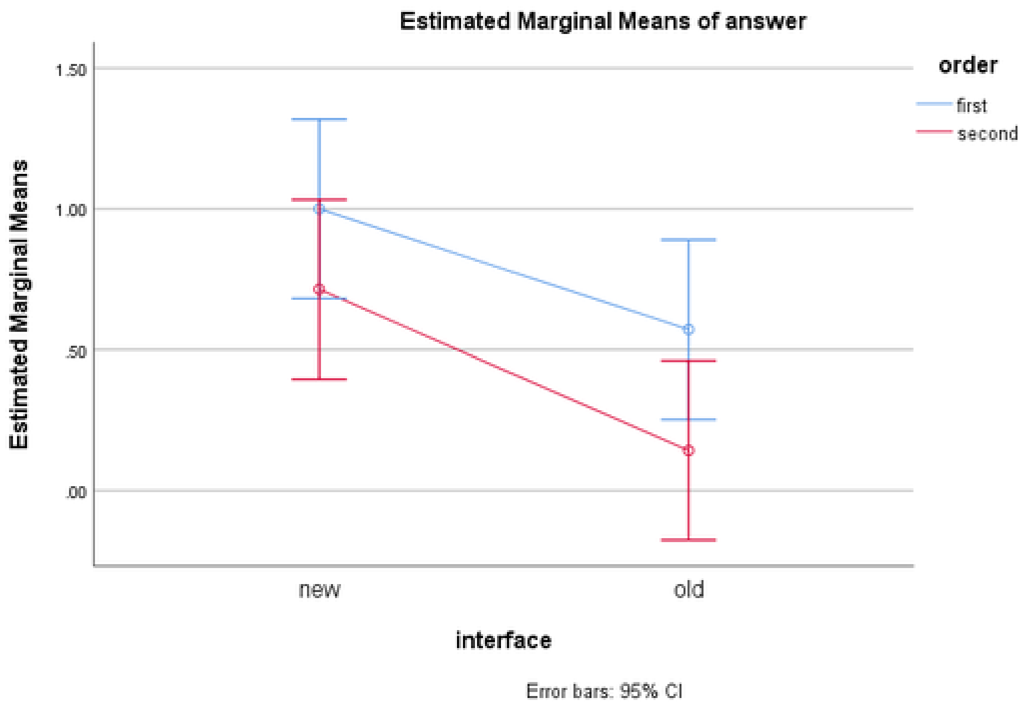


Figure 12. Experiment C, Time Variable: Factorial ANOVA Line Graph

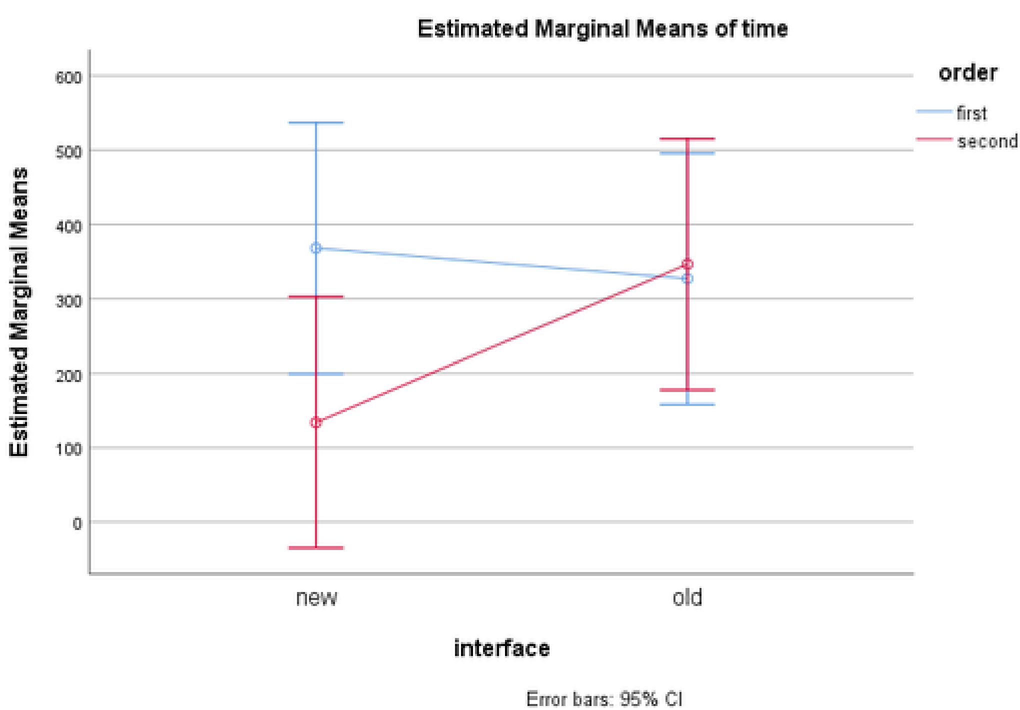


Table 23. Experiment D: NASA Sample Statistics

Paired Samples Statistics				
	Mean	**N**	**Std. Deviation**	**Std. Error Mean**
Old_MentalDemand	210.18	28	104.629	19.773
New_MentalDemand	153.04	28	115.368	21.803
Old_PhysicalDemand	101.61	28	108.227	20.453
New_PhysicalDemand	51.07	28	55.484	10.485
Old_TemporalDemand	155.00	28	110.696	20.920
New_TemporalDemand	111.25	28	94.913	17.937
Old_Performance	178.57	28	124.126	23.458
New_Performance	93.04	28	100.179	18.932
Old_Effort	187.32	28	127.435	24.083
New_Effort	115.89	28	94.927	17.940
Old_Frustration	101.79	28	112.950	21.346
New_Frustration	51.07	28	93.108	17.596

Table 24. Experiment D: NASA Sample Assay

Paired Samples Test									
	Paired Differences						**t**	**df**	**Sig. (2-tailed)**
		Mean	**Std. Deviation**	**Std. Error Mean**	**95% Confidence Interval of the Difference**				
					Lower	**Upper**			
Mental Demand	Old-New	57.143	125.133	23.648	8.621	105.664	2.416	27	**0.023**
Physical Demand	Old-New	50.536	97.414	18.410	12.762	88.309	2.745	27	**0.011**
Temporal Demand	Old-New	43.750	91.151	17.226	8.405	79.095	2.540	27	**0.017**
Performance	Old-New	85.536	131.651	24.880	34.487	136.585	3.438	27	**0.002**
Effort	Old-New	71.429	90.245	17.055	36.435	106.422	4.188	27	**0.000**
Frustration	Old-New	50.714	85.263	16.113	17.653	83.776	3.147	27	**0.004**

to extract product features in which users are interested to optimize the platform interface and help users achieve appropriate results more quickly.

CONFLICTS OF INTEREST

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

FUNDING STATEMENT

No funding was received for this work.

PROCESS DATES

March, 2024
Received: September 20, 2023, Revision: March 13, 2024, Accepted: February 18, 2024

CORRESPONDING AUTHOR

Correspondence should be addressed to Pei-Ju Lee, pjlee@nchu.edu.tw

REFERENCES

Archak, N., Ghose, A., & Ipeirotis, P. G. (2011). Deriving the pricing power of product features by mining consumer reviews. *Management Science*, *57*(8), 1485–1509. doi:10.1287/mnsc.1110.1370

Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SentiWordNet 3.0 : An enhanced lexical resource for sentiment analysis and opinion mining SentiWordNet. *Analysis*, *0*, 1–12.

Bagheri, A., Saraee, M., & DeJong, F. (2013). Care more about customers: Unsupervised domain-independent aspect detection for sentiment analysis of customer reviews. *Knowledge-Based Systems*, *52*, 201–213. doi:10.1016/j.knosys.2013.08.011

Bai, X., Padman, R., Airoldi, E., & Society, I. C. (2005). On learning parsimonious models for extracting consumer opinions. In *Proceedings of 38th Annual Hawaii International Conference on System Sciences*. IEEE.

Banerjee, S., & Chua, A. Y. K. (2016). In search of patterns among travellers' hotel ratings in TripAdvisor. *Tourism Management*, *53*, 125–131. doi:10.1016/j.tourman.2015.09.020

Benbasat, I., & Tan, J. K. (1990). Processing of Graphical information : A decomposition taxonomy to match data extraction tasks and graphical representations. *Information Systems Research*, *1*(4), 416–439.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Cao, Q., Duan, W., & Gan, Q. (2011). Exploring determinants of voting for the "helpfulness" of online user reviews: A text mining approach. *Decision Support Systems*, *50*(2), 511–521. doi:10.1016/j.dss.2010.11.009

Chang, Y. C., Ku, C. H., & Chen, C. H. (2017). Social media analytics: Extracting and visualizing Hilton hotel ratings and reviews from TripAdvisor. *International Journal of Information Management*, *48*, 263–279. doi:10.1016/j.ijinfomgt.2017.11.001

Daniel, K., Jörn, K., Mansmann, F., & Ellis, G. (2010). *Mastering the information age: Solving problems with visual analytics*. Eurographics Association.

Falschlunger, L., Lehner, O., & Treiblmaier, H. (2016). The impact of information overload on decision making outcome in high complexity settings. In *Proceedings of SIGHCI 2016*. AIS.

Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research*, *19*(3), 291–313. doi:10.1287/isre.1080.0193

Ghose, A., & Ipeirotis, P. G. (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*, *23*(10), 1498–1512. doi:10.1109/TKDE.2010.188

Giachanou, A., & Crestani, F. (2016). Like it or not: A survey of Twitter sentiment analysis methods. *ACM Computing Surveys*, *49*(2), 1–41. doi:10.1145/2938640

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the 2004 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '04*. Association for Computing Machinery.

Hwang, S.-Y., Lai, C.-Y., Jian, J.-J., & Chang, S. (2014). The Identification of noteworthy hotel reviews for hotel management. *Pacis*, *10*(2), 35–38. doi:10.17705/1pais.06402

Kelleher, C., & Wagener, T. (2011). Ten guidelines for effective data visualization in scientific publications. *Environmental Modelling & Software*, *26*(6), 822–827. doi:10.1016/j.envsoft.2010.12.006

Kim, S.-M., Pantel, P., Chklovski, T., & Pennacchiotti, M. (2006). Automatically assessing review helpfulness. In *Proceedings of the 2006 Conference of Empirical Methods in Natural Language Processing (EMNLP 2006)*. Association for Computational Linguistics. doi:10.3115/1610075.1610135

Lee, A. J. T., Yang, F. C., Chen, C. H., Wang, C. S., & Sun, C. Y. (2016). Mining perceptual maps from consumer reviews. *Decision Support Systems*, *82*, 12–25. doi:10.1016/j.dss.2015.11.002

Li, G., Law, R., Vu, H. Q., Rong, J., & Zhao, X. (2015). Identifying emerging hotel preferences using Emerging Pattern Mining technique. *Tourism Management*, *46*, 311–321. doi:10.1016/j.tourman.2014.06.015

Liu, B. (2010). Sentiment Analysis and subjectivity. In N. Indurkhya F. J. Damerau (Eds.), Handbook of Natural Language Processing (pp. 1–38). Chapman & Hall.

Liu, Y., Jin, J., Ji, P., Harding, J. A., & Fung, R. Y. K. (2013). Identifying helpful online reviews: A product designer's perspective. *Computer Aided Design*, *45*(2), 180–194. doi:10.1016/j.cad.2012.07.008

Lloret, E., Boldrini, E., Vodolazova, T., Martínez-Barco, P., Muñoz, R., & Palomar, M. (2015). A novel concept-level approach for ultra-concise opinion summarization. *Expert Systems with Applications*, *42*(20), 7148–7156. doi:10.1016/j.eswa.2015.05.026

Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., & McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. Association for Computational Linguistics. doi:10.3115/v1/P14-5010

Martin, L., & Pu, P. (2014). Prediction of helpful reviews using emotions extraction. In *Proceedings of 28th AAAI Conference on Artificial Intelligence*. AAAI Press. doi:10.1609/aaai.v28i1.8937

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. [PLEASE ADD JOURNAL AND VOLUME NUMBER HERE], 1–12.

Mudambi, S. M., & Schuff, D. (2010). What makes a helpful online review? A study of customer reviews on Amazon.com. *Management Information Systems Quarterly*, *34*(1), 185–200. doi:10.2307/20721420

O'Mahony, M. P., & Smyth, B. (2010). A classification-based review recommender. *Research and Development in Intelligent Systems XXVI: Incorporating Applications and Innovations in Intelligent Systems XVII*, *23*(4), 49–62. doi:10.1007/978-1-84882-983-1_4

Paget, S. (n.d.). Local Consumer Review Survey 2024: Trends, Behaviors, and Platforms Explored. Retrieved October 20, 2017, from https://www.brightlocal.com/learn/local-consumer-review-survey/

Pai, M.-Y., Chu, H.-C., Wang, S.-C., & Chen, Y.-M. (2013). Electronic word of mouth analysis for service experience. *Expert Systems with Applications*, *40*(6), 1993–2006. doi:10.1016/j.eswa.2012.10.024

Paivio, A., & Csapo, K. (1973). Picture superiority in free recall: Imagery or dual coding? *Cognitive Psychology*, *5*(2), 176–206. doi:10.1016/0010-0285(73)90032-7

Palakvangsa-Na-Ayudhya, S., Sriarunrungreung, V., Thongprasan, P., & Porcharoen, S. (2011). Nebular: A sentiment classification system for the tourism business. In *Proceedings of the 2011 8th International Joint Conference on Computer Science and Software Engineering, JCSSE 2011*. IEEE. doi:10.1109/JCSSE.2011.5930137

PowerReviews. (n.d.). Survey Confirms the Value of Reviews, Provides New Insights. Retrieved October 20, 2017, from https://www.powerreviews.com/blog/survey-confirms-the-value-of-reviews/

Rhee, H. T., & Yang, S. B. (2015). Does hotel attribute importance differ by hotel? Focusing on hotel star-classifications and customers' overall ratings. *Computers in Human Behavior*, *50*, 576–587. doi:10.1016/j.chb.2015.02.069

Riloff, E., & Wiebe, J. (2003). Learning extraction patterns for subjective expressions. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*. [PLEASE ADD PUBLISHER HERE]. doi:10.3115/1119355.1119369

Sacha, D., Stoffel, A., Stoffel, F., Kwon, B., Ellis, G., & Keim, D. (2014). Knowledge generation model for visual analytics. *Visualization and Computer Graphics. IEEE Transactions On*, *20*(12), 1604–1613.

Schmutz, P., Heinz, S., Métrailler, Y., & Opwis, K. (2009). Cognitive load in eCommerce applications—Measurement and effects on user satisfaction. *Advances in Human-Computer Interaction*, *2009*, 1–9. doi:10.1155/2009/121494

Shah, R. R., Yu, Y., Verma, A., Tang, S., Shaikh, A. D., & Zimmermann, R. (2016). Leveraging multimodal information for event summarization and concept-level sentiment analysis. *Knowledge-Based Systems*, *108*, 102–109. doi:10.1016/j.knosys.2016.05.022

Sinoara, R. A., Camacho-Collados, J., Rossi, R. G., Navigli, R., & Rezende, S. O. (2019). Knowledge-enhanced document embeddings for text classification. *Knowledge-Based Systems*, *163*, 955–971. doi:10.1016/j.knosys.2018.10.026

Stone, P. J., & Hunt, E. B. (1963). A computer approach to content analysis. In *Proceedings of the May 21-23, 1963, Spring Joint Computer Conference on - AFIPS '63 (Spring)*. Associaton for Computing Machinery. doi:10.1145/1461551.1461583

Tang, C., Mehl, M. R., Eastlick, M. A., He, W., & Card, N. A. (2014). A longitudinal exploration of the relations between electronic word-of-mouth indicators and firms' profitability: Findings from the banking industry. *International Journal of Information Management*, *36*(6), 1124–1132. doi:10.1016/j.ijinfomgt.2016.03.015

Wilson, T., Wiebe, J., & Hoffman, P. (2005). Recognizing contextual polarity in phrase level sentiment analysis. *Acl*, *7*(5), 12–21. doi:10.3115/1220575.1220619

Zhan, J., Loh, H. T., & Liu, Y. (2009). Gather customer concerns from online product reviews: A text summarization approach. *Expert Systems with Applications*, *36*(2), 2107–2115. doi:10.1016/j.eswa.2007.12.039

Zhu, F., & Zhang, X. (2010). Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of Marketing*, *74*(3), 133–148. doi:10.1509/jm.74.2.133

Pei-Hua Lee received her MD from the China Medical University. She was trained at the China Medical University Hospital for residency. She is now an attending radiologist at the China Medical University Hospital. Her research interests include clinical medical imaging, healthcare sciences, and data mining.

Yu-Kai Sun is currently an assistant professor in the Department of Power Mechanical Engineering at National Taitung Junior College, Taiwan. He received a PhD degree in Materials Sciences from National Taiwan University in 2022. His current research interests include materials sciences, industrial education, database management, and vehicle engineering.

Yin-Pei Ke received her MS degree in Graduate Institute of Information Management from National Chung Cheng University of Taiwan in 2018. Her research interests include data mining, information retrieval and text mining.

Pei-Ju Lee is currently an associate professor of the Institute of Data Science and Information Computing at National Chung Hsing University, Taiwan. She received a PhD degree in Information Sciences from University of Pittsburgh in 2015. Her current research interests include information fusion, data mining, database management, human-computer interaction, and human factor.