International Journal of Engineering & Technology, 7 (4.31) (2018) 341-347

SPC

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper



Know your Hotels Well! an Online Review Analysis using Text Analytics

Angela S.H Lee [1]*, Zaharin Yusoff[2], Zuraini Zainol[3], Pillai V[4]

[1]*School of Science and Technology, Sunway University, Malaysia*
[2]*Department of Computer Science, Faculty of Science and Defence Technology, National Defence University of Malaysia, Malaysia*
**Corresponding author E-mail: angelal@sunway.edu.my*

Abstract

Online travel forums have become an extremely popular platform for sharing travel information, with a large number of reviews being posted daily. Travel websites such as TripAdvisor and Booking.com have turned into very important resources for hotel operators and travellers alike, for promoting hotel rooms, choosing hotels as well as for soliciting and sharing feedback. Criticisms, compliments, dissensions, etc., are now accessible anytime and anywhere on the web, and can be readily amassed, while opinion mining techniques have developed rapidly. Together they provide the opportunity and capability to analyse and deduce factors that influence travellers in their choice of hotels. In this paper, we apply opinion mining on data collected from Tripadvisor websites. In total, 11,130 reviews on 4 hotels within the four-star and five-star categories in Kuala Lumpur are crawled, collected, and mined to identify the top-k most predominant information based on the most frequent and most related terms used in describing each of the chosen hotels. The results of this study would allow travellers to see the opinions of other travellers on these hotels, and hotel operators would be able to receive feedback to improve their services and in turn promote their hotels. This study is also carried out in view of future improvements in the techniques used and the analysis performed.

Keywords: *Opinion Mining; Text mining; Hotel Reviews; Customer Satisfaction; SAS TexMiner*

1. Introduction

Online travel forums have become an extremely popular platform for sharing travel information, with a large number of reviews being posted daily. Travel websites such as TripAdvisor and Booking.com have turned into very important resources for hotel operators and travellers alike. While hotel operators use the forums to promote their hotels as well as to solicit feedback in an effort to improve their services, travellers use them to decide on which hotels to stay at during their vacation or business travels, as well as to share their experience with other travellers on the services of the hotels.

A travel forum site, such as Tripadvisor, enables travellers to contribute descriptions, pictures, reviews and travelogues about the hotels they have stayed at and the tourist attractions they have visited. Individuals who visit a Tripadvisor site can gather numerous bits of information about a destination – lodging, amenities, food, best restaurants in the neighbourhood, places to visit, etc. It is also an online social network, enabling travellers to connect and be connected with other travellers around the world, to provide and read feedback/reviews on the satisfaction level for the hotels they stayed at. Content created by travellers on destinations have turned out to be increasingly essential. For instance, a recent study found that more than 60% of respondents checked online surveys, web journals, and other travellers' criticisms before purchasing an item or service, and an additional 20% said the reviews have significant influence on their purchase decisions [1].

Hotels operate in a competitive and dynamic environment [2]. They employ many different tools to assess and address customer satisfaction and their needs and wants in the hope of raising the level of satisfaction. They usually collect customer feedback by placing comment cards in the guest rooms, employing service recovery techniques to address in house service failures, distributing post departure satisfaction surveys, and introducing follow-up measures. Despite such efforts, the problem persists as guests tend to be unwilling to share their experience and provide feedback directly to the hotels. Instead, guests feel more comfortable and easier to share their experience via review websites, social media, blogs and other online platforms [3]. It is important for hotel managers to utilise customer review information that is available online in order to better understand their customers and improve their hotel performance. Furthermore, online media generally generate larger volumes of data, and as such opinion mining analyses of hotel reviews would be able to give a larger and better picture in terms of reviews on a given hotel. This will also ensure that each relevant factor can be studied and not missed out, thus potentially increasing business opportunities, customer satisfaction and customer loyalty [4]. Text analytics can show details such as which brands in an industry speaks the loudest, who are talking about which brand, products, industry, competitors, and what topics in an industry would generate the most buzz [4]. To have a competitive edge over other brands, the solution is to tap on all the information in the reviews, to understand travellers experiences, needs and wants, and to identify factors that affect customer satisfaction. Understanding the structure and a good portrayal of the online tourism domain is imperative for a better association of travel related data.

There are several ongoing research efforts to study hotel customer reviews, what customers really think about a hotel, and the factors that affect customer satisfaction [5][6]. Every traveller can rate hotels, leave comments and reviews on the good and bad experi-

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ences to share worldwide [7]. This paper reports on a text mining approach that allows for the extraction of meaningful patterns from large volumes of textual information, given that little attention has been afforded to hotel reviews in the Malaysian hotel industry [8]. This study is also carried out in view of future improvements in the techniques used and the analysis performed.

2. Literature Review

2.1 Online Reviews in the Hotel Industry

The hospitality industry, including the hotel industry, thrives on customer satisfaction, as that would lead to repeat purchases as well as good word-of-mouth promotions. The industry also depends a lot on customer feedback to improve services. Both would require the gathering of data from clients and other sources, and its analysis. In the early days, getting data was through manual means, such as comment cards, surveys, reviews, articles, site visit reports, etc. [9]. These will have to be manually perused, analysed, and some conclusions drawn. This exercise is very tedious and time consuming, and the data gathered is usually quite small in volume.
With the advent of the Internet, much more data is generated and is quite readily available. More and more services are also offered online. As a result, the customer's buying choice (or range) has expanded many-fold, and this would usually drive the customer to seek advice from various parties to make the final choice. On this point, it has been found that this decision-making is very much influenced by online networking, where many reviews of products and services are shared by netizens [9] [10].
With the rise of online tourism services, the hotel industry is said to be strongly affected by eWOM (Electronic Word-of-Mouth). There have been studies on the influence of eWOM in general, on discovering the factors that lead travellers to writing online hotel reviews, and on the impact of online hotel reviews [11][12].
In sum, for product and service providers, such as hotels, there is a clear need to keep track of online reviews given their major influence on would-be customers, including travellers.

2.2 Hotel Guest Satisfaction

The study of customer satisfaction and dissatisfaction has been an important research study among scholars from various disciplines, such as marketing, IT, Psychology, and Hospitality [4]. A satisfied customer will repeat purchases, whereas a dissatisfied customer will not only not repeat the purchases but may also spread negative promotions on the product or service.
Hotel guests generally look at several factors to evaluate the quality of service they receive during their stay, and that customer satisfaction may be affected by both tangible and intangible aspects of service quality [13] [14]. The growth of the Internet has enabled a rapid expansion of referrals by loyal customers through eWOM. The literature supports the fact that service recovery strategies increase customer loyalty [15], but on the other hand, if complaints are not addressed, it may result in dissatisfaction and low repeat purchases [16].
Online customer reviews empower guests by allowing them to access more accurate, up-to-date information about products and services [17]. From the operator perspective, user generated content creates opportunities for hotels to gain a better understanding of their guests. There have been studies on online complaints and reviews on hotels, and several categories for improvement were identified, in particular the physical attributes of hotel rooms, and the quality of the ambience of the room and hotel services, such as whether the hotel staff are friendly or rude [18][14][8].

2.3 Text Mining

Online reviews and other online materials can be very voluminous and of a very wide variety. It would be quite impossible to peruse all the documents manually. This is where text mining would be of great help. Text mining, sometimes also referred to as text analytics, is the process of deriving high-quality information from text, which is obtained through the formulation and display of patterns and trends in the input text.
Text mining has been used to classify pleasant reviews by satisfied customers and unpleasant reviews by dissatisfied customers [19]. The researchers utilize shallow natural language processing techniques as the means to identify emotion-based review categories. They suggest that hotel guest reviews can serve as a complementary and quite reliable source for the evaluation of hotel quality. Studies found certain common topics mentioned in the reviews, namely: hotel location, room size, staff, cleanliness, breakfast, comfort, temperature, cleanliness, and maintenance [20][8].
To date, not much emphasis has been put into studying online reviews on local hotels, which is the principal motivation for this paper.

3. Methodology

This study crawled data from TripAdvisor websites. The websites are free websites that allow travellers to access information about hotels, vacation, rental, as well as to express their reviews on hotels. The data on travellers' feedback is unstructured text. In total, 11,130 reviews on 4 hotels within the four-star and five-star categories in Kuala Lumpur are collected.
For carrying out the data analysis phase, this study uses the SAS Text Miner, which is a full integration add on to the SAS Eminer. The Text Parsing node is first selected to break every sentence into separate terms. This node builds a corpus dictionary and performs word stemming to associate terms that are from the same verb(s). The output of this node will be the input into to the next node, which is the Text Filer. This latter node filters associated terms into a dictionary-based table where multi-terms and spell-checking are applied onto the documents. With the spelling dictionary specified, the spellchecker will then list out the misspelt words from the highest to the lowest frequency. [This is just part of data cleansing, whereby the misspelt words may be manually corrected, or simply ignored] [21][22][8].
Based on the co-occurrence of words in the documents, concept links are then formed, which helps in understanding the connections amongst words. The software determines the list of key concepts and the relevant sub-concepts associated to each key concept. This is presented in the form of visual concept maps that show all the connections of each sub-term to the corresponding main terms.
Next, we connect the Text Cluster node to the Text Filter. This will discover themes and assign each document to one of these themes. The function of a cluster node is to split the entire corpus of documents into mutually exclusive clusters. Key descriptive terms from each of the documents are then automatically displayed in each cluster. The last node used in our analysis phase is the Text Topic. The Theme or underlying idea that occurs in a document will be derived from this node. For example: in a student evaluation survey, participants will sometime respond with: "The class can be conducted in a better way. Here are two suggestions: 1) Lecturer can relate real life case examples from industry to teach in the class; 2) More interactive teaching. In this case, the Text Cluster node will identify two themes: 1) Improve classroom teaching 2) Interactive. This Text Cluster node will assign comments related to the above themes, whereas the Text Topic node will find terms that occur frequently together within a single document. This phase allows this study to analyze and examine the key issues formed from the travellers' reviews, while the visual concept maps are used to link all the key terms and sub-terms

together. Table 1 shows the number of reviews and hotel names used for the data analysis in this study.

Table 1: Summary of Hotel Name and Reviews Collected

Hotel Type	Hotel names and No of Reviews	
	Name of Hotel	No of Reviews (2012-2017)
Four Star	Aloft KL Sentral	5501
	Novotel Kuala Lumpur	
	Hilton Kuala Lumpur	5629
Five Star	Shangri-La Hotel KL	
	Total Reviews Collected	11,130

4 Analysis and Results

This section explains the analysis of the results derived from the text analytics phases. Key terms are ranked using weights and this will be analyzed and explained based on the travellers' comments. From their comments, issues can be highlighted and improvements can be recommended. Table 2 shows the frequency of terms, number of documents, and weights of related terms based on the frequency of terms listed from the highest to the lowest.

Then, Figure 1 and Figure 2 show the corresponding bar charts of the factors by frequency of terms for four-star hotels and five-star hotels respectively. From these Figures, the highest mentioned term is comfort (staff friendliness), while the lowest is noise.

Concept links show the quality of relationship of related terms. The thickness of the line characterises the strength of the associations, with a thicker line showing a more grounded relationship between terms, in contrast to a more slender line. Each connection node can likewise be extended to view its sub connected node to have a bigger picture of the connection between the two noteworthy nodes. Concept links are discussed next in 4.1 and 4.2 below.

Table 2: Summary of Factor in Terms with Frequency of Terms, Documents and Weight

Hotel Type	Factors in Terms	Frequency of terms	Documents	Weight
Four - Star Hotel	Comfort (Staff Friendliness)	3633	2717	0.096
	Location	3018	2632	0.091
	Comfort (Service)	2230	1685	0.15
	Facilities (Pool)	1799	1454	0.164
	Cleanliness	1654	1452	0.161
	Facilities (Restaurant)	1071	874	0.224
	Price	843	719	0.243
	Noise	233	192	0.399
Five - Star Hotel	Comfort (Staff Friendliness)	3904	2813	0.094
	Comfort (Service)	3130	2199	0.124
	Facilities (Pool)	2562	1880	0.142
	Location	2168	1937	0.129
	Facilities (Restaurant)	1581	1254	0.185
	Cleanliness	1325	1134	0.194
	Price	698	577	0.272
	Noise	165	137	0.441

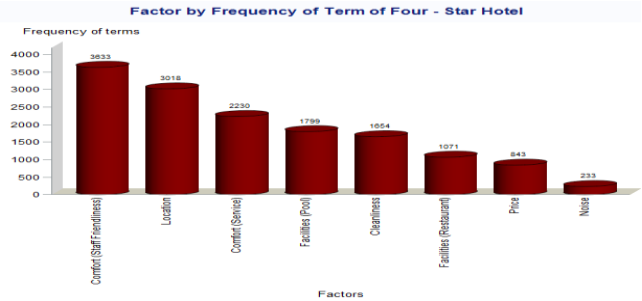

Fig.1: Factors by frequency of terms on Four-Star Hotel

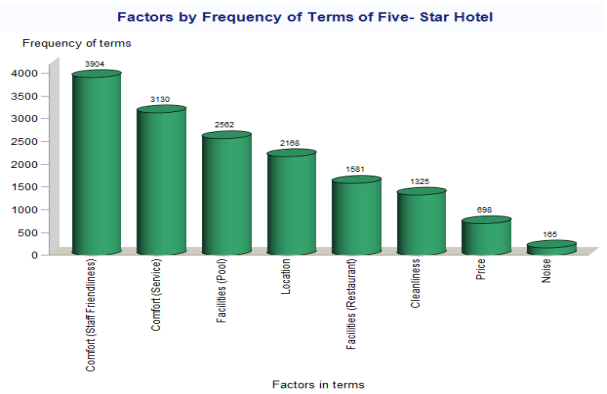

Fig.2: Factors by frequency of terms on Five-Star Hotel

4.1 Four Star Hotels Concept Link

Figure 3 shows the number of main terms derived from the concept links of the most popular sub-terms generated from the main terms. Figure 4 shows the concept links of the main term " Location" of Four Star Hotels. The term "location" is chosen to additionally explore alternate terms that are related with the term "location". Figure 4, is a context diagram derived for four-star hotels with terms associated with "location", which are "bukit bintang", "walk", "bukit", strategic", "klcc", "great", "bintang" and "mall". The terms that are unequivocally connected with the term "location" are "strategic", "great", and "bukit bintang".

Terms	Sub Terms (spokes)							
Location	strategic	great	bukit bintang	walk	klcc	bukit	bintang	mall
Price	reason	worth	range	bar	decent	little	pay	buffet
Facilities (Restaurant)	buffet	food	meal	road	shop	nook	eat	chinese
Facilities (Pool)	swim	relax	view	amazing	gym	rooftop	top	express
Noise	construction	noisy	sleep	loud	toilet	pipe	water	hear
Cleanliness	comfortable	room	nice	neat	modern	spacious	tidy	quick
Comfort (Staff Friendliness)	friendly	courteous	attentive	polite	smile	always	helpful	help
Comfort (Service)	housekeeping	excellent	customer	slow	front	order	standard	level

Fig.3: Sub Terms from the concept links for Four-Star Hotels

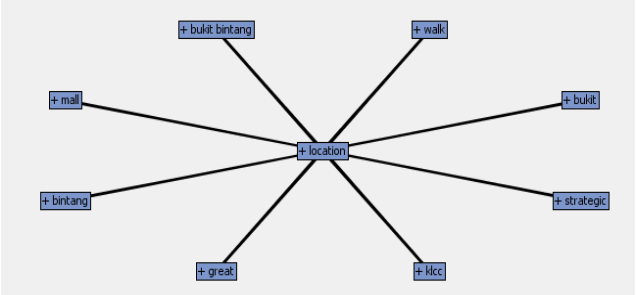

Fig. 4: Concept Links of the term "location" in Four-Star hotels

4.2 Five Star Hotels Concept Link

Figure 5 shows the number of main terms derived from the concept links and the most popular sub-terms generated from the main terms. Figure 6 shows the concept link for the main term " location". The terms that are associated with "location" are "convenient", "great", "walk", "station", "monorail", "central", "train" and

"prefect". The strongly associated terms with "location" are "great", "convenient", and "perfect".

Terms	Sub Terms (spokes)								
Location	great	convenient	perfect	walk	station	monorail	central	train	-
Price	reasonable	expensive	cost	high	star	cheap	pay	half	-
Facilities (Restaurant)	buffet	bar	japanese	main	eat	chinese	meal	food	-
Facilities (Pool)	swim	sun	jacuzzi	gym	monorail	big	tower	area	water
Noise	construction	loud	disturb	hear	noisy	first	floor	door	-
Cleanliness	spotlessly	tidy	comfortable	neat	big	room	bed	spacious	-
Comfort (Staff Friendliness)	courteous	professional	polite	helpful	smile	always	attentive	help	-
Comfort (Service)	impeccable	food	level	excellent	offer	prompt	customer	order	-

Fig.5: Sub-Terms from the concept link for Five-Star Hotel

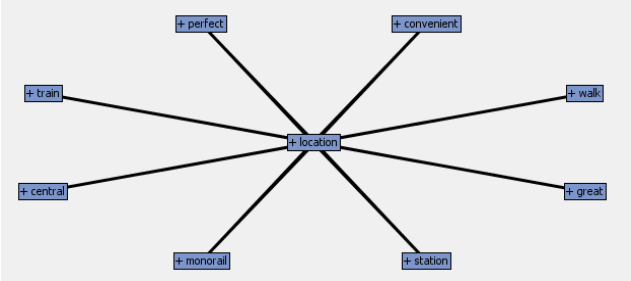

Fig. 6: Concept Links of term "location" in Four-Star hotels

4.3 Text Analytics Using R-Mining

Just to view results using a different tool other than the SAS Text Miner, R-Mining is used on the same data (or rather a subset of it), focusing on the capabilities of capturing the tacit knowledge within the reviews on the experiences of the travellers. Figure 7 illustrates the framework of the Context-Based Keyword Pattern Cluster Analysis (CBKPC) using R-Mining [23, 24]. The analysis is carried out only for Four-Star hotels as this part is only for illustrative purposes.

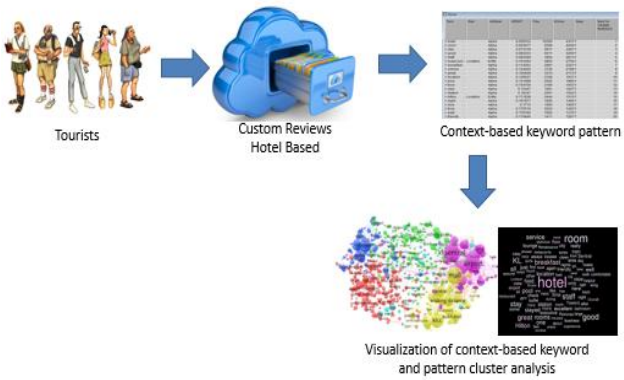

Fig. 7: Framework for Context-based Keyword Pattern Cluster Analysis

The dataset consists of 2500 reviews (extracted from the 11.130 used for the SAS Text Miner). A pre-processing task is required before the text analysis can be carried out, which is to remove errors and to prepare the data for further analysis. This will also improve the efficiency of the text mining process due to the higher quality input text. In this experiment, the pre-processing of data consists of 2 tasks, namely cleansing and structuring. The cleansing task consists of two sub-steps: (i) convert all text to lower case for standardisation, (ii) remove punctuation marks, white spaces, symbols, numbers, etc. After these, stop words are removed and word stemming is carried out on the data.

Terms	Docs hotel_v2.txt
breakfast	1338
good	1816
hotel	3694
locat	1843
novotel	1183
room	4086
servic	1092
staff	1679
stay	2257
walk	981

Fig. 8: Concept Links of the term "location" in Four-Star hotels

Figure 8 shows 10 keywords with the highest frequency terms, which are also given as a bar chart in Figure 9. The keywords are as follows: "breakfast" (1338), "good" (1816), "hotel" (3694), "locat" (1843), "novotel" (1183), "room" (4086), "service" (1092), "staff" (1679), "stay" (2257), "walk" (981).

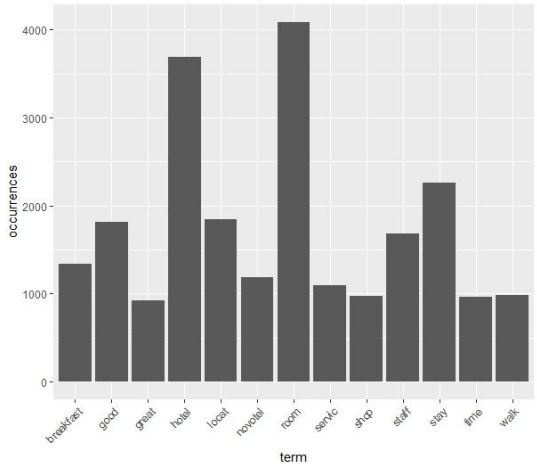

Fig. 9: Concept Links of the term "location" in Four-Star hotels

The frequently used keywords can also be visualised in the form of a word cloud. As presented in Figure 10, the size of the displayed words corresponds to the frequency of the keywords. From the figure, it can be concluded that the keywords "hotel", "room", "stay", "locat", "good", and "staff" are among the top six most important keywords in documents.



Fig.10: Word cloud with keywords occ

4.4 Text Clustering on Four-Star Hotels

Continuing the mainstream discussion from 4.2, cluster IDs are then formulated based on the descriptive terms derived from the text topic node. Table 3 shows the clustering results for four-star hotels. Themes are generated by the Text Miner as represented by

the text clusters. Within the text clusters, it is found that two clusters are very predominant, which are Cluster ID 1 and Cluster ID 2. Based on the descriptive terms, it clearly shows that travellers give higher attention to amenities/facilities and the service provided by the four-star hotels.

Table 3: Cluster ID and Descriptive Terms for Four Star hotels

Cluster ID	Descriptive Terms	Frequency	Percentage
1	+pool +bar +aloft +view +back +check +wait +dont +day +train +bed +experience +spacious +floor +time	1503	27%
2	+hotel +good +staff +stay +location +breakfast +service +great +friendly +shop +food +helpful +excellent +value +business	1076	20%
5	+'kuala lumpur' +sentral +station +train +pool +airport +mall +'kl sentral' +express nu +spacious +aloft +view +locate +easy	879	16%
6	+front +desk +water +shower +wait +bathroom +bed +old +guest +small +check +book +minute +day +floor	640	12%
3	+walk +location +shop +tower +mall +bintang +bukit +minute +distance +pavilion +petronas +klcc +twin +pavillion +'bukit bintang'	595	11%
7	+novotel +book +accor +bad +executive +member novotel +old +pay +card +lounge +standard +dont +guest +experience	544	10%
4	+lumpur +kuala +'kuala lumpur' +city +visit novotel +petronas +tower +distance +novotel +locate +value +stay +hotel +excellent	264	5%

4.5 Text Clustering on Five-Star Hotels

In contrast, for Five-Star hotels, the more predominant clusters are in the areas of bathroom cleanliness, lounge area and the services provided in the hotel. Table 4 shows the descriptive terms from the cluster node.

Table 4: Cluster ID and Descriptive Terms for Five Star hotel

Cluster ID	Descriptive Terms	Frequency
4	+check +bathroom +executive +bed +hilton +pay +book +back +first +floor +didnt +lounge +'executive lounge' +dont +bad	1660
5	+hotel +room +good +stay +staff +service +great +pool +food +location +friendly +excellent +restaurant +recommend +helpful	1170
3	+tower +walk +view +wing +petronas +east +twin west +minute +monorail +distance +renaissance +pool +breakfast +location	974
2	+'kuala lumpur' +train +station +airport +sentral +express +hilton +easy +city +'kl sentral' +klia +locate +business +convenient +shop	740
7	+star +bad +front +dont +pay +guest +didnt +manager +renaissance +book +check +first +look +day +experience	483
1	+lumpur +kuala +'kuala lumpur' +hilton +renaissance +visit +city +stay +friendly +hotel +locate +recommend +staff +special +view	307
6	+staff +lounge +executive +hilton +team +always +feel +special home +smile +'executive lounge' +manager +front +guest +service	295

A quick comparison can be made between the predominant clusters for the four-star and those for the five-star hotels. The two main clusters for both are facilities and customer service. The terms related with the two main clusters are frequently utilized since they have connections to the hotel's facilities, customer service and location. The highest recurrence of a cluster is Facilities, with 1503 for four-star hotels and 1660 for five-star hotels, while the lowest recurring primary cluster is Location for both four- and five-star hotels, with 879 and 974 respectively. This also shows that the recurrence rate for five-star hotels is higher than for four-star hotels, at least as far as these three primary clusters are concerned. In both cases, it can be readily concluded that these three main factors play critical roles when an individual is reserving or picking a place to stay on an excursion or business trip.

5 Discussion

5.1 Key Factors (Further Details)

Given the 8 main factors obtained from the analysis, some further details may be obtained by tracing the links from these factors (and their themes) to the actual review documents (unfortunately, the perusal will have to be carried out manually). The following are some extracts for each of the 8 main factors.

5.1.1 Comfort (Staff Friendliness)

As mentioned earlier, hotels thrive on customer satisfaction, and this is attained not only via the services rendered, but also on the way they are rendered. For this latter part, hotel guests would want to feel very comfortable with the hotel staff. Most guests would rely on the services of the concierge from handling the luggage upon arrival to settling down in the room. The concierge is a very significant asset and need to be strategically positioned in the lobby and are staffed with very competent and friendly personnel. The front desk staff also need to be very competent and friendly as they need to tend to inquiries, arrange for transport, be part tourist guide, etc. Even cleaning staff and restaurant waiters are included in travellers' reviews. A good experience from the hotel is set to warm the hearts of guests and would encourage returns as well as to spread positive word-of-mouth recommendations to friends and colleagues.

5.1.2 Comfort (Service)

Travellers' reviews would always touch on the quality of services at the hotel. Dependability and timeliness are very much expected, and any failure at any level would frustrate guests. The expectation level tends to be very high (when compared to many other products and services) as guests feel that they are not only paying for the lodging but also for the services and comfort. This is particularly true for four- and five-star hotels.

5.1.3 Location

The hotels involved in this study are at the very centre of Kuala Lumpur, within the well-known Golden Triangle, i.e. between the Petronas Twin Towers and Bukit Bintang. As such, hotel guests would consider the location as very strategic, and many world-class hotels are located there. Hotels are therefore expected to have the best facilities, readily available meeting venues, transport, excursions, easy access to shopping centres, and other attractions.

5.1.4 Facilities (Restaurant)

Having good restaurants in the neighbourhood is a very major factor for hotel guests. They also expect variety, At the very least, there should be at least one eatery available at any one time (preferably one that is open 24/7) within the area, if not provided by the hotel itself. [It is to be noted that this is not mandatory for four-star hotels, but many guests are not aware of that.]

5.1.5 Facilities (Pool)

One main consideration for choosing a hotel is the availability of a swimming pool. As this is mandatory for four- and five-star hotels, other expectations (or comparisons) arise, usually in the form of services around the pool area (drinks, food, sauna, gymnasium, etc.). Rooftop pools with a view are also getting very popular with hotel guests.

5.1.6 Cleanliness

For the hotels in this study, the majority of visitors are very content with the general cleanliness within and outside the hotel area. General expectations are the following (apart from a television, telephone, and refrigerator) – a spotless room, clutter and clamour free room, clean draperies, clean toilets, availability of heated water, and a safety door. These also tend to be expected even of budget hotels.

5.1.7 Price

For four- and five-star hotels, the more positive terms used in reviews for price are: "reasonable" and "worth it". On the end, it would be an outright "expensive". These terms are used principally for room rates, but some also refer to restaurant food, drinks from the bar, shopping outlet items, etc.

5.1.8 Noise

All the hotels in this study often do get upsetting reviews on this aspect, due partly to roadworks or building renovations nearby, partly to some talkative cleaning staff, but mostly due to the boisterous behaviour of other guests. Noise can also be due to rattling ventilation systems (fault of a lack of maintenance).

5.2 Future Work

As mentioned in the Introduction, apart from carrying out an analysis of online reviews on hotels, this study is also carried out

in view of future improvements in the techniques used and the analysis performed. For this purpose, Figure 10 gives a general view of a text mining process.

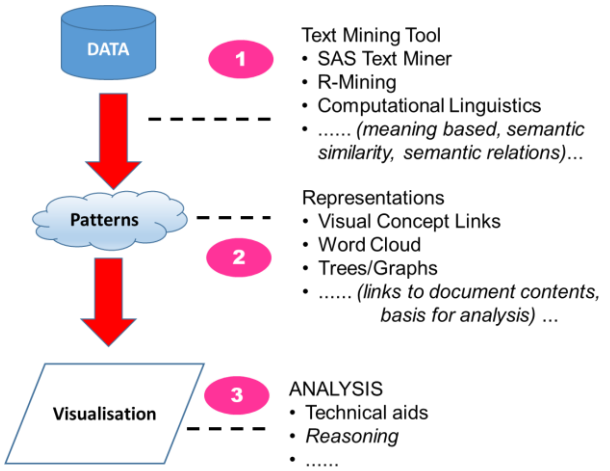

Fig.11: General Text Mining Process

1) From a given data set (text), a **Text Mining tool** is applied to the data (some may require pre-processing, usually in the form of cleansing and restructuring).
2) The output is a set of **patterns**, typically in the form of a set of primary terms (e.g. most frequent), with relations to other terms. The patterns may then be visualised in some form or another to help in understanding.
3) The patterns are then **analysed** for some conclusions to be drawn or some recommendations made.

Future improvements to the text mining process will have to be within the realms of (1), (2) and (3) above.

For (1), the text mining tool used for this study is the SAS Text Miner. An R-mining tool is also used for a view of possible alternative results. Both seem to be based on two computational linguistic algorithms that were first developed in the late 70s and early 80s, namely the word (frequency) count algorithm and the concordance (co-occurrence) algorithm. Since then, frequency counts in documents may have moved towards taking into account semantic similarities between words (akin to an advanced thesaurus), e.g. in determining the main or primary terms in a document (c.f. the 8 factors in Fig.1). Several tools can now be used to measure the semantic similarity between concepts, such as WNetSS API [28], which gives semantic similarity measurements based on the WordNet semantic resource. The relations amongst words (c.f. concept links) are more complex, but one may envisage the use of semantic relations to measure the affinity for association between two terms (c.f. valency).

For (2), the representation adopted in this study is the visual concept link for the SAS Text Miner, which is quite equivalent to the word cloud for the R-mining approach. These are quite adequate, just as earlier computational linguistic approaches would use trees and graphs. What would be very desirable are more sophisticated links from the representation to the detailed contents of the original documents. This will not only minimise the need for manual perusal of the documents to extract further information, it will also form a good technical aid for the analysis in (3).

For (3), the analysis carried out in this study is essentially what have been provided by the text mining tool (c.f. primary terms and concept links), plus some additional (manual) effort to trace the links to the original documents to provide some supporting details. As mentioned under (2), it would be good to have this tracing more automated. Better still, the supporting details should be predicted from the themes (clusters) via some mechanism to be formulated. Above all, a good analysis should result in additional information or knowledge, not just those that can be directly attributed to the text miner. As such, some reasoning process should be looked for in this area. This latter initiative would bring text mining to a totally different and definitely higher level.

6. Conclusion

This paper reports on a text (opinion) mining application on data consisting of online reviews on travel forums submitted by travellers using hotels. This is an effort to demonstrate the opportunity and capability to analyse and deduce factors that influence travellers in their choice of hotels. Eight main factors were identified, supported by their related themes (clusters) and links to the original documents for extracting further details if needed. The demonstration was carried out on data from Tripadvisor websites on 4 four-and five-star hotels in Kuala Lumpur. The study also shed some light into future work for improving the text mining process, namely in the areas of the text mining tool, the representation of the results, and most importantly the final analysis.

References

[1] Werbler, C. Online Consumer Reviews Significantly Impact Consumer Purchasing Decisions, New Opinion Research Corporation Survey Finds, (2008). [online] Businesswire.com. Available at: https://www.businesswire.com/news/home/20080624005206/en/Online-Consumer-Reviews-Significantly-Impact-Consumer-Purchasing.

[2] Verma, R., Victorino, L., Karniouchina, K and Feickert, J. Segmenting Hotel Customers based on the technology readiness index. Cornell Hospitability Report, (2007), pp, 4-13.

[3] Berezina, K., Cobanoglu, C., Miller, B. L., & Kwansa, F. A. The impact of information security breach on hotel guest perception of service quality, satisfaction, revisit intentions and word-of-mouth. International Journal of Contemporary Hospitality Management, (2012), 24(7), 991–1010.

[4] L. McGarrity, "What Sentiment Analysis Can Do for Your Brand," Marketing Profs, (2016). [Online]. Available: http://www.marketingprofs.com/opinions/2016/29673/what-sentiment-analysis-can-do-for-your-brand. [Accessed 11 April 2017].

[5] M. Darini and F. Khozaei, "The study of factors affecting customer's satisfaction with three star hotels in Dubai," International Journal of Advanced Engineering, Management and Science,(2016),2 (2), pp. 21-24.

[6] Z. Xiang, Z. Schwartz, J. H. G. Jr and M. Uysal, "What can big data and text analytics tell us about hotel guestexperience and satisfaction?," International Journal of Hospitality Management, (2015), vol. 44, pp. 120-130.

[7] R. Byng, "15 Years of TripAdvisor: How 200 Million Reviews Changed the Way We Travel," Independent UK, 25 February 2015. [Online]. Available: http://www.independent.co.uk/travel/15-years-of-tripadvisor-how-200-million-reviews-changed-the-way-we-travel-10070175.html. [Accessed 14 April 2017].

[8] Nicholas, C. K. W., & Lee, A. S. H. (2017). Voice of customers: Text analysis of hotel customer reviews (cleanliness, overall environment & value for money). In Proceedings of the 2017 International Conference on Big Data Research, (2017), pp. 104-111. ACM.

[9] Lau, K., Lee, K. and Ho, Y. Text Mining for the Hotel Industry. Cornell Hotel and Restaurant Administration Quarterly, (2005), 46(3), pp.344-362.

[10] Gémar, G. and Jiménez-Quintero, J. (2015). Text mining social media for competitive analysis. Tourism & Management Studies,(2015),11(1),pp.84-90. [Online] Available at: http://www.academia.edu/16536522/Text_mining_social_media_for_competitive_analysis.

[11] Cantallops, A. S., & Salvi, F. New consumer behavior: A review of research on eWOM and hotels. International Journal of Hospitality Management, (2014), 36, 41–51.

[12] Litvin, S. W. , Goldsmith, R. E. , & Pan, B. Electronic word-of-mouth in hospitality and tourism management. Tourism management, (2008), 29 (3), 458–468

[13] Pizam, A., & Ellis, T. Customer satisfaction and its measurement in hospital-ity enterprises. International Journal of Contemporary Hospitality Management, (1999),11(7), 326–339.

[14] Ekiz, E., & Au, N. Comparing Chinese and American attitudes towardscomplaining. International Journal of Contemporary Hospitality Management,(2011), 23(3), 327–343
[15] Cranage, D. A., & Mattila, A. S. Service recovery and pre-emptive strate-gies for service failure: Both lead to customer satisfaction and loyalty, but fordifferent reasons. Journal of Hospitality & Leisure Marketing, (2005), 13(3/4), 161–181
[16] Mattila, A. S., & Mount, D. J. The impact of selected customer characteristicsand response time on e-complaint satisfaction and return intent. International Journal of Hospitality Management, (2003), 22(2), 135–145.
[17] Kucuk, U. S., & Krishnamurthy, S. An analysis of consumer power on the Internet. Technovation,(2007), 27(1/2), 47–56
[18] Barreda, A., & Bilgihan, A. An analysis of user generated content for hotelexperiences. Journal of Hospitality and Tourism Technology, (2013), 4(3), 263–280.
[19] García-Barriocanal, E., Sicilia, M.-A., & Korfiatis, N. Exploring hotel ser-vice quality experience indicators in user-generated content: A case usingTripadvisor data. In Mediterranean Conference on Information Systems, (2010) Proceedings, 33.
[20] O'Connor, P. Managing a hotel's image on TripAdvisor. Journal of Hospitality Marketing & Management, (2010), 19(7), 754–772
[21] A. L. S. Hoong, T. M. Lim, S. K. Leow and J. L. R. Aun, "A study on the use of "Yams" for enterprise knowledge sharing," Digital Information and Communication Technology and it's Applications (DICTAP), 2012 Second International Conference *on*, Bangkok, 2012, pp. 183-188.doi: 10.1109/DICTAP.2012.6215348
[22] Lee A.S.H, K.L Chong & Nicholas, C. K. W. (2018). "Opinion-Seer: Text Visualization on Hotel Customer Reviews of Services and Physical Environment". 9th ICATSE Conference on Information Science and Applications. (In Press)
[23] Zainol Z, Nohuddin PNE, Noormanshah WMU & Hanafi AHM, M. "Visualization of context-based keyword pattern cluster analysis on tacit knowledge among officer cadets at Universiti Pertahanan Nasional Malaysia (UPNM)". *Advanced Science Letters*, Vol.24, No.3, (2018), pp.1550-1554, available online:https://www.ingentaconnect.com/contentone/asp/asl/2018/00000024/00000003/art00003, last visit:10.06.2018
[24] Zainol Z, Nohuddin PNE, Mohd TAT & Zakaria O, "Text analytics of unstructured textual data: a study on military peacekeeping document using R text mining package". *Proceeding of the 6th International Conference on Computing and Informatics (ICOCI17)*, (2017), pp. 1–7.