The use of post-purchase online consumer review in hotel attributes study was still scarce in the literature. Arguably, post purchase online review data would gain more accurate attributes that consumers actually consider in their purchase decision. This study aims to extract attributes from two samples of five-star hotel reviews (Jakarta and Singapore) with text mining methodology. In addition, this study also aims to describe positioning of five-star hotels in Jakarta and Singapore based on the extracted attributes using Correspondence Analysis. This study finds that reviewers of five star hotels in both cities mentioned similar attributes such as service, staff, club, location, pool and food. Attributes derived from text mining seem to be viable input to build fairly accurate positioning map of hotels. This study has demonstrated the viability of online review as a source of data for hotel attribute and positioning studies.

Keywords: text mining, correspondence analysis, hotel attribute, Jakarta, Singapore

Introduction

In the past few years, online Word of Mouth Communication (WOM) has become very pervasive in marketing. Hospitality industry becomes a good example of online WOM pervasiveness over the Internet. For instance, a traveler can get numerous reviews and tips about hotels and attractions around the world from review sites such as Tripadvisor.com, Booking.com, or Expedia.com. In addition, there are many other types of online information available online, such as social media posts made by individuals.

The significance of social media and other online resources in marketing of hospitality and leisure industry has been documented in the literature. Xiang and Gretzel (2010) suggest that social media had increasingly become an important source of information for travelers as they search for information on line before making their travel purchase decision. Travel reviews are used during travel idea generation stage as well as during final purchase decision. Most review readers perceived travel reviews help them in learning about travel products and destinations (such as accommodations or transportation), evaluating alternatives, and
avoiding undesirable places (Gretzel & Yoo, 2008). In support of Xiang and Gretzel (2010); Callarisa, García, Cardiff, and Roshchina (2012) also report that online reviews in social network affected consumer travel decision. Furthermore, they are considered more credible than other means of marketing communication.

Large amount of travel and accommodation information available online and widely used by travel consumers provide research opportunity for hospitality industry and academic community. One area where analysis of online information is useful is in hotel industry. According to Zhou, Ye, Pearce, and Wu (2014), online reviews could be used in hotel satisfaction studies to identify and rate aspects (attributes) of hotels that drive consumer satisfaction. Zhou et al. (2014) propose that internet-based hotel review offer efficient and data-rich alternative to survey based methods, which have some potential issues related to data collections such as biases, incomplete data, and expensive data collection process. In fact, the use of online consumer reviews in hospitality and travel research has become common practice in the last few years (see Table 1 for example of hospitality research involving online consumer reviews).

Even though it has full of potential to become rich source of research data, analysis of online WOM, such as travel reviews, can be challenging due to enormous quantity of unstructured consumer information. To deal with this, text mining technique can be used to identify important market terms and themes such as brand names and product features from online forums and social media. Netzer, Feldman, Goldenberg, and Fresko (2012) mentioned several past studies that applied text mining on product reviews, for instance to estimate consumer preferences, to reveal market structures and to predict firms’ stock performances.

Thus, the aim of this research is twofold. First, to explore the viability of online hotel review as source of valid data regarding hotel attributes whereby the author would examine reviews of five star hotel reviews in two different locations from a travel review website and then identify hotel attributes from the reviews using text mining methodology. Since this study is exploratory in nature, the author would compare the results with hotel attributes identified in past studies. In this regards, evaluation regarding similarity or differences between this study and past studies (that applied survey/ in-depth interview) can be made to prove the validity of attributes gained from text mining application. Second, the author would like to explore the viability of using attributes generated from text mining hotel reviews to construct positioning map of hotels in both cities via correspondence analysis. The author argues that positioning map generated through input from text mining of reviews can accurately reflect consumer perception, thus providing hotels with insights regarding their position relative to competitors.

Literature review

On-line word of mouth communication and hotel marketing

Online Word of Mouth Communication (WOM) has been reported to aid in consumer decision making and used by consumers as means to manage risk (O’Connor, 2010). Godes and Mayzlin (2004) report that WOM is often a prominent driver of consumer behavior such as decision to adopt a new technology, or decision to watch a new TV show. Gruen, Osmonbekov and Czaplewski (2006) propose that WOM have significant impact on consumer choice and post purchase product perceptions, as well as more effective than traditional marketing tools.

The Role of user generated contents and social networks in marketing have been documented. Decker and Trusov (2010) report that user generated contents such as consumer reviews can strongly influence decision making of potential buyers who search for information online prior to purchase. Similarly, Callarisa et al. (2012) report that online reviews in social network can affect consumer travel decision as well as considered to be more credible than other means of marketing communication.

Hotel and tourism sectors are also influenced by user generated contents. Social media is increasingly becoming important source of information for travelers as they search for information online before making their travel
| Author(s)                          | Year | Context                                                                 | Method                                                                 |
|-----------------------------------|------|--------------------------------------------------------------------------|------------------------------------------------------------------------|
| Xiang, Schwartz, Gerdes, and Uysal| 2015 | Used 60,648 customer reviews of 10,537 hotels in 100 largest US Cities to reveal several dimensions of hotel guest experience. In addition, the authors tested the association between the dimensions of guest experience and satisfaction. | Data downloaded using web crawler and processed using text mining method. Subsequently, data were factor analyzed and the results used as input for ANOVA to test association between the dimensions and guest satisfaction |
| Li, Law, Vu, Rong, and Zhao        | 2015 | Used 118,300 records from online reviews of 1740 hotels located in Hong Kong, Singapore, Shanghai, Bangkok and Sydney, the authors identified emergent hotel features | Text mining and specifically applied Emergent Pattern Mining to identify emerging patterns from the reviews. |
| Zhou, Ye, Pearce, and Wu          | 2014 | Used over 1000 user generated review of 4 and 5 star hotels in Hangzhou, China from Agoda.com website to identify 17 attributes influencing customer satisfaction, classified based on types of influence exerted. | Content analysis using two researchers to sample the comments and coded the review content into six broad themed categories |
| Xie, Zhang, and Zhang             | 2014 | Used consumer review data, management responses and hotel information from 843 hotels in five major hotel markets in Texas, USA downloaded from Tripadvisor.com | Using volume of review, valence of overall review, valence of individual attribute rating items, and review variation from consumer review as input for linear regression analysis |
| Li, Ye, and Law                    | 2014 | Revealed attributes that determine customer satisfaction from 774 star rated hotel in China | Automated Text mining and Hierarchical Clustering |
| Ye, Law, Gu and Chen              | 2011 | Used hotel bookings and review information of 1639 hotels in 10 randomly selected cities in China from Ctrip.com website to identify impact of online WOM on sales in the tourism and hotel industries. Valence and variance of online consumer generated review and number of online bookings were used as variables | Review and hotel booking data downloaded with web crawler were processed and then analyzed using Log-Linear Regression Model |
| O’Connor                          | 2010 | Used sample of 100 randomly selected hotels (from which five most current reviews were examined) in London from Tripadvisor.com website to identify factors that cause reviewer satisfaction and dissatisfaction | Content analysis supported by NVIVO7 software with two coders identifying the themes and coded the reviews into the themes |
| Ye, Zhang, and Law                 | 2009 | Used corpus of 591 negative reviews and 600 positive reviews of seven most popular US and Europe travel destinations downloaded from Yahoo.com’s travel column, the authors compared three methods of sentiment analysis. | Sentiment analysis |
purchase decision (Gretzel & Yoo, 2008; Xiang & Gretzel, 2010). According to Gretzel and Yoo (2008), travel reviews are used during travel idea generation stage as well as during final purchase decision, moreover, most review readers perceive travel reviews help them to learn about travel products and destinations (e.g. accommodations), to evaluate alternatives, and to avoid undesirable places. Online travel review is also perceived as a way to reduce risk (Gretzel & Yoo, 2008) and to increase efficiency of travel planning (Hwang, Jani, & Jeong, 2013). Travelers tend to consider consumer reviews important for accommodations decision making, and in some cases, reviews by other travelers can change a person’s hotel preference (Gretzel & Yoo, 2008).

Due to the prevalence of online information search behavior amongst tourists, analysis of online consumer reviews is essential for hotels. Callarisa et.al (2012) suggest that virtual environment (such as travel review sites) can be used to help hotel management to improve existing products, to create new products and to generate new business and differentiate in the market. By analyzing consumer reviews, hotel management can be informed of areas valued by their customers, and subsequently they can fine tune their hotel’s offerings to deliver exactly what customer wants.

In the past few years, there have been a number of hospitality and travel studies using online consumer reviews as their source of data, and some of them applied text mining methodology. The studies are summarized in Table 1.

Online review can be used as alternative source of information to study hotel attributes and positioning. Past research has used this approach. For example, studies by Li, Law, Vu, Rong, and Zhao (2015); Li, Ye, and Law (2014); Xiang and Gretzel (2010); Xiang, Schwartz, Gerdes, and Uysal (2015); Zhou et al. (2014); Xie, Zhang, and Zhang (2014) extracted hotel attributes from reviews.

In hospitality industry, it is crucial to identify the correct set of attributes that consumers regard as relevant factor in their purchase decision making since product attributes play pivotal role in developing customers’ positive affective evaluation toward the product. For instance, Han (2013), reports that distinguishable attributes of a restaurant can help customers assign value to the product. Furthermore, the attributes can influence customers to develop favorable affective evaluation and post-purchase intentions. Past research in hotel industry finds that hotel attributes influence hotel room rate (Zhang, Ye, & Law, 2011), traveler’s hotel selection (Jones & Chen, 2011) and guest satisfaction (Ogle, 2009). Understanding hotel attributes would in turn help hotel managers to develop and position their hotels in the market in accordance to attributes that matter for consumers.

The author argues that online review can be good source of attributes data. Hotel reviews are normally based on actual visit experience. Zhou et al., (2014) suggest to use online review as data source since it allows consumers freedom to report anything they deem relevant to their experience. This argument is also in line with consumer decision making process. For instance, Jones and Chen (2011) propose that customers use different attributes at different stages of the choice process, and accordingly, focusing on attributes used to evaluate purchase decision (after purchase/use experience) would likely yield more accurate and realistic picture of the attributes that affect the actual decision making. Thus, following Jones and Chen’s (2011) reasoning, the attributes extracted from online review would accurately reflect attributes relevant in hotel customer’s purchase decision. The use of attributes mentioned in reviews can also fill some gaps in hotel attributes research. Dolnicar and Otter (2003) report that most hotel attributes studies were focused on the “before purchase” attributes, and that future studies need to focus on attributes “during” and “after stay”. As for attributes from reviews, it can be considered as attributes “after stay” since reviews are normally written after actual hotel stay.

**Hotel attributes research**

Research effort to study hotel attributes seemed to be very extensive during decades between 1980s and early 2000s. Dolnicar and Otter (2003) reviewed 21 studies published between 1984 and 2000 in hospitality, tourism research and business journals. They extracted 173 attributes categorized into nine categories (Image, Price/Value, Hotel, Services, Room, Marketing, Food and Beverages, Security, Location), and then identified 23 attributes (out of the initial 173 attributes) that were
used most frequently in past studies. Another study was conducted by Ogle (2009) who used email-based qualitative questionnaire sent to 29 General Managers of deluxe hotels in Asia Pacific area, asking a series of question, one of which was about General Managers’ perception of primary criteria used by business guests to evaluate the guests’ stay. Jones and Chen (2011) conducted an experiment involving 53 subjects which simulated a real search and selection of hotels in Las Vegas area for a given travel date for four nights, for a number of travelers, with a given amount of travel budget.

More recent study about hotel attributes focused on online review. O’Connor (2013) employed content analysis on five most recent comments in TA website about 100 hotels in London. He identified important issues/attributes that serve as most common causes of satisfaction and dissatisfaction among reviewers based on most frequent words in the review. The attributes identified in the four aforementioned studies are displayed in Table 2.

Understanding which hotel attributes that matter for customers is important for hotel management so they can make optimal hotel development decision (Dolnicar & Otter, 2003).

Hotel attributes is important part of hotel selection criteria. Hotel attributes are important factors that might influence hotel room rate (Zhang, Ye, & Law, 2011), hotel selection (Jones & Chen, 2011) and guest satisfaction (Ogle 2009). Studies have found relationship between customer satisfaction and service quality, which include staff services quality and facilities (Choi & Chu, 2000). Past research about satisfaction and service quality suggest that hotel consumers form satisfaction evaluation based on their evaluation of attributes associated with the hotel (e.g. whether

| Table 2. Hotel Attributes Identified from Past Research |
|-------------------------------------------------------|
| Dolnicar and Otter (2003)                             |
| Restaurant facilities                                 |
| F&B quality                                           |
| Hotel aesthetics                                      |
| Meeting rooms available                               |
| Cleanliness of hotel                                  |
| Swimming pool                                         |
| Parking facilities                                    |
| Health /fitness facilities                            |
| Reputation                                            |
| Convenient location                                   |
| Price of accommodation                                |
| Room size                                             |
| Non-smoking rooms                                     |
| Cleanliness of bathroom                               |
| Quiet / soundproof rooms                              |
| Bathroom amenities                                    |
| Comfort of bed                                        |
| Cleanliness of room                                   |
| Service professionalism / quality                     |
| Friendliness of staff                                 |
| Business center                                       |
| Service speed (24 hour) room service                  |
|                                                      |
| Ogle (2009)                                           |
| Location                                              |
| Service/Service quality                               |
| Problem resolution                                    |
| Recognition                                           |
| Condition of room                                     |
| Responsiveness                                        |
| Cleanliness                                           |
| Value of money                                        |
| Safety and security                                   |
| Business Centre                                      |
| Meeting Facilities                                    |
| Price                                                |
| Comfort                                              |
| Communication                                        |
| Facilities                                           |
| Reputation                                           |
| Awards                                               |
| Special Deals                                        |
|                                                      |
| Jones and Chen (2011)                                 |
| Non smoking (rooms)                                   |
| Swimming pool                                         |
| High speed internet                                   |
| Hot tub                                              |
| Fitness Centre                                       |
| Room Service                                         |
| (Note: above list is the most popular attributes of 24 available attributes—complete attribute list were not provided in Jones and Chen’s article) |
|                                                      |
| O’Connor (2013)                                       |
| Hotel location                                       |
| Room size                                            |
| Staff                                                |
| Cleanliness                                          |
| Breakfast                                            |
| In room facilities                                    |
| Comfortable                                          |
| Temperature                                          |
| Dirty                                                |
| Maintenance                                          |
| Noise                                                |
| Expense                                              |
| Room Amenities                                       |
| Public Areas                                         |
| Booking                                              |
| Reservation issue                                    |
| Check-in                                             |
| Elevator Problems                                    |
| Restaurant Quality                                   |
| Quiet                                                |
| Internet access                                      |
| Upgrade                                              |
| (Note: above list are themes identified from most frequent words in online reviews) |
consumer’s expectation of a certain attribute is fulfilled or not); and these attributes may be key factor that affect consumer’s decision for future repurchase (Choi & Chu, 2000).

Dolnicar and Otter (2003) classify hotel attributes research into three streams: (1) research focused on relevant hotel attributes prior to booking, (2) research focused on relevant attributes during the stay at hotel, and (3) research focused on relevant attributes after the hotel experience. In their review, Dolnicar and Otter (2003) report that a majority of past studies were focused on the first two streams, while the third stream is still lacking in number. Amongst the studies focused on after visit experience (carried out by means of survey), it is found that staff/service qualities (e.g. related to staff friendliness, helpfulness, responsiveness), room quality (cleanliness, temperature control, quietness), value (e.g. room and food/beverages value for money, convenient location, brand), business facilities/services (e.g. meeting rooms, secretarial services), general amenities (e.g. room service, laundry, etc), and security were amongst the top attributes (Choi & Chu, 2000; Chu & Choi, 2000).

Methods

Scope of study

This study focuses on top ten five-star hotels in Jakarta, complemented with top five five-star hotels in Singapore (based on rankings in TripAdvisor website). The author is interested to limit the scope of study to a sample of five-star hotels located in Jakarta and complemented with a sample of hotels from Singapore to check external validity. The author intends to focus on five star hotels in Jakarta (as main sample) and Singapore (as additional sample to check external validity) because study that focuses on this type of hotel in the aforementioned cities in South East Asia is relatively scarce. The author argues that study on hotel attributes focused on five-star segment of hotel industry would be important as five-star hotels mostly cater the high-end corporate market and the wealthy tourism market (since these markets are most likely the only markets that can afford premium rate offered in five-star hotels). This market, although relatively small, represents lucrative market. In this kind of market, the way hotel attributes are designed and executed may be much more critical to satisfy the consumers.

Jakarta and Singapore were selected as the focus of study since both are business cities as well as the center of each respective country’s government. In addition, both cities also offer some tourist attractions or serve as gateway to other tourism destinations, thus having potential to attract business travelers as well as leisure travelers. According to Statistics Indonesia, in 2014 Jakarta was visited by 2.3 million foreigners that came mostly from Malaysia, China, Japan, South Korea and USA (Pariwisata DKI Jakarta, 2015a). Meanwhile, during the same period, Singapore was visited by 15 million foreign visitors that mostly came from Asia, Americas and Europe (Singapore Tourism Board, 2015a, 2015b).

Research on five-star hotels in Jakarta is also warranted because five-star hotels tend to be more preferred by foreign visitors to Indonesia. According to Statistics Indonesia (Pariwisata DKI Jakarta, 2015b), five-star hotels in Jakarta experience occupancy rate that varies between 50 and 60 percent during period of June and July each year, with ratio of foreign to domestic guests around 0.82 (approximately 8 foreign guests per 10 domestic guests), which is significantly higher incidence that that occurred in other classes of hotel (which varies between 0.02 in one-star hotel to 0.35 in four-star hotel). This indicates that five-star hotel is more preferred by foreign guests. Thus understanding which attributes are considered salient is imperative for five star hotels in Jakarta to attract its clientele, which consist of foreign visitors as well as wealthy Indonesia travelers.

Data collection and analysis methods

This study applies text mining method to analyze user-generated review from online review website. Netzer, Feldman, Goldenberg, and Fresko (2012) suggest that “text mining refers to the process of extracting useful, meaningful, and nontrivial information from unstructured text”. Text mining, which originated in computer science literature, has been used in marketing applications, for instance to study consumer preferences from product reviews, estimate demand for hotels and understanding market structure based on product attributes mentioned in reviews (Netzer et al., 2012).
Text mining process can be summed into two phases, namely (1) Preprocessing and integration of unstructured data, and (2) Statistical analysis of the preprocessed data to extract content from the text (Francis & Flynn, 2010).

This study consists of several phases, namely data collection, preprocessing, content extraction and data analysis (attribute extraction and position map construction). The first phase involved getting the raw text data from the review website. In this study, the author focused on top ten five-star hotels in Jakarta according to TripAdvisor (TA) website as of June 2013. Random samples were taken from each of the ten five-star hotels with help of Research Randomizer (Urbaniak & Plous, 2013), an online random number generator software. The number of sample size taken from each hotel varies according to its population (i.e. number of total reviews in a particular hotel). Sample size was calculated using online sample size calculator published by Australia National Statistical Service (“Sample size calculator,” n.d.). The sample size was taken by considering confidence level and standard error desired. In this study, the author used 95% confidence level and desired standard error of 0.05. Table 3 describes the five-star hotels (using their initials) and number of reviews sampled from the review website. In total, 662 reviews were included in Jakarta sample and 500 reviews in Singapore sample.

In order to establish external validity, (i.e., to demonstrate that text mining works in other location/data), second sample is needed. Thus, the author gathered sample of 500 reviews from top five hotels in Singapore, which would be analyzed in a similar manner as the Jakarta sample. However, due to time constraint, the author had to resort to convenience sampling of the reviews whereby 100 most recent reviews (at the time of data collection) from each of the top five hotels (according to TA during the time of data collection) were taken. The author decided to use arbitrary figure of 100 reviews per hotel after comparing with the sample size figures generated by online sample size calculator published by Australia National Statistical Service. Using the same criteria as used in Jakarta sample (confidence level of 95% and standard error of 0.05), on average the sample size calculator prescribed sample size under 100 reviews per hotel. Thus, the author rounded the sample size upwards to 100 reviews per hotel. Selection of Singapore for comparison is due to a relative similarity between Jakarta and Singapore in terms of both being business cities located in South East Asia. The relatively close distance between Singapore and Jakarta also affects travel behavior of visitors, particularly those on business trips. Some travelers would tend to visit Jakarta and Singapore on the same business trip and the author argue that this might bring more similarity of perceived attributes between five star hotels located in both cities. The result from both samples would be compared to see whether common attributes can be found from both samples.

Reviews selected were limited only to reviews in English. Each review was put into separate text files and given unique identification code that allows identifying the hotel to which
the review belongs. Since the author focused on finding attributes (which is usually in form of unigrams belonging to noun word class), the reviews would be analyzed at word level. Thus syntactic problems that may occur due to imperfect grammar or mixing English with non-English terms (e.g. in case of Singlish) would not be a problem. Any words not recognized as English word (or misspelled) were put aside from the analysis. In principle, text mining methodology could be carried out for texts written in non-English language. For instance, Chiu, Chiu, Sung, and Hsieh (2013) conducted opinion mining of hotel consumer-generated contents in Chinese Weblogs. Nevertheless, not all text mining software have the capability to directly handle non English texts.

One of the challenges in text mining is to handle word stemming (eliminating morphological endings and inflections from words (Feinerer, 2008). Some software such as GATE (General Architecture for Text Engineering), and KH Coder are reported to be able to handle texts written in non English languages such as Spanish, Chinese, Arabic, and Japanese (“Top 20 Free Software for Text Analysis, Text Mining, Text Analytics,” n.d.). In R software (R Core Team, 2013) tm package (Feinerer & Hornik, 2013), by default handles English using Porter’s stemmer algorithm. But according to Feinerer (2008), other languages such as German or Russian can also be used with some modifications in the software and package’s syntax and settings, which may be cumbersome for R users with less experience and technical inclination.

After collecting the reviews, the next step was preprocessing the texts. The author followed guidelines from Zhao (2012) which included the following steps: building corpus from the collected text, removing undesirable characters (e.g. whitespace, punctuations, stop-words, numbers, etc) , changing all words into lowercase format and building term-document matrix. Preprocessing and subsequent process of data analysis was carried out using “tm: Text Mining Package” (Feinerer & Hornik, 2013) in R statistical software version 3.01 (R Core Team, 2013)

The content extraction phase was conducted on the aggregate data. In this study, the author analyzed data using count-based evaluation, which identifies terms with highest frequency of occurrence. According to Feinerer (2008), words with highest occurrence frequencies in a text are deemed important. This kind of analysis, although simple, is commonly used in text mining and has advantage of being easy to interpret and computationally inexpensive (Feinerer, 2008).

The reviews from both sample sets were processed separately using tm package (Feinerer & Hornik, 2013) in R statistical software (R Core Team, 2013). After preprocessing (changing all words to lower case, removing white spaces, removing punctuation, numbers and stop-words), separate Term-Document Matrix (TDM) for each sample was constructed. The TDMs for both samples were calculated using TF-IDF (term frequency-inverse document frequency). TF-IDF is a means of normalizing term frequency data by taking into account the number of documents in which the term appears. TF-IDF is a popular weighting scheme used in text mining since it reduces the impact of irrelevant terms and highlights discriminative terms (Feinerer, 2008). Since each review has different length, a lengthy review would probably contain more of a certain term than a shorter review, consequently if a certain term got high raw frequency, there was a possibility that the high frequency came from a few lengthy reviews that repeatedly mentioned the term, not from many reviews. Therefore the author suggest that calculating frequency of occurrence using TF-IDF may be more appropriate as it would adjust the term frequency by taking into account the document length.

Data processing resulted in a list of terms. Word Cloud in Figure 1 and Figure 2 display the terms found in review data for Jakarta and Singapore samples respectively. Please note that the words contained in Figure 1 and Figure 2 are stemmed words. In text mining, stemming is the process of removing word suffixes to retrieve their radicals; for words in English, stemming removes common morphological and inflectional endings (Feinerer, 2008). Stemming is used so that words having different morphological and inflectional endings that belong to the same root are not counted as different words.

The terms contained in Word Cloud still needs to be further filtered to highlight hotel attributes. Judging from their part of speech,
attributes highlighted in past literature seems to be comprised of nouns. Thus, the next step involve identifying the part of speech that each word on the list of terms from text mining belong and then focus only on nouns, since they are more likely represent hotel attributes.

In order to do this process, the author first conducted stem completion using tm package (Feinerer & Hornik, 2013) in R version 3.01 (R Core Team, 2013), and then input the words into POS (part of speech) tagger provided online by University of Copenhagen’s Center for Sprogteknoologi to identify the part of speech, and then using the results to build list of most frequently occurring nouns in each data set. Once the part of speech had been identified, adjectives and other words such as proper nouns are excluded from the analysis, since those terms are less likely represent hotel attribute. The author first classified the terms using POS (part of speech) tagger (“CST’s online tools,” n.d.), which is available online. After the classification, the author deleted terms other than nouns, including proper nouns (nouns that refer to names of specific people, places, organizations, and sometimes things). To make sure accuracy of categorization results, the author then rechecked each words with English dictionary from Merriam-Webster online dictionary (http://www.merriam-webster.com/). Table 4 displays the 25 most frequently mentioned nouns from the review data.

**Results and Discussion**

The author applied word cloud using packages in R software (R Core Team, 2013) to visualize terms occurring in the reviews (see Figure 1 and Figure 2). Word cloud offers fast and practical method to data mine and synthesize large amount of data to yield key and significant themes (Huynh & Wilson, 2014). Viégas and Wattenberg (2008) suggest that word cloud evolved from tag-cloud, a technique that saw wider use in 2001 by Fortune magazine when they mapped 500 largest corporations in the world in circular masses of text whereby each cloud representing companies in a particular country. Viegas and Wattenberg (2008) adds that subsequently, tag-clouds were used in various social media as a way to display popular tags or topics used by social media users; subsequently, this technique (which is now called “word cloud”) was then found its way to academic articles and books. In business research, word cloud was also used, for instance in studies related to business intelligence (e.g. Ishikiriyama, Miro, & Gomes, 2015; Moro, Cortez, & Rita, 2015).

From the word cloud in Figure 1 and Figure 2, it could be seen that the reviews contain different kinds of words, such as nouns (e.g. hotel, room, service) and adjectives (e.g. good, great) with words denoting location (e.g. Jakarta and Singapore). In word clouds, terms with higher term frequency, appear larger in the
word cloud. Note however that terms appearing in the word cloud are terms that are still in its stemmed form. To have a clearer sense of the terms appearing in the reviews, the stemmed words were returned back to its original form using tm package’s stem completion procedure. The results are displayed in Table 4.

As could be seen from Table 4, location, staff, service, restaurant and food are five most frequently mentioned attributes in five-star hotels in Jakarta data set, while service, staff, location, view and room are the five most frequently mentioned attributes in Singapore data set. Apparently, terms mentioned by reviewers of Singapore and Jakarta Five-Star hotels are mostly similar, although with different rank. However, the two sets of sample seem to agree on top two terms, which were service and staff, respectively.

In terms of external validity, it seems that attributes identified in this study concurred with attributes that were reported on past studies that also used user-generated review (see Table 5). Studies by Li et al. (2015) and Xiang et al. (2015) reported similar set of attributes such as service, staff, room, location, etc. According to Xiang et al. (2015), these terms can be classified into several groups, namely: (1) core product (example in this study: “room”, “bed”, and “bathroom”); (2) hotel amenities (example in this study: “restaurant”, “pool”, etc); (3) hotel attributes (example in this study: “location”, “convenience”, etc); (4) hotel staff-related descriptors (example in this study: “staff”); (5) hotel service encounters (example in this study: “check-in”, “shopping”, etc); and (6) evaluation of experience (note: example from this study not included in top 25 most frequent words). It apparent from Table 4 that the top 10 terms in this were centered around the first three categories (core product, hotel amenities and attributes) proposed by Xiang et al. (2015). Study by Xiang et al. (2015) focused on hotels located in US, while Li et al. (2015) focused

| Terms     | Jakarta Sample | Singapore Sample | Terms     | Jakarta Sample | Singapore Sample |
|-----------|----------------|------------------|-----------|----------------|------------------|
|           | TF-IDF         | TF-IDF           |           |                |                  |
| 1         | Service        | 8.641            | Service   | 6.623          |                  |
| 2         | Staff          | 7.779            | Staff     | 6.454          |                  |
| 3         | Club           | 7.730            | Location  | 6.234          |                  |
| 4         | Location       | 7.711            | View      | 6.216          |                  |
| 5         | Pool           | 7.686            | Room      | 6.135          |                  |
| 6         | Food           | 7.054            | Hotel     | 6.012          |                  |
| 7         | Restaurant     | 7.046            | Food      | 5.667          |                  |
| 8         | Breakfast      | 6.870            | Club      | 5.525          |                  |
| 9         | Hotel          | 6.452            | Breakfast | 5.441          |                  |
| 10        | Check-in       | 6.027            | Shop      | 5.182          |                  |
| 11        | Mall           | 5.928            | Place     | 5.076          |                  |
| 12        | Floor          | 5.774            | Pool      | 5.042          |                  |
| 13        | Lounge         | 5.721            | Check-in  | 4.685          |                  |
| 14        | Facility       | 5.720            | Bar       | 4.529          |                  |
| 15        | Shop           | 5.641            | Restaurant| 4.426          |                  |
| 16        | Room           | 5.340            | Facilities| 4.416          |                  |
| 17        | Gym            | 5.299            | Lobby     | 3.410          |                  |
| 18        | Bathroom       | 5.165            | Bathroom  | 2.996          |                  |
| 19        | Buffet         | 5.045            | Convenience| 2.871          |                  |
| 20        | Traffic        | 4.882            | Expensive | 2.679          |                  |
| 21        | View           | 4.584            | Mall      | 2.517          |                  |
| 22        | Bed            | 4.526            | Minute    | 2.516          |                  |
| 23        | Work           | 4.274            | Spa       | 2.498          |                  |
| 24        | Spa            | 4.268            | Butler    | 2.413          |                  |
| 25        | Bar            | 4.117            | Person    | 2.405          |                  |
Table 5. Comparison of Attributes Found in Hotel Attributes Study Using Consumer Review as 
Data Source

| O’Connor (2013) | Xiang et al. (2015) | Li et al. (2015) | Current Study (Jakarta & Singapore) |
|-----------------|---------------------|------------------|-----------------------------------|
| Hotel location  | Room                | Room             | Service                           |
| Room size       | Clean               | Staff            | Staff                             |
| Staff           | Staff               | Location         | Club                              |
| Cleanliness     | Location            | Breakfast        | Location                          |
| Breakfast       | Comfortable         | Service          | Pool                              |
| In room facilities | Service             | Clean            | Food                              |
| Comfortable     | Friendly            | Food             | Restaurant                        |
| Temperature     | Close               | Pool             | Breakfast                         |
| Dirty           | Breakfast           | Floor            | Hotel                             |
| Maintenance     | Helpful             | View             | View                              |
| Noise           | Bed                 | Bed              | Room                              |
| Expense         | Price               | Bathroom         | Check-in                          |
| Room Amenities  | Restaurants         | Station          | Mall                              |
| Public Areas    | Walking             | Restaurant       | Floor                             |
| Booking         | Area                | Bar              | Lounge                            |
| Reservation issue | Parking            | Buffet           | Shop                              |
| Check-in        | Bathroom            | Price            | Gym                               |
| Elevator Problems | Pool               | Airport          | Bathroom                          |
| Restaurant Quality | Free               | Lobby            | Place                             |
| Quiet           | Convenient          | Road             | Buffet                            |
| Internet access | Downtown            | Reception        | Traffic                           |
| Upgrade         | Airport             | Desk             | Bar                               |
| (Note: above list are themes identified from most frequent words in online reviews) | Desk | Street | Bed |
| View            | Taxi                | Facilities       |                                   |
| Recommend       | Lounge              | Work             |                                   |
| Noise           | Door                | Spa              |                                   |
| Quiet           | Water               | Lobby            |                                   |
| Food            | Train               | Convenience      |                                   |
| Distance        | Access              | Expensive        |                                   |
| Shuttle         | Club                | Minute           |                                   |
| Street          | Coffee              | Butler           |                                   |
| Shopping        | Internet            | Person           |                                   |

Note: Difference in rankings of terms may result from different weighting scheme used in analysis. Some studies may have used raw term counts, while this study used TF-IDF weighting.

on hotels in Hong Kong, Shanghai, Bangkok and Sydney (note: both were not focused on specific hotel class). This fact indicates that hotel consumers in different cities in the world seem to have same salient set of attributes that mark their experience.

Result from studies conducted using user generated reviews in Table 5 could also be compared to studies conducted using survey and other conventional data collection method shown in Table 2. Despite of some differences in exact terms used in each study, some similarities of terms can be seen. For example, Dolnicar and Otter (2003) reported convenient location, service quality, friendliness of staffs, and price as some of the important hotel attributes. Similarly, O’Connor (2013) found that room, breakfast, staff, location, bed and shower as some of the most frequently mentioned attributes. Meanwhile, location and service were also
identified in Ogle’s (2009) study. Apparently, location, service, room, staffs, and various hotel facilities such as gym and swimming pools are consistently reported from study to study. It seems that these factors may be critical factors that hotel managers need to pay close attention to, since these factors consistently appear in studies that conducted before purchase, during purchase and after purchase phase as highlighted by Dolnicar and Otter (2003) and Jones and Chen (2011).

Positioning of Five Stars Hotels using Correspondence Analysis

Once the attribute set had been derived from most frequently mentioned terms, the next step was performing correspondence analysis to analyze association between hotels and attributes. Correspondence analysis was conducted using FactoMineR package (Husson, Josse, Le, & Mazet, 2014) in R statistical software version 3.01 (R Core Team, 2013). Input data for Correspondence Analysis were derived from raw term frequency of each hotel in both Jakarta and Singapore samples (analyses were run separately). The author used terms that were most frequently mentioned and developed contingency table as input data for correspondence analysis in FactoMineR (Husson et al., 2014).

For Jakarta sample, the raw term frequency was rescaled to account for disproportionate number of reviews for each hotel. This step follows suggestion from Kudlats, Money, and Hair (2014), who propose that rescaling is required since disproportionate number of reviews for each hotel would cause disproportionate mentions of attributes, which in turn might cause correspondence analysis result to be skewed toward hotels with more reviews. However, the same treatment was not carried out for the Singapore sample since there is already same number of reviews for each hotel. After rescaling the frequency of terms, correspondence analysis (CA) was carried out.

Once CA solution was obtained, the author checked the presence of outlier. According to Bendixen (1996), outliers could be found by examining rows and columns that have both high absolute coordinate values (more than +/- 1) and high contributions. There were outliers in both sample (“lobbies” in Jakarta sample and “butler” in Singapore sample). Presence of outlier could dominate the interpretation toward one or more of the axes and cause the remaining row and column points tightly clustered together and difficult to interpret. Thus, outlier terms were removed, and another CA was carried out without the offending terms.
Table 6. Summary of Correspondence Analysis Result Jakarta Sample (overall) Source

| Dimension | Singular value | Inertia | Chi-Square | Sig | Percent of inertia |
|-----------|---------------|---------|------------|-----|-------------------|
|           |               |         |            |     | Accounted for      | Cumulative |
| 1         | 0.23          | 0.051   | 38.08      |     | 38.08             |
| 2         | 0.17          | 0.029   | 22.03      |     | 60.11             |
| 3         | 0.15          | 0.021   | 15.80      |     | 75.91             |
| 4         | 0.11          | 0.013   | 9.52       |     | 85.44             |
| 5         | 0.08          | 0.006   | 4.61       |     | 90.05             |
| 6         | 0.07          | 0.006   | 4.21       |     | 94.26             |
| 7         | 0.06          | 0.003   | 2.60       |     | 96.86             |
| 8         | 0.05          | 0.002   | 1.72       |     | 98.58             |
| 9         | 0.04          | 0.002   | 1.42       |     | 100               |
| 10        | 0.23          | 0       | 0          |     | 100               |
| Total     |               | 0.133   | 996.35     | 0.00| 100%              | 100%       |

Note: Square-root of total inertia: 0.365 with Chi-Sq 996.35 with df = (26-1)*(10-1) = 225 yielding p-value <0.001

Table 7. Summary of Correspondence Analysis Result (Jakarta sample, row points)

| Attributes | Mass | Coordinates | Inertia | Contribution | Cos² | Total |
|------------|------|-------------|---------|--------------|------|-------|
|            | Dim 1 | Dim 2       | Dim 1   | Dim 2        | Dim 1 | Dim 2 |
| Amenities  | 0.017 | -0.161      | 0.134   | 0.002        | 0.848 | 1.018 |
| Bar        | 0.010 | 0.211       | 0.353   | 0.005        | 0.85  | 4.137 |
| Bathroom   | 0.034 | -0.045      | 0.16    | 0.004        | 0.133 | 2.977 |
| Bed        | 0.023 | 0.149       | -0.065  | 0.003        | 1.007 | 0.334 |
| Breakfast  | 0.050 | 0.177       | -0.009  | 0.003        | 3.079 | 0.013 |
| Buffet     | 0.022 | 0.295       | -0.127  | 0.004        | 3.799 | 1.206 |
| Check-In   | 0.009 | -0.189      | -0.123  | 0.001        | 0.654 | 0.48  |
| Club       | 0.030 | -0.064      | -0.432  | 0.008        | 0.241 | 18.937|
| Facilities | 0.015 | -0.016      | -0.017  | 0.002        | 0.007 | 0.015 |
| Floor      | 0.023 | 0.02        | -0.072  | 0.002        | 0.017 | 0.396 |
| Food       | 0.051 | -0.055      | -0.019  | 0.000        | 0.301 | 0.064 |
| Gym        | 0.027 | -0.345      | 0.102   | 0.005        | 6.425 | 0.968 |
| Internet   | 0.017 | 0.142       | -0.215  | 0.004        | 0.677 | 2.707 |
| Location   | 0.042 | -0.028      | -0.034  | 0.002        | 0.067 | 0.164 |
| Lounge     | 0.030 | 0.419       | -0.486  | 0.013        | 10.231| 23.778|
| Mall       | 0.032 | -0.769      | -0.272  | 0.023        | 37.738| 8.139 |
| Pool       | 0.042 | -0.186      | 0.093   | 0.006        | 2.842 | 1.234 |
| Price      | 0.024 | 0.31        | 0.216   | 0.009        | 4.513 | 3.787 |
| Restaurant | 0.051 | 0.16        | 0.172   | 0.005        | 2.53  | 5.081 |
| Room       | 0.179 | 0.013       | 0.061   | 0.003        | 0.059 | 2.238 |
| Security   | 0.029 | 0.44        | -0.27   | 0.009        | 10.954| 7.132 |
| Service    | 0.098 | 0.065       | 0.015   | 0.001        | 0.82  | 0.071 |
| Shopping   | 0.034 | -0.349      | -0.148  | 0.007        | 8.094 | 2.507 |
| Shower     | 0.010 | 0.22        | 0.345   | 0.006        | 0.913 | 3.869 |
| Staff      | 0.084 | -0.045      | 0.173   | 0.004        | 0.337 | 8.605 |
| View       | 0.018 | -0.281      | 0.048   | 0.003        | 2.864 | 0.142 |

Active Total | 1.000 | 0.134 | 100 | 100

*Attributes that were well represented on two dimensional map (i.e Total Cos² > 0.5)
Table 8. Summary of Correspondence Analysis Results (Jakarta sample, column points)

| Hotels | Mass | Coordinates Dim 1 | Coordinates Dim 2 | Inertia Dim 1 | Chi-Square | Cos^2 Dim 1 | Cos^2 Dim 2 | Total |
|--------|------|-------------------|-------------------|---------------|------------|-------------|-------------|-------|
| MO     | 0.042| 0.078             | 0.173             | 0.008         | 0.511      | 0.033       | 0.164       | 0.197 |
| D      | 0.044| 0.141             | 0.226             | 0.008         | 1.73       | 0.015       | 0.269       | 0.374 |
| FS     | 0.048| 0.034             | 0.206             | 0.007         | 0.109      | 0.009       | 0.315       | 0.324 |
| ML     | 0.041| 0.237             | 0.111             | 0.006         | 4.539      | 0.037       | 0.082       | 0.454 |
| RCPP   | 0.056| -0.159            | -0.342            | 0.016         | 2.779      | 0.087       | 0.405       | 0.492 |
| SL     | 0.040| 0.185             | -0.03             | 0.006         | 2.71       | 0.021       | 0.005       | 0.219 |
| RCMK   | 0.059| 0.179             | -0.011            | 0.009         | 3.728      | 0.022       | 0.001       | 0.223 |
| KRT    | 0.100| -0.09             | 0.352             | 0.020         | 1.593      | 0.04        | 0.611       | 0.651* |
| GH     | 0.350| -0.249            | -0.057            | 0.025         | 42.765     | 0.868       | 0.046       | 0.914* |
| JWM    | 0.218| 0.304             | -0.12             | 0.028         | 39.535     | 0.716       | 0.112       | 0.828* |

**Active** 1.000 0.134 100 100

*Attributes that were well represented on two dimensional map (i.e Total Cos^2 > 0.5)

Table 9. Summary of Correspondence Analysis Result Singapore Sample (overall)

| Dimension | Singular value | Inertia | Chi-Square | Sig | Percent of inertia |
|-----------|----------------|---------|------------|-----|--------------------|
|           |                |         |            |     | Accounted for Cumulative |
| 1         | 0.254          | 0.064   |            |     | 57.07              | 57.07 |
| 2         | 0.149          | 0.022   |            |     | 19.66              | 76.72 |
| 3         | 0.129          | 0.017   |            |     | 14.78              | 91.50 |
| 4         | 0.098          | 0.010   |            |     | 8.50               | 100.00 |
| 5         | 0.000          | 0.000   |            |     | 0.00               | 100.00 |
| Total     |                | 0.113   | 491.21     | 0.00| 100%               | 100%  |

*Note: Square-root of total inertia: 0.336 with Chi-Sq 491.21 with df = (23-1)*(5-1) = 88, yielding p- value <0.001

CA factor map of Jakarta sample is displayed in Figure 3 with corresponding output tables presented in Table 6 through Table 8, while the CA factor map of Singapore sample is displayed in Figure 4 with corresponding output tables presented in Table 9 through Table 11.

From Table 6 and Table 9, total square root of inertia for Jakarta and Singapore samples were 0.365 and 0.336 respectively. Thus, it could be concluded that in Jakarta and Singapore samples, CA results indicate significant association between rows and columns, since square root of inertia exceed 0.2 (Bendixen, 1996). After confirming significant row and column association, the second step is to determine the appropriate number of dimensions. Bendixen (1996) suggest the appropriate number of dimensions could be determined by examining the percent of inertia accounted for by each dimension, and then compare those ratios with expected ratio that would occur if the data were purely random without significant row-column dependencies (for details see Bendixen, 1996). Following Bendixen’s suggestion, in Jakarta sample, there are three dimensions that can be analyzed, but the author decided to take the first two dimensions since they account for most variance (both account for 60.11 percent variance). In Singapore sample, dimension 1 and 2 also account for most of the variance (76.72 percent), thus in Singapore sample dimension 1 and 2 will also be selected for analysis.
Table 10. Summary of Correspondence Analysis Result (Singapore sample, row points)

| Attributes | Mass     | Coordinates Dim 1 | Coordinates Dim 2 | Inertia Dim 1 | Inertia Dim 2 | Contribution Dim 1 | Contribution Dim 2 | Cos² Dim 1 | Cos² Dim 2 | Total     |
|------------|----------|-------------------|-------------------|---------------|---------------|--------------------|--------------------|------------|------------|-----------|
| Service    | 0.0838   | 0.238             | -0.025            | 0.0057        | 7.402         | 0.23               | 0.834              | 0.009      | 0.843*     |
| Staff      | 0.0813   | -0.024            | -0.052            | 0.0012        | 0.074         | 1.005              | 0.04               | 0.188      | 0.228      |
| Location   | 0.0365   | -0.054            | 0.205             | 0.0048        | 0.167         | 6.931              | 0.023              | 0.321      | 0.344      |
| View       | 0.0487   | -0.661            | -0.16             | 0.0238        | 33.078        | 5.601              | 0.893              | 0.052      | 0.945*     |
| Room       | 0.1736   | -0.037            | -0.037            | 0.0005        | 0.372         | 1.061              | 0.492              | 0.484      | 0.976*     |
| Hotel      | 0.2036   | 0.119             | 0.004             | 0.0034        | 4.498         | 0.017              | 0.846              | 0.001      | 0.847*     |
| Food       | 0.0381   | 0.119             | 0.036             | 0.0008        | 0.842         | 0.218              | 0.667              | 0.059      | 0.726*     |
| Club       | 0.0317   | -0.742            | -0.159            | 0.0206        | 27.107        | 3.626              | 0.846              | 0.039      | 0.885*     |
| Breakfast  | 0.0494   | -0.032            | 0.038             | 0.0018        | 0.078         | 0.328              | 0.029              | 0.041      | 0.07       |
| Shopping   | 0.0207   | 0.261             | 0.101             | 0.0044        | 2.188         | 0.958              | 0.317              | 0.048      | 0.365      |
| Place      | 0.0243   | 0.079             | -0.049            | 0.0002        | 0.238         | 0.265              | 0.677              | 0.26       | 0.937*     |
| Pool       | 0.0354   | 0.077             | 0.294             | 0.0055        | 0.328         | 13.746             | 0.038              | 0.553      | 0.591*     |
| Check-In   | 0.0347   | -0.084            | -0.036            | 0.0008        | 0.383         | 0.2                | 0.322              | 0.058      | 0.38       |
| Bar        | 0.0225   | 0.64              | -0.59             | 0.0171        | 14.322        | 35.336             | 0.538              | 0.457      | 0.995*     |
| Restaurant | 0.0266   | 0.046             | 0.208             | 0.0015        | 0.088         | 5.182              | 0.037              | 0.76       | 0.797*     |
| Facilities | 0.0133   | -0.114            | 0.237             | 0.0022        | 0.271         | 3.376              | 0.079              | 0.339      | 0.418      |
| Lobby      | 0.0122   | -0.006            | 0.201             | 0.0031        | 0.001         | 2.222              | 0                 | 0.159      | 0.159      |
| Bathroom   | 0.0186   | -0.077            | 0.008             | 0.0009        | 0.171         | 0.006              | 0.018              | 0.001      | 0.119      |
| Convenience| 0.0076   | -0.197            | 0.415             | 0.0028        | 0.457         | 5.877              | 0.103              | 0.458      | 0.561*     |
| Expensive  | 0.0092   | 0.324             | -0.264            | 0.0023        | 1.498         | 2.895              | 0.416              | 0.277      | 0.693*     |
| Mall       | 0.0073   | -0.603            | -0.107            | 0.0044        | 4.155         | 0.381              | 0.602              | 0.019      | 0.621*     |
| Minute     | 0.0135   | 0.281             | 0.149             | 0.0017        | 1.666         | 1.353              | 0.641              | 0.179      | 0.82*      |
| Spa        | 0.0076   | -0.229            | 0.519             | 0.0031        | 0.615         | 9.188              | 0.129              | 0.665      | 0.794*     |

Active Total  1.000  0.138  100  100

*Attributes that were well represented on two dimensional map (i.e Total Cos² > 0.5)

Table 11. Summary of Correspondence Analysis Results (Singapore sample, column points)

| Hotels | Mass     | Coordinates Dim 1 | Coordinates Dim 2 | Inertia Dim 1 | Inertia Dim 2 | Contribution Dim 1 | Contribution Dim 2 | Cos² Dim 1 | Cos² Dim 2 | Total     |
|--------|----------|-------------------|-------------------|---------------|---------------|--------------------|--------------------|------------|------------|-----------|
| FSS    | FSS      | 0.168             | 0.166             | 0.018         | 6.913         | 19.576             | 0.252              | 0.245      | 0.497      |
| RCM    | RCM      | -0.269            | -0.061            | 0.023         | 26.366        | 3.945              | 0.737              | 0.038      | 0.775*     |
| SHLA   | SHLA     | 0.068             | 0.214             | 0.016         | 1.378         | 39.268             | 0.056              | 0.551      | 0.607*     |
| PPS    | PPS      | -0.218            | -0.072            | 0.018         | 17.434        | 5.498              | 0.619              | 0.067      | 0.686*     |
| RFLS   | RFLS     | 0.412             | -0.197            | 0.038         | 47.909        | 31.713             | 0.808              | 0.184      | 0.992*     |

Active Total Active Total  0.138  100  100

*Attributes that were well represented on two dimensional map (i.e Total Cos² > 0.5)

As can be seen in Figure 3 and Figure 4, the hotels were mapped near attributes associated with them. The closer a hotel to attributes, the more they are associated. However, the resulting CA output still needed to be refined to yield more meaningful interpretation. To get meaningful interpretation, there are several steps that needed to be carried out (Kuldats et al., 2014). First, checking how well the row and column points are well represented...
in the two dimensional map using total $\cos^2$ value for row and column points. Second, interpreting the dimensions by examining the row points that contribute significantly to that dimension. This was carried out by inspecting the contribution of each (row or column) element and check if the contribution exceeded the level of contribution that can be expected if there were random distribution of row (or column) elements over the axes (Bendixen, 1996). Third, interpreting the meaning of association between hotels and dimensions. The association was examined based on quality of representation indicators ($\cos^2$ values) for column points (Table 8 and Table 11 for Jakarta and Singapore samples, respectively). If a particular hotel receive higher $\cos^2$ value on a certain Dimension, then the hotel tend to be associated with the Dimension. For instance, in Jakarta sample, MO have $\cos^2$ value of 0.03 for Dimension 1 and 0.16 for Dimension 2, then MO tend to be more associated with Dimension 2. Lastly, interpreting the association of row points and column points graphically. Greenacre (2007) recommended that correspondence map should be interpreted starting from the horizontal axis then move toward vertical axis. Positions of hotels in the map indicate their relationship to each other as well as the association of hotels relative to the significant terms/attributes (cf. Kudlats et al., 2014).

Table 12 and Table 13 displays summary of significant terms on Dimension 1 and Dimension 2 for Jakarta and Singapore sample respectively, while Figure 5 and 6 display the final correspondence map that depict only significant attributes and hotels. In subsequent paragraphs, these graphs and figures will be analyzed in more detail.

**Table 12. Significant Terms on Dimension 1 and Dimension 2 (Jakarta sample)**

| Dimension 1 | Terms on the left side | Terms on the right side |
|-------------|------------------------|-------------------------|
| Mall (37.74%) | Lounge (10.23%)         |
| Shopping (8.1%) | Security (10.95%)      |
| Gym (6.43%) | Price (4.51%)           |

| Dimension 2 | Terms on top half perceptual map | Terms on lower half of perceptual map |
|-------------|----------------------------------|-------------------------------------|
| Staff (8.6%) | Mall (8.14%)                     |
| Restaurant (5.08%) | Lounge (23.78%)      |
| Price (3.79%) | Security (7.13%)               |
| Bar (4.14%) | Club (18.94%)                    |
| Shower (3.87%) |                                |

| Dimension 1 | Terms on the left side | Terms on the right side |
|-------------|------------------------|-------------------------|
| View (33.08%) | Bar (14.32%)           |
| Club (2.11%) | Service (7.40%)        |
|              | Hotel (4.50%)          |

| Dimension 2 | Terms on top half perceptual map | Terms on lower half of perceptual map |
|-------------|----------------------------------|-------------------------------------|
| Pool (13.75%) | Bar (35.34%)                     |
| Spa (9.19%) | View (5.60%)                     |
| Location (6.93%) |                                |
| Convenience (5.88%) |                                |
Interpretation of the Correspondence Map

The majority of hotels in Jakarta and Singapore samples belongs to international chain of hotels. In Jakarta sample, only D that does not seem to be affiliated to any hotel chain (domestic or international). Some hotels in Jakarta and Singapore are part of same international hotel chains. For instance, FS (Jakarta) and FSS (Singapore) are part of same international chain. Similarly, RCPP, RCMK (Jakarta) and RCM in are also affiliated. The same is also true for SL (Jakarta) and SHLA (Singapore).

From Figure 5, GH plots closer to left side of Dimension 1 (Mall, Shopping, and Gym), than to Lounge, Security and Price. The same observation can be applied of RCPP. Both are also closer to lower side of Dimension 2 (Mall, Lounge, Security, and Club) than to upper side of Dimension 2 (Staff, Restaurant, Price, Bar, and Shower). It appears that the two hotels were perceived to be closer to shopping centers. The association is confirmed as the two hotels are actually located near major shopping centers in Central Jakarta area. For five star hotels in Jakarta it seems that customers regard placement in strategic location, and situated close to malls/shopping areas as important features, this is evident from remarks in reviews about GH:

“The main reason for me choosing this hotel is the location, considering the bad traffic in Jakarta, this hotel is above PI mall, 2 minutes walking to the big GI shopping mall, where lots of good shops and places to eat.” (GH Guest)

Importance of location in hotel industry was highlighted by Zhang et al., (2011) who found that location convenience and room quality were predictors of hotel room price in midscale and luxury hotels. They found that the more desirable the location, the higher the room rate that a particular hotel can charge. The finding in this study confirms with previous findings related to the importance of location in hotel industry as one of the attributes consumers consider.

KRT was the only hotel on the top left quadrant. It is closer to the left side of Dimension 1 (Mall, Shopping, and Gym), and upper side of Dimension 2 (Staff, Restaurant, Price, Bar, and Shower). KRT is also located near shopping mall, thus confirming the depicted association in perceptual map. It seems that closeness to shopping malls is a differentiating factor that separated KRT with FS, MO, D and ML. The latter group of hotels are closer to the right side of Dimension 1 (lounge, security and price) while similarly associated with upper side of Dimension 2 (Staff, Restaurant, Price, Bar and Shower).

Besides being somewhat associated to “lounge, security and price” attributes, FS, MO, D and ML are also closely associated with amenities such as restaurant, bar, shower, and to the term “price”. Apparently, the guests have made notable mentions of restaurants, bars and lounge in these hotels.

“...The restaurants are among the best in Jakarta...” (ML Guest)

“...The French restaurant and Chinese restaurant are wonderful...” (MO Guest)

It is interesting to note that even though the four hotels are associated with “price”, when we examine the reviews individually, some refer the hotel to offer value for money in spite of being expensive, but there are some who referred to paying price above perceived value. For instance different guests from the same hotel expressed:

“...Its pricey for Indonesia standards (still so much cheaper than FS in Sydney or Singapore) but service is always excellent......” (FS Guest)

“...Probably won’t stay there again as for the same price you can get the service and the comfortable rooms at other place...” (FS Guest)

It is not possible to gauge whether there are more people that perceive the hotel were overpriced or justly priced with the current method. Complementary method such as opinion mining and sentiment analysis can be applied to capture sentiment from the reviews. Opinion mining and sentiment analysis focus on detecting polarity and recognize emotion (Cambria, Schuller, Xia, & Havasi, 2013). An example of study employing this method is study by Mostafa (2013), which analyzed sentiment polarity of more than 3500 social media tweets expressing attitudes towards sixteen global brands. However, Mostafa (2013) also noted that sentiment polarity analysis might need to be complemented with sentiment topic recognition (STR) in order to reveal the underlying reason.
behind forming such opinions. In the current research, due to time and scope limitation, the author did not apply any of the aforementioned methods and only limited the use of text mining methodology to identify attributes and generate positioning maps.

The fourth cluster consists of RCMK, SL, and JWM. The three hotels are similar to GH and RCPP in terms of association with shopping malls (the hotels are relatively close to shopping malls, although require more travel efforts than GH and RCMK). Nevertheless, the differentiating attribute of the fourth cluster is in terms of mention about security. The three hotels in this cluster are more associated with security. Some of the comments regarding security are as follows:

"Security is very thorough - both on the driveway, at the front door, and apparently at certain times of day..." (RCMK Customer)

"We arrived at the entrance to the SL and were initially stunned at the amount of security that is felt necessary to protect staff and guests..." (SL Customer)

In the Singapore sample, RCM and PPS are associated with left side of Dimension 1 (View, Club, Convenience, and Location) and bottom side of Dimension 2 (Bar, Club, Service, and View). RFLS, is positioned on the bottom right quadrant, being closely associated with bar and service. FSS is associated closely with pool, hotel and service. SHLA is associated with location, convenience and spa. It is apparent from Figure 6 that the top five hotels in Singapore are relatively close to each other despite of being positioned in different quadrants. These hotels are differentiated along the combination of key attributes such as being in convenient location, offering good view, having distinctive clubs, pool, or spa, or offering great service. However, out of the five hotels depicted in the positioning map, RFLS conspicuously stands in one key distinguishing attribute, which is being the only hotel positioned the closest to “bar”. Upon further examination of the hotel, it turns out that the hotel has several bars; one of them is a famous tourist destination in Singapore that many reviewers recommend to visit.

When we compare some of the hotels that belong to similar chain in Indonesia and Singapore, there is similarity between RCPP and RCM (both belong to same international hotel chain), in terms of its association with dimensions concerning bar, club and location.

Figure 5. Final Correspondence Map, Jakarta Sample (only significant attributes and hotels are shown)
If the reader examine Figure 4 (where all attributes are plotted with the hotels regardless the significance of attributes), RCM and RCPP are positioned relatively close to shopping mall. From Figure 4, SL and SHLA are also appear to share some similarity, for instance, in both locations they are closely associated with restaurant, location and pool.

One notable difference in positioning map of hotels in Jakarta and Singapore is in the absence of “security” term in Singapore positioning map. In Singapore sample, the reviewers might not write security as a notable attribute in the hotels that were sampled, perhaps because Singapore is perceived as one of the safest cities in the world. On the other hand, in Jakarta, the tight security measures in some of the hotels are noticed by the reviewers. This is perhaps due to past terrorist attacks and the generally less positive perception of the city’s safety level.

Managerial Implications

This study confirmed other studies that have attempted to use text mining methodology to gather information regarding business entities. In this study, the author managed to identify attributes most frequently mentioned in hotel reviews and describe how the attributes are associated with the hotels under review. Thus, the author would like to make a few suggestions to hotel managers.

First, hotel managers need to apply text mining methodology in their business intelligence activities to monitor consumer perception and subsequently use the information to adjust the marketing mix to reach and maintain desired position in the hotel market. The pervasiveness of travel review forums and review sites present opportunities for hotel managers to gauge consumer perception and find out what relevant attributes they consider important.

Second, paying attention to elements such as location, staff and service that were mentioned the most in the reviews. For five-star hotel segment that primarily caters high end corporate/business travelers and upscale tourist market, consumers would likely demand convenience and top-notch service that reflect the premium price they pay. Failure to provide that could mean losing customers as some of the guests in the review has mentioned their
intention not to return. This study found that similar set of attributes was evident in Jakarta and Singapore. These attributes were also similar to past studies on hotels in US cities, Hong Kong, Bangkok, and Sydney. This fact indicates that five-star hotel market seems to be relatively similar. This is probably due to similar profile of travelers that frequent five-star hotels (i.e. in Jakarta, based on data from Statistics Indonesia, foreign visitors seem to prefer five-star hotels, and looking at the demographic makeup of foreign visitors to Jakarta and Singapore from their respective authorities, it seems that both cities were visited by people from Asia-Pacific, Europe, and Americas). These people were likely to be frequent travelers who travel around the world, thus these kinds of travelers would probably have a similar frame of reference as to what five-star hotel everywhere ought to offer. Thus hotel managers (particularly those from international hotel chains) need to maintain strong focus on these attributes, since these attributes could serve as hygiene factor that might make or break their business.

**Conclusion**

This study applied text mining to uncover attributes used by hotel customers in Jakarta and Singapore to evaluate their post-purchase experience with hotels. In addition, this study also applied correspondence analysis to create positioning map of the hotels from Jakarta sample based on attributes derived from hotel reviews using text mining method. The attributes identified from terms mentioned in reviews turned out to resemble attributes mentioned in past hotel attribute research. Dolnicar and Otter (2003) noted that most hotel attributes used in past hotel attribute studies were used in studies conducted at pre-purchase stage. This study, on the other hand was based on post-purchase data (online hotel reviews by customers who had stayed in the hotels). The attributes identified from this study turned out to be overlapping with attributes reported in Dolnicar and Otter’s literature review.

This study also demonstrated that attributes distilled from online review via text mining could be used to create a relatively accurate two-dimensional positioning map. The association between attributes and hotels as depicted on the map could be confirmed with the actual condition (e.g. by cross-checking with actual words mentioned in the reviews as well as with facts pertaining to the hotels). For instance, some hotels that were associated with shopping malls in the two-dimensional map were actually located near shopping malls. This is verifiable by looking at Jakarta City map as well as cross-checking with the individual reviews content.

However, this study exhibited some limitations. First, two-dimensional map produced on Jakarta sample could only account for 60.1 percent of data variation, while the Singapore sample, though slightly better, could only account for 76.72%. In addition, only a few terms (attributes) and hotels demonstrated quality of representation above minimum criteria of 0.5 as suggested by Kudlats et al. (2014). With this limitation, caution needs to be paid when interpreting the two-dimensional map solution.

Another limitation of this study is that this study did not analyze the sentiments of the statements. Thus, one cannot make conclusion about the performance of each hotel in terms of each attribute. The correspondence map only displayed association between hotels and attributes based on the number of mentions about the attributes, not based on positivity or negativity of the attributes. Further research also could be directed into analyzing sentiments from the reviews to see how each hotel fare in terms of the attributes used in the review. In addition, future research can also attempt to use review data to predict consumer hotel preference or identify other hotels that consumer normally consider when they think of a certain brand (i.e. brands of hotels in consideration set) as demonstrated by Netzer et al. (2013) in sedan car and diabetes drug market.
References

Bendixen, M. (1996). *A Practical Guide to the Use of Correspondence Analysis in Marketing Research* (pp. 16–36). Marketing Research On-line, vol.1.

Callarisa, L., Garcia, J. S., Cardiff, J., & Roshchina, A. (2012). Harnessing social media platforms to measure customer-based hotel brand equity. *Tourism Management Perspectives, 4*, 73–79. doi:10.1016/j.tmp.2012.04.005

Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New Avenues in Opinion Mining and Sentiment Analysis. *IEEE Intelligent Systems, 2*(April), 15–21.

Chiu, C., Chiu, N.-H., Sung, R.-J., & Hsieh, P.-Y. (2013). Opinion mining of hotel customer-generated contents in Chinese weblogs. *Current Issues in Tourism, 0*(0), 1–19. doi: 10.1080/13683500.2013.841656

Choi, T. Y., & Chu, R. (2000). Levels of satisfaction among Asian and Western travellers. *International Journal of Quality and Reliability Management, 17*(2), 116–131.

Chu, K. K. S., & Choi, T. (2000). An importance-performance analysis of hotel selection factors in the Hong Kong hotel industry: a comparison of business and leisure travellers. *Tourism Management, 21*(4), 363–377. doi:10.1016/S0261-5177(99)00070-9

CST’s online tools. (n.d.). Center for Sprogteknologi, University of Copenhagen. Retrieved from http://ada.sc.ku.dk/tools/index.php?lang=en

Decker, R., & Trusov, M. (2010). Estimating aggregate consumer preferences from online product reviews. *International Journal of Research in Marketing*, 27, 293–307.

Dolnicar, S., & Otter, T. (2003). Which Hotel attributes Matter? A review of previous and a framework for future research. *Proceedings of the 9th Annual Conference of the Asia Pacific Tourism Association (APTA), 1*, 176–188.

Feinerer, I. (2008). *A text mining framework in R and its applications* (Unpublished Doctoral Thesis). WU Vienna University of Economics and Business. Retrieved from http://epub.wu.ac.at/1923/

Feinerer, I., & Hornik, K. (2013). tm: Text Mining Package (R package version 0.5-9.1) [Computer Program]. Retrieved from http://cran.r-project.org/package=tm

Francis, L., & Flynn, M. (2010). Text mining handbook. *Casualty Actuarial Society E-Forum*, (Spring 2010), 1–61.

Godes, D., & Mayzlin, D. (2004). Using Online Conversations to Study Word-of-Mouth Communication. *Marketing Science, 24*(4), 545–560.

Greenacre, M. (2007). *Correspondence analysis in practice* (2nd ed.). Boca Raton, FL: Chapman and Hall/CRC.

Gretzel, U., & Yoo, K. (2008). Use and impact of online travel reviews. *Information and Communication Technologies in Tourism*, 35–46. Retrieved from http://www.springerlink.com/index/x740x784r6w527t0.pdf

Gruen, T. W., Osmonbekov, T., & Czaplewski, A. J. (2006). eWOM: The impact of customer to customer online know-how. *Journal of Business Research*, 59(4), 449–456.

Han, H. (2013). The healthcare hotel: Distinctive attributes for international medical travelers, 56(2013).

Husson, F., Josse, J., Le, S., & Mazet, J. (2014). FactoMineR: Multivariate Exploratory Data Analysis and Data Mining with R. (R package version 1.26) [Computer Program]. Retrieved from http://cran.r-project.org/package=FactoMineR

Huynh, G., & Wilson, J. A. J. (2014). Vietnamese Attitudes and Behavioural Patterns towards Counterfeit Brands. *ASEAN Marketing Journal*, 6(2), 89–104.

Hwang, Y. H., Jani, D., & Jeong, H. K. (2013). Analyzing international tourists’ functional information needs: A comparative analysis of inquiries in an on-line travel forum. *Journal of Business Research*, 66(6), 700–705. doi:10.1016/j.jbusres.2011.09.006

Ishikiriyama, C. S., Miro, D., & Gomes, C. F. S. (2015). Text Mining Business Intelligence: A small sample of what words can say. *Procedia Computer Science, 55*(Itqm), 261–267. doi:10.1016/j.procs.2015.07.044
Jones, P., & Chen, M.-M. (2011). Factors Determining Hotel Selection: Online Behaviour by Leisure Travellers. *Tourism and Hospitality Research, 11*(1), 83–95. doi:10.1057/thr.2010.20

Kudlats, J., Money, A., & Hair, J. F. (2014). Correspondence analysis: A promising technique to interpret qualitative data in family business research. *Journal of Family Business Strategy, 5*, 30–40. doi:http://dx.doi.org/10.1016/j.jfbs.2014.01.005

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

Li, H., Ye, Q., & Law, R. (2014). Determinants of Customer Satisfaction in the Hotel Industry: An Application of Online Review Analysis. *Asia Pacific Journal of Tourism Research, 18*(7), 784–802.

Moro, S., Cortez, P., & Rita, P. (2015). Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation. *Expert Systems with Applications, 42*(3), 1314–1324. doi:10.1016/j.eswa.2014.09.024

Mostafa, M. M. (2013). More than words: Social networks’ text mining for consumer brand sentiments. *Expert Systems with Applications, 40*, 4241–4251.

Netzer, O., Feldman, R., Goldberg, J., & Fresko, M. (2012). Mine Your Own Business: Market-Structure Surveillance Through Text Mining, 31(3), 521–543.

O’Connor, P. (2010). Managing a Hotel’s Image on TripAdvisor. *Journal of Hospitality Marketing & Management, 19*(7), 754–772. doi:10.1080/19368623.2010.508007

Ogle, A. (2009). Making sense of the hotel guestroom. *Journal of Retail and Leisure Property, 8*(3), 159–172. doi:10.1057/rpl.2009.7

Pariwisata DKI Jakarta. (2015a). *Berita Resmi Statistik No.07/02/31/ Th. XVII, 02 Februari 2015, Badan Pusat Statistik Propinsi DKI Jakarta*, 1–7.

Pariwisata DKI Jakarta. (2015b). *Berita Resmi Statistik No.35/08/31/Th. XVII, 03 Agustus 2015, Badan Pusat Statistik Propinsi DKI Jakarta*, 1–7.

R Core Team. (2013). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/.

Sample size calculator. (n.d.). Australian Bureau of Statistics. Retrieved from http://www.nss.gov.au/nss/home.nsf/pages/Sample+size+calculator

Singapore Tourism Board. (2015a). International Visitor Arrivals Statistics. Retrieved September 25, 2015, from https://www.stb.gov.sg/statistics-and-market-insights/marketstatistics/ivas-tat_dec_2014%28updated 07apr15%29.pdf

Singapore Tourism Board. (2015b). *Tourism Sector Performance, Q4 2014 Report* (pp. 1–7). [Adobe Digital Edition Version]. Retrieved from https://www.stb.gov.sg/statistics-and-market-insights/MarketStatistics/STB_Tourism Sector Performance Q4 2014_Final.pdf

Tang, C., & Guo, L. (2013). Digging for gold with a simple tool: Validating text mining in studying electronic word-of-mouth (eWOM) communication. *Marketing Letters*. doi:10.1007/s11002-013-9268-8

Top 20 Free Software for Text Analysis, Text Mining, Text Analytics. (n.d.). Retrieved September 21, 2015, from http://www.predictiveanalyticstoday.com/top-free-software-for-text-analysis-text-mining-text-analytics/

Urbaniak, G. C., & Plous, S. (2013). Research randomizer (version 4.0). Retrieved from http://www.randomizer.org/

Viégas, F. B., & Wattenberg, M. (2008). Tag Clouds and the Case for Vernacular Visualization. *Interactions*, (July-August), 49–52.

Xiang, Z., & Gretzel, U. (2010). Role of social media in online travel information search. *Tourism Management, 31*(2), 179–188. doi:10.1016/j.tourman.2009.02.016

Xiang, Z., Schwartz, Z., Gerdes, J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management, 44*(2015), 120–130. doi:10.1016/j.ijhm.2014.10.013
Xie, K. L., Zhang, Z., & Zhang, Z. (2014). The business value of online consumer reviews and management response to hotel performance. *International Journal of Hospitality Management*, 43, 1–12.

Ye, Q., Law, R., Gu, B., & Chen, W. (2011). The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings. *Computers in Human Behavior*, 27(2), 634–639. doi:10.1016/j.chb.2010.04.014

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

Zhang, Z., Ye, Q., & Law, R. (2011). Determinants of hotel room price: An exploration of travelers’ hierarchy of accommodation needs. *International Journal of Contemporary Hospitality Management*, 23(7), 972–981. doi:10.1108/09596111111167551

Zhao, Y. (2012). *R and Data Mining: Examples and case studies*. Waltham, MA: Academic Press.

Zhou, L., Ye, S., Pearce, P. L., & Wu, M. (2014). Refreshing hotel satisfaction studies by reconfiguring customer review data. *International Journal of Hospitality Management*, 38(2014), 1–10.