

## CONTENT ANALYSIS OF SOCIAL MEDIA: A GROUNDED THEORY APPROACH

Linda S.L. Lai  
School of Business  
Macao Polytechnic Institute  
Macao SAR, China  
[sllai@ipm.edu.mo](mailto:sllai@ipm.edu.mo)

W.M. To  
School of Business  
Macao Polytechnic Institute  
Macao SAR, China  
[wmt@ipm.edu.mo](mailto:wmt@ipm.edu.mo)

### ABSTRACT

Social media has become a vital part of social life. It affects the beliefs, values, and attitudes of people, as well as their intentions and behaviors. Meanwhile, social media enables governments and organizations to engage people while allowing consumers to make informed decisions. Therefore, converting social media content into information, key concepts, and themes is crucial for generating knowledge and formulating strategies. In this paper, we introduce a grounded theory approach that involves (i) defining the goal and scope of a study; (ii) logically and systematically identifying social media sources, total sample size, and the sample size of every source category; (iii) employing computer-aided lexical analysis with statistical and graphical methods to identify the key dimensions of the topic while minimizing human errors, as well as coding and categorization biases; and (iv) interpreting the findings of the study. This systematic approach was illustrated with the destination image of Macao as an example. The proposed methodology with its hybrid nature can quantitatively analyze qualitative social media content (e.g., impressions, opinions, and feelings) and can identify emergent concepts from the ground up. This paper contributes to electronic commerce research by presenting a novel approach for extracting, analyzing, and understanding social media content.

Keywords: Social media; Content analysis; Lexical and statistical approaches; Concept formation

### 1. Introduction

The Internet has attracted the attention of research communities [Comley, 2008; Dwivedi et al., 2008; Zwass, 1996]. In particular, the significant role of analyzing social media and networks to advance our understanding of information sharing, communication [Averya et al., 2010; Chiu et al., 2006; Turri et al., 2013], opinion formation, and dissemination has been recognized [Abrahams et al., 2012; Airoidi et al., 2006; Bai, 2011; Jansen et al., 2011; Lane et al., 2012]. Nevertheless, rigorous, quantitative studies on social media content, particularly on electronic commerce and information management, remain scarce [Bai, 2011]. The most considerable barrier to social media usage is the lack of a versatile methodology for selecting, collecting, processing, and analyzing contextual information obtained from social media sites. However, several software companies have developed proprietary text mining systems for data visualization [Arnold, 2012], and researchers have developed expert systems for sentiment analysis [Abrahams et al., 2012; Lane et al., 2012].

Nevertheless, social media content is widely accessible, up-to-date, and available in electronic format. Therefore, a systematic approach is necessary, as it helps electronic commerce researchers, organizations, and governments understand the commonality in various online text data that appear in social media. Using the information obtained from social media, researchers can gain valuable insights into the beliefs, values, attitudes, and perceptions of social media users with regard to the utility of user-generated content and trust formation [Karimov et al., 2011; Kim et al., 2012; Wang & Li, 2014]. Consequently, such information can help marketers monitor the perceptions of people regarding social networks and aid organizations in strategic planning.

To address the gap between the availability of user-generated raw text and the contextual information of aggregated data, the present study introduces a grounded theory approach [Strauss & Corbin, 1988] to analyze social media content to identify the underlying factor structure of the collected information and to interpret the identified

structure in relation to the study objective. Strauss and Corbin [1998] defined the grounded theory approach as a research method that employs a systematic set of procedures to develop an “inductively derived” grounded theory about a particular phenomenon. This grounded theory can likewise be used to explore concepts and develop themes based on qualitative data [Arnold et al., 2006; Glaser & Strauss, 2009; Witz, 2007].

Social media is one of the most important components of electronic commerce and information management [Lai & Turban, 2008]. Thus, computer-based methods that employ lexical analysis and statistical approaches should be developed for handling large qualitative datasets (bottom-up), or at least, a smaller representative subset of the data. Specifically, our approach contributes to the existing literature by creating a structured goal-driven process and adopting both computer-based lexical analysis and statistical approaches to achieve interpretable text mining (i.e., deriving concepts and themes that bear significant meanings to users). Although several analysis methods have focused on the underlying structure, algorithms, and architectures of employed technology, our approach regards the technical process as a means to an end. As such, we aim to generate attainable recommendations of business value from social media content analysis.

The remainder of the paper is structured as follows. First, a review of the literature on social media, content analysis approach, and computer-aided lexical and statistical analysis methods is presented. Then, we propose a grounded theory methodology to address the gap in the content analysis of social media and subsequently use the destination image of Macao to illustrate the application of the methodology. Finally, we present the implications and limitations of the study.

## 2. Literature Review

### 2.1. Social Media

Social media comprise Internet-based applications that are developed based on the ideological and technological foundations of Web 2.0. Social media enables the creation and exchange of user-generated content [Hinchcliffe, 2008; Karimov et al., 2011; Turban et al., 2015]. Using Internet- and web-based technologies, social media transform broadcast media monologues (i.e., one-to-many) into social media dialogues (i.e., many-to-many). Through social media, users can upload photos, videos, music, images, and texts to share ideas, feelings, opinions, and experiences with other members [Lai & Turban, 2008; Turban et al., 2015]. In developing countries such as China and India, social media has undergone phenomenal growth [Lai & To, 2012; To et al., 2014; Srivastava & Pandey, 2013].

In particular, social networking sites, online forums, instant messaging services, and mobile smart platforms have grown exponentially, resulting in the widespread use of social media. In this regard, social media has become a powerful force of democratization. Social media enabled communication and collaboration among individuals at a massive scale without geographical, time, and system constraints [Hinchcliffe, 2008; Lai & Turban, 2008]. The personal elements of social media communities induce high levels of trust. Such trust results in the perception of the received information’s reliability [Karimov et al., 2001]. Trust and information exchange are essential components of decision-making.

### 2.2. Social Media Content Analysis

The rapidly increasing amount of social media information and consumer views on a product or service, which can be either positive or negative, has a considerable effect on an organization. Thus, researchers have developed sophisticated tools for topic modeling and document clustering [Banerjee & Basu, 2007; Becker et al., 2009; Blei et al., 2003; Ramage et al., 2011; Ma et al., 2013], as well as text mining tools [Aggarwal & Wang, 2011; Hu & Liu, 2012; Morinaga et al., 2002; Ye et al., 2009]. The unsupervised learning of latent topics is useful for different online applications, such as organizing documents according to topic-based clustering, and information filtering based on user preferences [Banerjee & Basu, 2007]. The topic model utilizes the Bayesian model for text document collection [Blei et al., 2003]. It automatically learns a set of thematic topics from collected documents and then assigns a number of these topics to each collected document. The topic model can be considered as a probabilistic version of latent semantic analysis [Deerwester et al., 1990; Newman & Block, 2006].

Scholars focused on the topic modeling community [Morinaga et al., 2002; Ramage et al., 2011; Wang et al., 2009; Ye et al., 2009] have suggested methods to incorporate meta-data and hierarchy into their models, which can be considered as partially supervised text mining. This methodology is similar to the approach that we describe in the present study. Particularly, several researchers have developed sentiment classification techniques that automatically mine and classify the text of written comments or opinions in social media as positive or negative [Morinaga et al., 2002; Saggion et al., 2007; Saggion & Funk, 2009; Ye et al., 2009]. Mathematically, sentiment classification labels a passage  $p$  according to its general sentiment  $s$ , where  $s \in \{-1, 1\}$ ;  $-1$  indicates unfavorable and  $1$  represents favorable description. Hence, a collection of passages can be classified into two. Similarly, Saggion and Funk (2009) proposed a five-point fine grain classification scale (from very bad to excellent) with a supervised machine learning framework. Such classification methods have been applied to fields related to computing, such as natural language processing and

information retrieval [Pang et al., 2002; Godbole et al., 2007; Turney & Littman, 2003], as well as to the real-time monitoring of candidates' performance in the debates prior to the 2008 US elections [O'Connor et al., 2010]. Lane et al. [2012] integrated sentiment analysis and machine learning to analyze positive or negative opinions expressed in social media. However, Abrahams et al. [2012] argued that sentiment analysis is insufficient to identify, categorize, and prioritize vehicle defects discussed in online forums, a particular form of social media. Thus, they developed a system to identify and prioritize vehicular defects based on text mining. However, Abrahams et al. [2012] and Lane et al. [2012] relied on experts to develop a defect and component classification scheme for defect identification or sentiment analysis.

Overall, existing approaches to sentiment analysis, whether rule-based or learning-based, have focused on determining sentiment scores for a collection of documents. By contrast, our approach focuses on identifying themes and links among those themes through exploratory factor analysis and thematic mapping. In particular, exploratory factor analysis is employed to identify the factors or themes, as well as the variables or words that belong under specific categories [Hair et al., 2006]. Thematic mapping creates links based on co-occurrences among keywords in sampled texts. Our approach can be used to extract both facts and opinions from social media content.

### 2.3. Content Analysis

Content analysis is "any technique for making inferences by objectively and systematically identifying specified characteristics of messages" [Holsti, 1969, p. 14]. Content analysis is the process of "summarizing, quantitative analysis of messages that relies on the scientific method (including attention to objectivity, intersubjectivity, a priori design, reliability, validity, generalizability, replicability, and hypothesis testing) and is not limited as to the types of variables that may be measured or the context in which the messages are created or presented" [Neuendorf, 2002, p. 10].

Used as a research method, content analysis has several advantages [Merriam, 2009]. First, it provides profound insight into a situation which is not limited by existing viewpoints or methodologies, thus allowing new theories on the topic to be discovered. Second, content analysis is highly effective when applicable models, which serve as a basis for quantitative research projects, are unavailable. Finally, the observational approach allows for the participants' opinion to be taken into consideration, which is impossible in the generalized view provided by quantitative research.

Traditionally, content analysis involves the following steps: (1) selecting a topic, (2) deciding on the sample, (3) defining concepts or units to be counted, (4) constructing categories, (5) creating coding forms, (6) training coders, (7) collecting data, (8) determining inter-coder reliability, (9) analyzing data, and (10) reporting results. Traditional content analysis involves human subjective interpretation. Thus, the classification procedure should be reliable to ensure consistency among different coders and the same coders over time. For this reason, validity, inter-coder reliability, and intra-coder reliability have been the subject of extensive research [Krippendorff, 2004].

### 2.4. Computer-Aided Lexical Analysis

In traditional content analysis, a research team should formulate a coding scheme and train coders prior to analyzing message characteristics. Linguistics researchers and scientists have developed algorithms and software to aid in content analysis [Evans, 1996; Krippendorff, 2004; Scott, 1996, 2008; Smith, 2000], which helped reduce subjective interpretation among coders. Computer-aided lexical software has several categories. Some simple lexical software produce word counts, whereas others produce both word counts and co-occurrences. Recognizing the limitations of content analysis, Scott [1996] independently developed WordSmith, a lexical software, and proposed keyness as an approach for identifying keywords with unusually high frequency counts, with the British National Corpus as lexical reference. Smith [2000] introduced an automated content analysis technique that groups keywords according to themes and presents the identified themes graphically.

### 2.5. Statistical Factor Analysis

Most lexical software packages generate a list of keywords along with their respective frequency counts. A keyword can be considered a key attribute, and frequency counts can be considered the strength of this attribute embedded in the text. From this perspective, content analysis through lexical software can be applied repeatedly in several similar documents to identify a frequency table. In this table, columns represent key attributes (i.e., keywords obtained from a computer-aided lexical analysis of a master file), whereas rows represent the strength of attributes among different writers or authors. Once the frequency table is completed, the table is subjected to factor analysis. Hair et al. [2006] explained exploratory and confirmatory factor analyses. Given that contextual information evolves from content analysis, exploratory factor analysis is the most appropriate method for examining underlying factors.

## 3. Methodology for Social Media Content Analysis

The establishment of a systematic methodology is important in gathering, analyzing, and grouping descriptive information available in social media into interpretable concepts for various decision support applications, such as crowdsourcing, profiling, web mining, social recommendation, and social reputation modeling. Analyzed results from

the proposed methodology should be reproducible and generalizable to allow for the collection, analysis, and longitudinal monitoring of web information as a function of time. This phenomenon can be attributed to the profound effects of shifts in the views of citizens, specifically “netizens,” regarding the components of decision making.

Figure 1 shows the four-phase “social media-to-concepts” methodology developed by the authors. Decision-makers employ computer-aided technologies to understand the contextual information available on the Internet. The proposed methodology is divided into four phases, namely, definition of goal and scope, data collection, data transformation, and results interpretation.

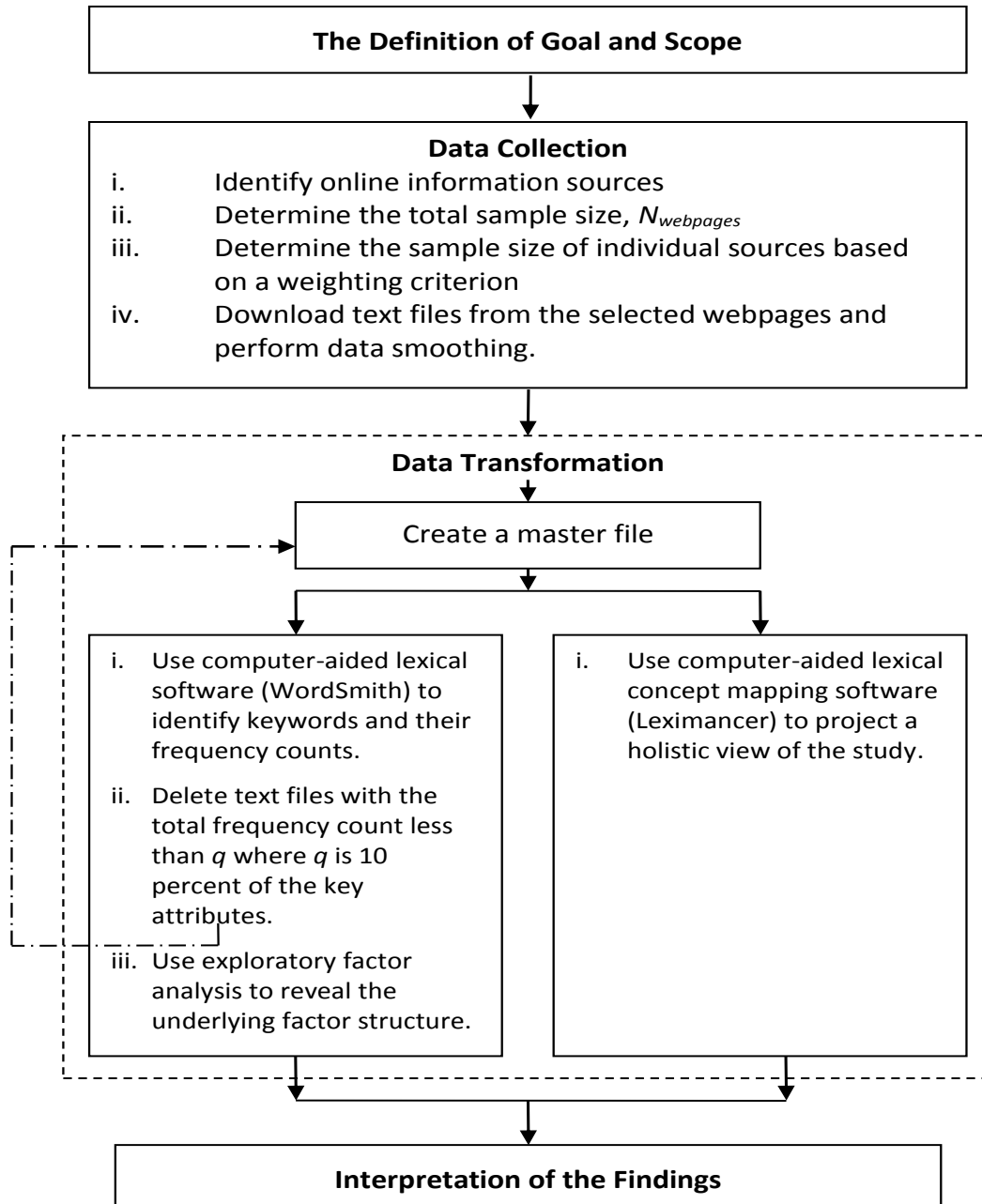


Figure 1: Four phases of the “social media-to-concepts” approach

### 3.1. Phase 1: Definition of Goal and Scope

The most critical phase in any project is the definition of the goal and scope of the study. Several specific criteria on the dataset can be identified and delineated. A content analysis can become an unbounded study if the objectives

of the investigation are unspecified. For example, when the investigator decides to study the destination image of a city or country based on information from social media, the goal of the project is “the destination image of the PLACE,” and the scope under investigation can be defined as the “content analysis of the PLACE using the latest online information provided by social media sites where tourists who have visited the place express their ‘true’ feelings.” Thus, a holistic and balanced view of the PLACE can be obtained because the information expressed by tourists or visitors with different interests is simultaneously and naturally collected.

### 3.2. Phase 2: Data Collection

Having defined a goal and a clear scope, the investigator can then identify the criteria, including the sources and the number of webpages to be downloaded. The sources of webpages should be identified based on the scope of the study. Using the aforementioned example, an investigator can determine whether the sources meet the criteria. These sources are webpages, including blogs and online forums in popular websites related to the nature of this study (e.g., Tripadvisor.com for the tourism industry). The webpage location can be identified based on the results of a keyword search in Google or Bing. For example, “PLACE” + “tourist” + “blogs/forums” in Google search can identify the webpages of travel-related blogs and forums. The most popular social media websites on travel can be located through Alexa.com.

After identifying the sources of information available in social media, the next step is to determine the number of webpages to be downloaded for content analysis based on the study goal. In the above example, “the destination image of the PLACE” provides valuable information on the extensiveness of the collection domain [Pike, 2002].

Destination image is formed by a multitude of dimensions and a number of attributes [Beerli & Martin, 2004; Echtner & Ritchie, 2003; Ryan & Cave, 2005]. In particular, Beerli and Martin [2004] suggested that the individual assessments of destination images can be broadly classified according to the following nine dimensions: natural resources, general infrastructure, tourist infrastructure, tourist leisure and recreation, culture and history, political and economic factors, natural environment, social environment, and ambience in the place. Beerli and Martin [2004] selected 33 attributes to measure the cognitive and affective components of the destination image for Lanzarote, Spain. The researchers found that the 33 attributes can be classified under the following six dimensions: natural and cultural resources, tourist and leisure infrastructures, social setting, natural environment, atmosphere, and affective image. Echtner and Ritchie [2003] found that most researchers in the past have conceptualized destination image in terms of physical attributes, rather than holistic impressions. The number of attributes used by researchers ranges from 10 [Goodrich, 1977] to 32 [Phelps, 1986]. By providing a list of 34 attributes found along a functional–psychological continuum, Echtner and Ritchie [2003] suggested that destination image should be operationalized in terms of functional and psychological characteristics. Using 9 dimensions (or factors) and 34 attributes, the minimum number of responses per attribute is 10 [Hair et al., 2006; Nunnally, 1978]. Thus, the minimum number of responses (webpages) in the present study is 340. A considerable number of responses increases the power associated with inferential statistics [Hair et al., 2006]. Mathematically, this criterion is expressed by Eq. (1).

$$N_{webpages} \geq 10 * N_{att}, \quad (1)$$

where  $N_{webpages}$  represents the total number of webpages and  $N_{att}$  represents the number of attributes.

The number of webpages for each category of social media sources can be determined to meet the objective of the study, given multiple social media sources, such as blogs and forums on service providers. For example, if the study aims to obtain the destination image of the PLACE in each category, the researcher can download an equal proportion of blogs and forums from service providers. Eq. 2(a) and 2(b) illustrate how the number of webpages from every category of social media sources is determined.

$$p_i = w_i * N_{webpages}, \quad 2(a)$$

$$\text{and } \sum w_i = 1, \quad 2(b)$$

where  $p_i$  represents the number of webpages per category of social media sources, whereas  $w_i$  is the weighting function for  $i$ th type of sources. For multiple sources with similar popularity,  $w_i$  can be set as a constant by simply dividing  $N_{webpages}$  by the total number of sources.

### 3.3. Phase 3: Data Transformation

In the past, content analysis was mostly conducted manually, with investigators interpreting text by classification, categorization, and subjective interpretation. With the advancement of lexical software, semantic software, and statistical tools, investigators can objectively interpret qualitative information from a wider perspective by identifying underlying key attributes, factors, and themes [Bondi & Scott, 2010; Scott, 2008; Scott & Smith, 2005]. In the present study, we introduce two complementary approaches to analyze user-generated content obtained from social media sites. The first approach identifies keywords based on their keyness through lexical software [Scott, 1996, 2008] and employs exploratory factor analysis in grouping keywords into several factors. Scott [2008] defined keyness as “a quality possessed by words, word-clusters, phrases, etc., a quality which is not language-dependent but text-dependent”

[Bondi & Scott, 2010, p.24]. Keyness is a log-likelihood (LL) measure of the relatedness of one or more specified words or keywords in the master file to a reference corpus [Hurt, 2010; Scott, 2008]. Rayson and Garside [2000] presented a contingency table of two corpora (Table 1), such as the master file and reference corpora for word frequencies.

Table 1: Contingency table for word frequencies

	Corpus one	Corpus two	Total
Frequency of the "word"	$a$	$b$	$a+b$
Frequency of other words	$c-a$	$d-b$	$c+d-a-b$
Total	$c$	$d$	$c+d$

In Table 1, the values  $a$  and  $c$  correspond to the total number of a particular word and the total number of words in corpus one, respectively, while  $b$  and  $d$  correspond to the total number of that particular word and the total number of words in corpus two, respectively. Moreover,  $a$  and  $b$  pertain to the observed values ( $O$ ), and expected values ( $E$ ) can be derived with Eq. (3).

$$E_i = \frac{N_i \sum_j O_j}{\sum_j N_j} \quad (3)$$

In the above case,  $N_1 = c$ , and  $N_2 = d$ . Thus, for a particular word,  $E_1 = c^*(a+b)/(c+d)$ , and  $E_2 = d^*(a+b)/(c+d)$ . This calculation considers the sizes of the two corpora.  $LL$  can be calculated with Eq. (4).

$$-2 \ln \lambda = 2 \sum_i O_i \ln \frac{O_i}{N_i} \quad (4)$$

The above equation yields the same result in calculating  $LL$  as follows:

$$LL = 2 \left[ a \log \frac{a}{E_1} + b \log \frac{b}{E_2} \right] \quad (5)$$

Exploratory factor analysis is a method for condensing information contained in a number of original variables into a smaller set of underlying dimensions (or factors), with minimal loss of information [Hair et al., 2006]. Hair et al. [2006] reported that this technique can be applied to a response table, in which columns represent variables or attributes and rows are the responses to these variables. Using the above example, all webpages (i.e., 340) are combined to form a master file. WordSmith is used to identify the keyness of words by comparing the master file with a lexical reference. WordSmith is equipped with the British National Corpus, a 100 million-word collection of samples of written and spoken language from several sources designed to represent a wide cross-section of contemporary British English. This lexical software identifies keywords in terms of their keyness values when these words have exceptionally higher frequency counts than the reference corpus. As in the above example, the first 68 words for 34 attributes with high keyness are retained, whereas some keyness words with an anti-image correlation of less than 0.5, factor loading of less than 0.4, or cross-loading of more than 0.4 [Hair et al., 2006] are deleted. We can thus apply a safety factor, namely, oversample ratio [Shannon, 1949] of 2:

$$N_{keyness} = 2 * N_{att}, \quad (6)$$

where  $N_{keyness}$  represents the number of retained keyness in words, which are regarded as the key dimensions of a purposeful study. Once the list of keyness words is obtained, each webpage is treated as a response. The frequency of each keyness word for every response is identified through WordSmith. This methodology is appropriate because the information provider, whether a tourist with family members or a casual traveler, can independently create a webpage. A frequency table of keyness in words is obtained by applying the procedure to all identified webpages. The frequency table is then subjected to factor analysis with SPSS, a statistical software program.

As suggested by Hair et al. [2006], the Kaiser-Meyer-Olkin (KMO) index of sampling adequacy and the test of Bartlett's test of sphericity are used to ensure the adequacy of the samples for factor analysis. Varimax and oblimin rotations are applied to the dataset to reveal underlying factors. The latent root criterion, that is, an eigenvalue greater than one, is used to retain the factors that explain most variances and maintain the simplicity of the factor structure. Exploratory factor analysis decomposes an adjusted correlation matrix whose variables are standardized, with mean = 0 and standard deviation = 1. The amount of variance explained is then equal to the matrix trace, the sum of the adjusted diagonals, or commonalities. Thus, an exploratory factor analysis is given by:

$$Y = X\beta + E, \quad (7)$$

where  $Y$  is the matrix of measured variables,  $X$  is the matrix of common factors,  $\beta$  is the matrix of weights (factor loadings), and  $E$  is the matrix of unique factors that represent error variations.

The second approach employs lexical mapping software such as Leximancer to identify themes based on the co-occurrences of keywords across a text database. This approach is similar to the first method, except that the software derives the concept map from cross-tabulation, rather than principal component analysis or principal axis factoring. This approach is versatile because the orthogonal relationship or other constraints among factors is unnecessary, and the co-occurrence produces links that create a concept map similar to human walkthrough experiences [Watson et al., 2005]. Therefore, this approach is appropriate in characterizing the experiential feelings of writers.

#### 3.4. Phase 4: Interpretation of Results

In the final phase of the methodology, the investigator ensures that the results obtained in Phase 3 are interpretable, specifically with regard to the goal and scope defined in Phase 1. The investigator summarizes the findings, identifies the managerial and practical implications of the findings to aid in decision-making, and acknowledges the limitations of the study.

### 4. Application of the Methodology to an Illustrative Case

#### 4.1 Case of Macao, China

Macao is one of the two special administrative regions in the People's Republic of China. The other administrative region is Hong Kong. As a former Portuguese colony, Macao was both the first and last European colony in China. Its land area is 29.9 square kilometers. Its population is approximately 580,000 in 2014. As a popular tourist destination in the region, Macao is the only Chinese city where casino gaming is legal.

When the government of Macao legalized gaming in 2002, casino concessions were granted to Sociedade de Jogos de Macau (SJM), Wynn, and Galaxy, which signed sub-concessions with MGM, Melco-PBL, and Las Vegas Sands, respectively. In eight years, Las Vegas Sands, Galaxy, Wynn, Melco-PBL, MGM, and SJM have developed mega-casinos, hotels, resorts, and shopping malls, as well as convention and exhibition centers. Macao has consequently transformed itself into a world-renowned gaming center. Moreover, the United Nations Educational, Scientific, and Cultural Organization (UNESCO) designated the historic center of Macao as a World Heritage Site in July 2005. In 2013, Macao attracted more than 29.3 million visitors, which contributed to the annual gross gaming revenue of US\$36.1 million [DSEC, 2013]. The tourist-to-population ratio was 50:1, which was considerably high compared with the ratio of 6.8:1 in Hong Kong [CenStatd, 2013] and 1.1:1 in Guangdong Province [GdGov, 2013]. Other Asian countries, such as Singapore, Vietnam, and Japan, followed Macao in building casinos with entertainment facilities to attract affluent tourists. In order to diversify its gaming-related tourist attractions, the government marketed Macao as an exciting place where East meets West, and as a destination for relaxation, entertainment, and gaming [Lai & To, 2010]. The tourism board of Macao uses its online platform to promote activities, places to visit, and participation in events. Casino and event operators, as well as tourism agencies, produce rich media to promote their services through the Internet. Much valuable tourism information is available to tourists who visit Macao and share their experiences in blogs, community sites, discussion forums, and review bulletin boards in travel-specific social media sites.

In Macao, the government and business operators must be aware of the activities of visitors and their experiences during their trips. Key questions include, "Where did the visitors go?" "What were their experiences?" and "What are the overall images of Macao that they perceived?" By identifying one's experiences when traveling to Macao, government departments, such as the Macao Tourism Board, and gaming conglomerates can establish gaming-related and other infrastructures to maintain the competitiveness of Macao. Therefore, the present study focuses on invariance among collected documents, instead of focusing on that in document-level analysis and simple keyword approaches that have failed to explore interconnectedness among keywords or key-phrases.

#### 4.2 Content-Based Analysis of Social Media Sources in Macao

With its unique reputation of being the world's gaming center and a cultural heritage site, we use Macao as an illustrative example of the application of our social media analysis methodology (Figure 1). Macao's image in the West has a significant role in the development of its tourism industry. Thus, business and government decision-makers need to know how tourists perceive the city.

##### 4.2.1. Phase 1: Definition of goal and scope

The study aims to identify the destination image of Macao. The scope involved a content-based analysis of Macao tourism-related social media websites.

##### 4.2.2. Phase 2: Data collection

We identified the sources of online information. Webpages in Travelblog.org and Travepod.com were collected and listed after entering keywords, such as "Macao + travel + blogs," in Google search. Other webpages that described travel experiences in Macao were identified with "travel forum" and "travel review" as categories. The top three websites were Tripadvisor.com, Virtualtourist.com, and Travelblog.org. All webpages were reviewed to ensure that the content pertained to tourist experiences in Macao.

Subsequently, the total sample size,  $N_{webpages}$ , is determined. By using Eq. (1) and the number of attributes as 34 or less [Beerli & Martin, 2004; Echtner & Ritchie, 2003], we calculated the minimum number of webpages to be downloaded as 340. Text documents were downloaded from 500 webpages to expand the contextual information for this study.

The next step aimed to determine the number of webpages per category of social media sources with Eq. (2). We examined 125 webpages from Tripadvisor.com, Virtualtourist.com, Travelblog.org, and Travepod.com. The popularity of these four selected sites made them containing several-thousand active comments about Macao. Likewise, text documents were downloaded.

Data smoothing operations were performed to obtain a consistent interpretation by computer-aided lexical and/or statistical analysis. These operations included stripping HTML tags, using consistent spelling for frequently used words in all files (e.g., Macao instead of Macau), changing multiple words to one-word format (e.g., from Hong Kong to Hongkong), changing plural nouns to singular forms (e.g., from casinos to casino), and ensuring that descriptive words (e.g., helpful and nice) were used in a positive context. The standard word processors features, such as search and replace, were used to complete the data smoothing step.

#### 4.2.3. Phase 3: Data transformation

The downloaded social media content was subjected to the following three steps of computer-aided data transformation: compilation of a list of key image attributes through WordSmith, grouping of attributes into factors using SPSS, and generation of a concept map of the image using Leximancer.

First, keywords (key attributes) were identified through WordSmith. As stated in the previous section, the 500 text files selected were combined to form a master file, which was subsequently compared with the British National Corpus. Hundreds of keywords were identified and sorted based on their keyness values. The first 68 keywords, twice the expected number of key attributes, were identified. The word “Macao” was excluded from keywords because it is the domain of the study, rather than an attribute. The 68 keywords were regarded as the key image attributes of Macao.

Second, the 68 identified keywords or key attributes were grouped into a few underlying factors through SPSS. Each text file from a webpage was entered separately, and the frequency counts of the keywords were determined. A frequency table of 68 attributes (i.e., keywords) and 500 responses (i.e., webpages) was created. We checked the frequency counts of the keywords and the total frequency count of each response. Responses with a total frequency count of less than 7 (10% of the number of key attributes) were deleted. Thus, 68 attributes and 440 responses were left in the frequency table. The refined text files were combined to form a new master file, which was compared with the lexical reference. The analyzed results showed no changes in the keywords based on keyness values.

Factor analysis identifies clusters of closely related variables. Hence, a factor analysis of the frequency table was conducted with IBM SPSS 20.0. KMO and Bartlett’s test of sphericity were used to measure the sampling adequacy for factor analysis. The calculated KMO value was 0.78, and Bartlett’s test of sphericity revealed statistically significant results ( $p < 0.001$ ); factor analysis was thus appropriate [Hair et al., 2006]. Principal component analysis was conducted via varimax rotation, with factors determined by the first elbow of the Scree plot [Hair et al., 2006] retained. We retained only the attributes with rotated factor loadings higher than 0.40 and cross-loadings less than 0.40 as constitutive attributes. With 4 items dropped, our final list comprised 64 attributes (i.e., keywords). Principal component analysis decomposed the matrix  $X$  into three components (i.e.,  $X = U\Sigma V^T$ , where  $U$  is an  $N \times N$  orthonormal matrix,  $V$  is an  $m \times m$  orthonormal matrix, and  $\Sigma$  is an  $N \times m$  diagonal elements of which consist of  $\sigma$  non-zero singular values of  $X$ ) [To, 1994]. The maximum number of  $\sigma$  equals to the measured variables, that is, 64 after dropping 4 items for our case. Furthermore, the number of factors depended on the number of eigenvalues that were greater-than-one [Cliff, 1988].) The rotated solution provided a list of factors that were easily interpretable. Another analysis was conducted by oblimin rotation, and a similar factor structure was determined. Table 2 shows the results of the exploratory factor analysis by varimax rotation.

The nine factors derived from the 64 key image attributes after four items were dropped include: 1) heritage traits, 2) hotel facilities, 3) transportation and accessibility, 4) history and culture, 5) casinos and gaming, 6) landmarks and attractions, 7) entertainment and shows, 8) sightseeing and tour, and 9) tourism development (Table 2). The labeling of the factors was determined by the nature of the items with high loadings, following the suggestions of Hair et al. [2006]. These nine factors accounted for 55.44% of the total variance. The Cronbach’s alpha coefficients ranged from 0.61 to 0.87, with all values exceeding the minimum acceptable level of 0.6 for internal consistency in the early stage of research. Factors 2, 5, and 7 represented the theme “casino hotels with entertainment and shows” and constituted 19.10% of the total variance. The theme “cultural heritage” was formed by factors 1, 4, and 6, which constituted 20.92% of the total variance.



Table 2: Factor analysis of key attributes obtained from social media sites

Factor	Attribute	Factor Loading	Alpha	Variance (%)	Cumulative (%)
1	Jesuit	0.853	0.811	10.15	10.15
	Built	0.827			
	Church	0.770			
	Guia Fortress	0.705			
	Facade	0.677			
	Ruins of St. Paul's	0.663			
	Guia Lighthouse	0.639			
	Building	0.627			
	Fortress	0.602			
	Chinese	0.592			
	Architecture	0.571			
	Square	0.557			
	City	0.455			
2	Facility	0.940	0.871	8.65	18.80
	Multi	0.928			
	Room	0.887			
	International	0.782			
	Centre	0.673			
	Guest	0.639			
	Service	0.556			
	Taipa	0.476			
Located	0.432				
3	Ferry	0.800	0.710	6.53	25.33
	Bus	0.715			
	Hong Kong	0.709			
	Ticket	0.588			
	Harbor	0.552			
	Taxi	0.546			
	Zhu-hai	0.537			
Visit	0.476				
4	Historic	0.772	0.672	5.81	31.14
	Cultural	0.766			
	Heritage	0.605			
	Museum	0.573			
	Macanese	0.507			
Grand Prix	0.492				
5	Casino	0.931	0.683	5.69	36.83
	Baccarat	0.733			
	Gaming	0.698			
	Las Vegas	0.553			
	Lisboa	0.543			
	Cotai Strip	0.457			
	Wynn	0.452			
	Gambling	0.419			
MGM	0.417				
6	Statue	0.780	0.614	4.96	41.79
	Kun-iam Temple	0.703			
	A Ma Temple	0.550			
	Restaurant	0.528			
	Garden	0.496			
7	Dancing Water	0.792	0.661	4.86	46.65
	Show	0.731			
	City of Dreams	0.662			
	Entertainment	0.597			
	3D Dragon	0.551			
	Zaia	0.450			
8	Tour	0.812	0.620	4.61	51.26
	Na Tcha Temple	0.785			
	Macao Tower	0.669			
	Mandarin	0.484			
9	Mainland	0.701	0.615	4.18	55.44
	Tourism	0.669			
	Asia	0.629			
	Destination	0.505			

Finally, the master text file (with 440 responses) was analyzed using Leximancer and concepts formed based on the co-occurrences of the keywords or key attributes. The primary concepts and their relationships with other concepts were visually represented on a concept map (Figure 2), which is a big picture of the impression of tourists on Macao. Figure 2 is a heat map. Hot colors, that is, red and orange, denote more important image themes, whereas cold colors, such as blue and green, denote less important image themes. The map also displays the frequency and extent of the connectedness of concepts by the size of the concept point. Several key concepts of the destination image of Macao include “city,” “center,” “heritage,” “building,” “church,” “fortress,” and “casino.” We can observe two clusters of interrelated destination images, namely, casino hotels and cultural heritage. City basics, such as transportation and location, serve as links between the two clusters.

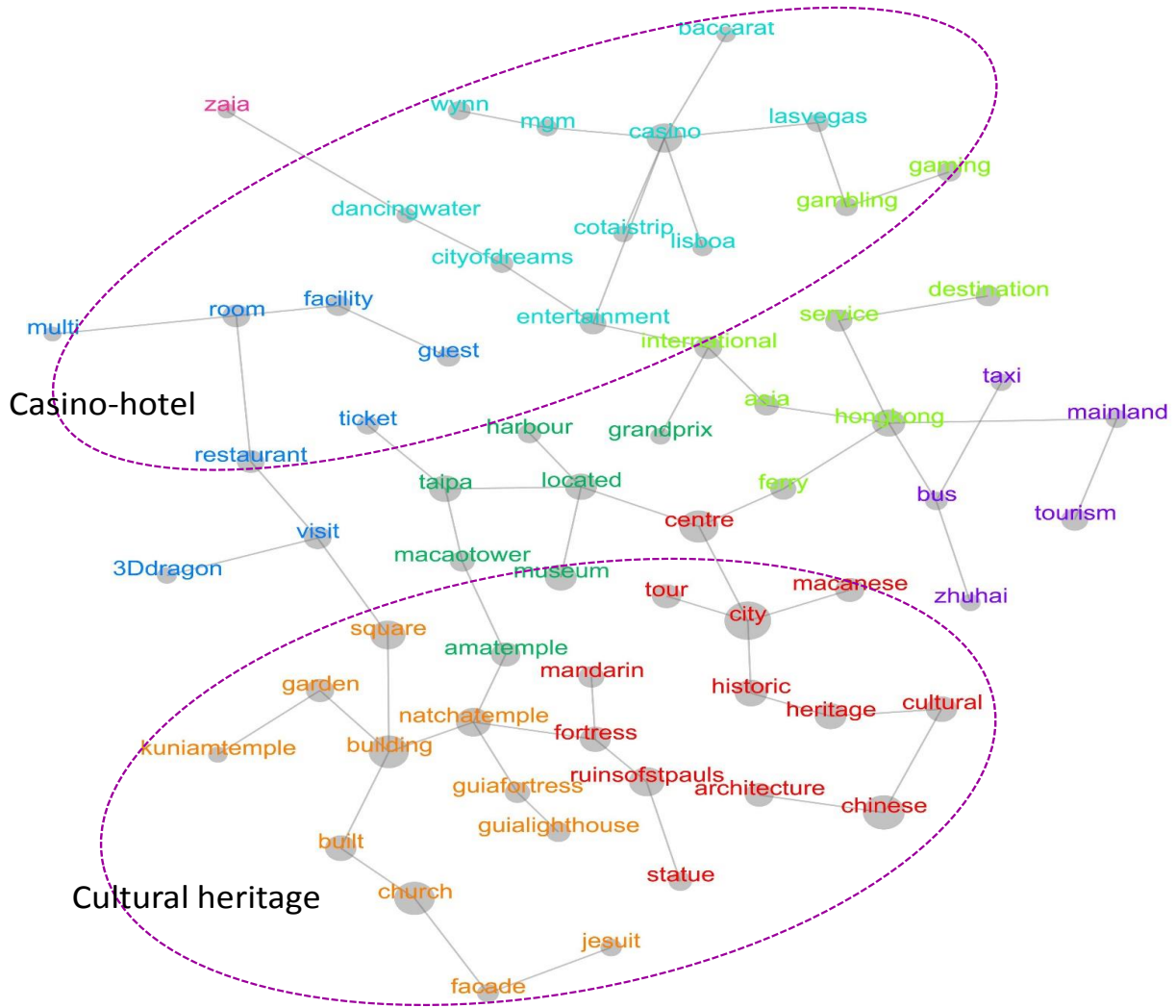


Figure 2: Concept map of the tourists' impression of Macao

#### 4.2.4. Phase 4: Interpretation of results

The destination image of Macao, as determined by the content analysis of social media, is a casino city with cultural heritage as its backdrop. Aside from being known as the world's casino city, Macao is also a city of culture with its unique Sino-Portuguese architectural and cultural heritage. The cultural appeal of Macao is often overshadowed by the image of the casino city. Macao is often cited as a city with more churches than the Vatican, while having more gambling tables than Las Vegas. The marriage of two seemingly incompatible features of gambling and heritage lends uniqueness to the destination image of Macao. Web-based social media has provided a different perspective of what Macao offers.

However, this case study has limitations. One is its cross-sectional research nature, which implies that any changes in web content may result in different analysis outcomes. The other is its reliance on the capability of source engines to identify information sources. As time passes, new social media content emerges. Thus, the entire process can be applied iteratively to monitor the evolution of identified concepts and themes.

## 5. Discussion

The decision making of managers is crucial to the survival of an organization, including government departments and business enterprises. With regard to the decision support process, the application of the proposed methodology (Figure 1) highlights two areas of discussion, namely, social media content analysis as an information management tool and social media as a platform for building trust and information exchange.

### 5.1. Social Media Content Analysis as Information Management Tool

The effectiveness of decision making significantly depends on how managers define problems they face, as well as whether they can achieve a particular objective and accurate information, and subsequently, derive and implement appropriate solution(s). Traditionally, organizations use market research techniques, such as focus groups and surveys, to gain information about the attitudes, behaviors, and characteristics of customers. However, using these techniques allows participants to be fully aware of the objectives of the studies, and thus, do not usually express their true feelings. With the advent of social media, people today openly express their feelings with regard to a variety of topics. Social media creates opportunities for organizations to listen to customers without interfering with their thought processes. Thus, customer interactions in social media can be observed naturally [Trusov et al., 2009].

The proposed approach for social media content analysis has several advantages. First, recall and respondent bias are minimized because the posts analyzed in this study are made by the participants prior to the selection of these posts [Herring, 2010]. Second, the approach integrates comprehensive information into the study compared with other quantitative techniques. Third, the approach identifies completed actions, whereas structured methods (e.g., a survey) determine only what people believe they would do. Fourth, the analysis captures “the wisdom of the crowds” and provides a user-generated perspective of the situation. Finally, the proposed methodology is a grounded theory approach [Kuziemsky et al., 2007; Strauss & Corbin, 1998] that limits the scope of a study while allowing for concept development from the ground up as the decision maker gathers and analyzes data.

### 5.2. Social Media as a Platform for Trust Building and Information Exchange

Regardless of the nature of the decisions involved (e.g. whether personal or business), trust and information exchange are essential components of the decision support process. Thus, individuals must obtain relevant information from trusted sources. Karimov et al. [2011] asserted that social media applications effectively induce the initial trust of people toward unfamiliar e-retailers; nevertheless, additional studies must be conducted to confirm this emergent effect. The interactivity and communication that transpire in social media sites may lead to a high degree of trust among participants [Tang et al., 2012]. Information exchange, which is often expressed as positive or negative word-of-mouth recommendations, is also a typical communal norm in social media. Social media sites are a platform for information exchange that is characterized by mutual trust among participants.

## 6. Conclusion

As the Internet increases in size, mode, and diversity, exploring web content, particularly social media, and transforming such content into concepts have become a challenge to electronic commerce researchers, business practitioners, and policymakers. In this paper, we introduced a systematic methodology to convert text files from social media to concepts that are repeatable, easily interpretable, and visible with a concept map. We have likewise established several criteria to identify sources, minimum sample size (i.e., total number of webpages), the sample size of each category of sources, and the number of key variables (or attributes).

The grounded theory approach indicates that aside from collecting relevant data for analysis, we also need to allow concepts and themes to emerge from the ground up. User-generated content, such as word-of-mouth, can be systematically monitored to understand the beliefs, values, attitudes, perceptions, intentions, and behaviors of users. The proposed methodology, whether quantitative or qualitative in nature, benefits business practitioners, considering that cost, time, and human errors are kept to a minimum during data processing and analysis.

### 6.1. Implications for Academics and Business Practitioners

Despite the important role of social media in shaping societies, content analysis in media remains an under-researched area in electronic commerce. In the past, its poor quality has been the subject of criticism, as evidenced by issues such as duplicated content and spam. Nevertheless, as the number of people who use social media and create content increases, high-quality content has become increasingly available [Agichtein et al., 2008]. Thus, the challenge for electronic commerce research is to identify the type of information or text to be extracted from social media and

to utilize such information for a particular purpose. The methodology we present in this paper contributes to the resolution of this issue.

The tourism industry has significantly contributed to the gross domestic and national product of several economies in the world. As such, this topic was selected to illustrate the implementation of the proposed methodology. Other applications may include reputation analysis and opinion extraction.

#### 6.2. Limitations and Future Research

The present study is not without limitations. The proposed methodology requires validation and confirmation in future studies, particularly in light of the existence of contributor or self-selection bias in online postings. If social media were manipulated (say, by reputation managers), opinions captured in the related content analysis would be manipulated as well. As previously noted, the proposed methodology can be applied in similar analyses of other economies, such as Hong Kong, Singapore, and Shanghai. Moreover, researchers can extend this methodology to areas in business and the social sciences, such as marketing, branding, politics, and sociology. Software developers can integrate and automate techniques mentioned in the paper, such as keyness, exploratory factor analysis, and latent and relational semantic analysis, for comprehensive social media content analysis.

#### Acknowledgement

This research was supported by a grant (RP/ESCE-01/2013) from Macao Polytechnic Institute.

#### REFERENCES

- Aggarwal, C.C., and H. Wang, "Text Mining in Social Networks," In *Social Network Data Analytics* (pp. 353-378), Springer US, 2011.
- Abrahams, A.S., J. Jiao, G.A. Wang, and W.G. Fan, "Vehicle Defect Discovery from Social Media," *Decision Support Systems*, Vol. 54, No. 1:87-97, 2012.
- Agichtein, E., C. Castillo, D. Donato, A. Gionis, and G. Mishne, "Finding High-Quality Content in Social Media," *Proceedings of the International Conference on Web Search and Web Data Mining*, 183-194, 2008.
- Airoldi, E.M., X. Bai, and R. Padman, "Markov-blanket and Meta-heuristic Search: Sentiment Extraction from Unstructured Text," *Lecture Notes in Computer Science*, Vol. 3932:167-187, 2006.
- Arnold, E., A. Bruton, and C. Ellis-Hill, "Adherence to Pulmonary Rehabilitation: A Qualitative Study," *Respiratory Medicine*, Vol. 100, No. 10:1716-1723, 2006.
- Arnold, S.E., Beyond Search-and-Retrieval: Enterprise Text Mining with SAS®. SAS white paper [www.sas.com/technologies/analytics/datamining/textminer](http://www.sas.com/technologies/analytics/datamining/textminer), <http://www.cxoamerica.com/pastissue/printarticle.asp>, 2012 .
- Averya, E., R. Lariscyb, and K.D. Sweetserb, "Social Media and Shared- or Divergent-uses? A Coorientation Analysis of Public Relations Practitioners and Journalists," *International Journal of Strategic Communication*, Vol. 4, No. 3:189-205, 2010.
- Bai, X., "Predicting Consumer Sentiments from Online Text," *Decision Support Systems*, Vol. 50, No. 4:732-742, 2011.
- Banerjee, B., and S. Basu, "Topic Models over Text Streams: A Study of Batch and Online Unsupervised Learning," *Proceedings of the SIAM International Conference on Data Mining (SDM)*, 2007.
- Becker, H., M. Naaman, and L. Gravano, "Event Identification in Social Media," *Proceedings of the ACM SIGMOD Workshop on the Web and Databases (WebDB '09)*, 2009.
- Berli, A. and J.D. Martin, "Factors Influencing Destination Image," *Annals of Tourism Research*, Vol. 31, No. 3:657-681, 2004.
- Blei, D.M., A.Y. Ng, and M.I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning*, Vol. 3: 993-1022, 2003.
- Bondi, M. and M. Scott, *Keyness in Texts*, Amsterdam: John Benjamins B.V., 2010.
- CenStatd, Hong Kong Annual Digest of Statistics - 2013 Edition, Hong Kong: Hong Kong Census and Statistics Department, 2013.
- Chiu, C.M., M.H. Hsu, and E.T.G. Wang, "Understanding Knowledge Sharing in Virtual Communities: An Integration of Social Capital and Social Cognitive Theories," *Decision Support Systems*, Vol. 42, No. 3:1872-1888, 2006.
- Cliff, N., "The Eigenvalues-Greater-than-One Rule and the Reliability of Components," *Psychological Bulletin*, Vol. 103, No. 2:276-279, 1998.
- Comley, P., "Online Research Communities: A User Guide," *International Journal of Market Research*, Vol. 50, No. 5:679-694, 2008.

- Deerwester, S., S.T. Dusmais, T.K. Landauer, G.W. Furnas, and R.A. Harshman, "Indexing by Latent Semantic Analysis," *Journal of the American Society for Information Science*, Vol. 41: 391-407, 1990.
- DSEC, Statistical Information System of Macao – 2013 Edition, Macao: Statistics Section, Civil Affairs Department of Macao, 2013.
- Dwivedi, Y.K., M. Kiang, B. Lal, and M.D. Williams, "Profiling Research Published in the Journal of Electronic Commerce Research," *Journal of Electronic Commerce Research*, Vol. 9, No. 2:77-91, 2008.
- Echtner, M.C. and J.R.B. Ritchie, "The Meaning and Measurement of Destination Image," *Journal of Tourism Studies*, Vol. 14, No. 1:37-48, 2003.
- Evans, W., "Computer-supported Content Analysis: Trends, Tools, and Techniques," *Social Science Computer Review*, Vol. 14, No. 3:269-279, 1996.
- GdGov, Monthly Statistics on Tourism, Guangdong: Guangdong Provincial Government, 2013.
- Glaser, B.G. and A.L. Strauss, *The Discovery of Grounded Theory: Strategies for Qualitative Research*, New Jersey: Transaction Publishers, 2009.
- Godbole, N., M. Srinivasiah, and S. Skiena, "Large-Scale Sentiment Analysis for News and Blogs," *Proceedings of the International Conference on Weblogs and Social Media (ICWSM'2007)*, 2007.
- Goodrich, J.N., "A New Approach to Image Analysis through Multidimensional Scaling," *Journal of Travel Research*, Vol. 16, No. 3:3-7, 1977.
- Hair, J.F. Jr., W.C. Black, B.J. Babin, R.E. Anderson, and R.L. Tatham, *Multivariate Data Analysis*, 6<sup>th</sup>ed, N.J., Upper Saddle River: Prentice Hall, 2006.
- Herring, S.C., "Web Content Analysis: Expanding the Paradigm," In *International Handbook of Internet Research*, New York: Springer, 233-249, 2010.
- Hinchcliffe, D. Profitably running an online business in the Web 2.0 era. Luettavissa: [http://Web2.socialcomputingmagazine.com/running\\_an\\_online\\_business\\_profitably\\_in\\_the\\_Web\\_20\\_era.htm](http://Web2.socialcomputingmagazine.com/running_an_online_business_profitably_in_the_Web_20_era.htm), 2008.
- Holsti, O.R., *Content Analysis for the Social Sciences and Humanities*, M.A., Reading: Addison-Wesley, 1969.
- Hu, X., and H. Liu, "Text Analytics in Social Media," In *Mining Text Data* (pp. 385-414), Springer US, 2012.
- Hurt, C.D., "Automatically Generated Keywords: A Comparison to Author-Generated Keywords in the Sciences," *Journal of Information and Organizational Sciences*, Vol. 34, No. 1:81-88, 2010.
- Jansen, B.J., K. Sobel, and G. Cook, "Classifying Ecommerce Information Sharing Behaviour by Youths on Social Networking Sites," *Journal of Information Science*, Vol. 37, No. 2:120-136, 2011.
- Karimov, F.P., M. Brengman, and L. Van Hove, "The Effect of Website Design Dimensions on Initial Trust: A Synthesis of the Empirical Literature," *Journal of Electronic Commerce Research*, Vol. 12, No. 4:272-301, 2011.
- Kim, C., M.H. Jin, J. Kim, and N. Shin, "User Perception of the Quality, Value, and Utility of User-Generated Content," *Journal of Electronic Commerce Research*, Vol. 13, No. 4:305-319, 2012.
- Krippendorff, K., *Content Analysis: An Introduction to Its Methodology*, 2<sup>nd</sup> ed., CA, Thousand Oaks: Sage Publications, 2004.
- Kuziemsky, C.E., G.M. Downing, F.M. Black, and F. Lau, "A Grounded Theory Guided Approach to Palliative Care Systems Design," *International Journal of Medical Informatics*, Vol. 76S:S141-S148, 2007.
- Lai, L.S.L. and E. Turban, "Groups Formation and Operations in the Web 2.0 Environment and Social Networks," *Group Decision and Negotiation*, Vol. 17, No. 5:387-402, 2008.
- Lai, L.S.L. and W.M. To, "Importance-performance Analysis for Public Management Decision Making: An Empirical Study of China's Macao Special Administrative Region," *Management Decision*, Vol. 48, No. 2: 277-295, 2010.
- Lai, L.S.L. and W.M. To, "Internet Diffusion in China: Economic and Social Implications", *IT Professional*, Vol. 14, No. 6:16-21, 2012.
- Lane, P.C.R., D. Clarke, and P. Hender, "On Developing Robust Models for Favourability Analysis: Model Choice, Feature Sets and Imbalanced Data," *Decision Support Systems*, Vol. 53, No. 4:712-718, 2012.
- Ma, B., D. Zhang, Z. Yan, and T. Kim, "An LDA and Synonym Lexicon Based Approach to Product Feature Extraction from Online Consumer Product Reviews," *Journal of Electronic Commerce Research*, Vol. 14 No. 4:304-314, 2013.
- Merriam, S.B., *Qualitative Research: A Guide to Design and Implementation*, New York: John Wiley and Sons, 2009.
- Morinaga, S., K. Yamanishi, K. Tateishi, and T. Fukushima, T., "Mining Product Reputations on the Web," *Proceedings of the 8<sup>th</sup> ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 341-349, 2002.
- Neuendorf, K.A., *The Content Analysis Guidebook*, CA, Thousand Oaks: Sage Publications, 2002.

- Newman, D.J., and S. Block, "Probabilistic Topic Decomposition of an Eighteenth-century American Newspaper," *Journal of the American Society for Information Science and Technology*, Vol. 57: 753-767, 2006.
- Nunnally, J.C., *Psychometric Theory*, 2<sup>nd</sup> Ed., New York: McGraw-Hill, 1978.
- O'Connor, B., R. Balasubramanian, B. Routledge, and N. Smith, "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series," *Proceedings of the International AAAI Conference on Weblogs and Social Media (ISWSM)*, 2010.
- Pang, B., L. Lee, and S. Vaithyanathan, "Thumbs Up? Sentiment Classification using Machine Learning Techniques," *Proceedings of the 2002 Conference on Empirical Methods in Natural Language*, 79-86, 2002.
- Phelps, A., "Holiday Destination Image - the Problem of Assessment: An Example Developed in Menorca," *Tourism Management*, Vol. 7, No. 3:168-180, 1986.
- Pike, S., "Destination Image Analysis - a Review of 142 papers from 1973 to 2000," *Tourism Management*, Vol. 23, No. 5:541-549, 2002.
- Ramage, D., C.D. Manning, and S. Dumais, "Partially Labeled Topic Models for Interpretable Text Mining," *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2011.
- Rayson, P. and R. Garside, "Comparing Corpora using Frequency Profiling," *Proceedings of the Workshop on Comparing Corpora, Association for Computational Linguistics*, 1-6, 2000.
- Ryan, C. and J. Cave, "Structuring Destination Image: A Qualitative Approach," *Journal of Travel Research*, Vol. 44, No. 2:143-150, 2005.
- Saggion, H. and A. Funk, "Extracting Opinions and Facts for Business Intelligence," *RNTI Journal*, Vol. E (17):119-146, 2009.
- Saggion, H., A. Funk, D. Maynard, and K. Bontcheva, "Ontology-based Information Extraction for Business Applications," *Proceedings of the 6<sup>th</sup> International Semantic Web Conference (ISWC 2007)*, 2007.
- Scott, M., "Developing WordSmith," *International Journal of English Studies*, Vol. 8, No. 1:95-106, 2008.
- Scott, M., *Wordsmith Tools*, Oxford: Oxford University Press, 1996.
- Scott, N.R. and A.E. Smith, "Use of Automated Content Analysis Techniques for Event Image Assessment," *Tourism Recreation Research*, Vol. 30, No. 2:87-91, 2005.
- Shannon, C.E., "Communication in the Presence of Noise," *Proceedings of the Institute of Radio Engineers*, Vol. 37, No. 1:10-21, 1949.
- Smith, A.E., "Machine Mapping of Document Collections: the Leximancer System," *Proceedings of the Fifth Australasian Document Computing Symposium*, Sunshine Coast, Australia, 2000.
- Srivastava, A. and K.M. Pandey, K.M., "Social Media Marketing: An Impeccable Approach to e-Commerce", *Management Insight*, Vol. 8, No. 2, 2013.
- Strauss, A. and J. Corbin, *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, CA, Thousand Oaks: Sage Publications, 1998.
- Tang, J., H. Gao, and H. Liu, "mTrust: Discerning Multi-faceted Trust in a Connected World," *Proceedings of the fifth ACM International Conference on Web Search and Data Mining*, 93-102, 2012.
- To, W.M. "Application of Singular Value Decomposition to Direct Matrix Update Method," *AIAA Journal*, Vol. 32, No. 10:2124-2126, 1994.
- To, W.M., L.S.L. Lai, and A.W.L. Chung, "Cloud Computing in China: Barriers and Potential," *IT Professional*, Vol. 15, No. 3:48-53, 2013.
- Trusov, M., R.E. Bucklin, and K. Pauwels, "Effects of Word-of-mouth Versus Traditional Marketing: Findings from an Internet Social Networking Site," *Journal of Marketing*, Vol. 73, No. 5:90-102, 2009.
- Turney, P.D. and M.L. Littman, "Measuring Praise and Criticism: Inference of Semantic Orientation from Association," *ACM Transactions on Information Systems*, Vol. 21, No. 4:315-346, 2003.
- Turban, E., J. Strauss, and L.S.L. Lai, *Social Commerce – An IS and Marketing Perspective*, Springer, 2015.
- Turri, A.M., K.H. Smith, and E. Kemp, "Developing Affective Brand Commitment through Social Media," *Journal of Electronic Commerce Research*, Vol. 14, No. 3:201-214, 2013.
- Wang, C., B. Thiesson, C. Meek, and D. Blei, "Markov Topic Models," *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2009.
- Wang, X., and Y. Li, (2014), "Trust, Psychological Need, and Motivation to Produce User-Generated Content: A Self-Determination Perspective," *Journal of Electronic Commerce Research*, Vol. 15, No. 3:241-253, 2014.
- Watson, M., A.E. Smith, and S. Watter, "Leximancer Concept Mapping of Patient Case Studies," *Lecture Notes in Computer Science*, Vol. 3683:1232-1238, 2005.
- Witz, K.G., "Awakening to" an Aspect in the Other On Developing Insights and Concepts in Qualitative Research," *Qualitative Inquiry*, Vol. 13, No. 2:235-258, 2007.

- Ye, Q., Z. Zhang, and R. Law, "Sentiment Classification of Online Reviews to Travel Destinations by Supervised Machine Learning Approaches," *Expert Systems with Applications*, Vol. 36:6527-6535, 2009.
- Zwass, Z., "Electronic Commerce: Structures and Issues," *International Journal of Electronic Commerce*, Vol. 1, No. 1:3-23, 1996.