

# Journal of Travel Research

<http://jtr.sagepub.com/>

---

## **Facilitating Content Analysis in Tourism Research**

Svetlana Stepchenkova, Andrei P. Kirilenko and Alastair M. Morrison

*Journal of Travel Research* 2009 47: 454 originally published online 13 November 2008

DOI: 10.1177/0047287508326509

The online version of this article can be found at:

<http://jtr.sagepub.com/content/47/4/454>

---

Published by:



<http://www.sagepublications.com>

On behalf of:



[Travel and Tourism Research Association](http://www.tta.org.uk)

**Additional services and information for *Journal of Travel Research* can be found at:**

**Email Alerts:** <http://jtr.sagepub.com/cgi/alerts>

**Subscriptions:** <http://jtr.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.com/journalsPermissions.nav>

**Citations:** <http://jtr.sagepub.com/content/47/4/454.refs.html>

# Facilitating Content Analysis in Tourism Research

Svetlana Stepchenkova  
*Purdue University*

Andrei P. Kirilenko  
*University of North Dakota*

Alastair M. Morrison  
*Purdue University*

The article proposes a new methodological approach to facilitate content analysis of electronic textual data in a more efficient and transparent way. The textual data are processed iteratively by two software products, CATPAC and WORDER. This approach permits smoothing of the original textual data, identification of the variables of interest, frequency count of the occurrences of these variables in the texts being processed, storage of frequency results in general purpose statistical packages, and subsequent dimensional reduction of word-frequency data by means of factor analysis. Application of this methodology is illustrated on three examples of destination-image studies, which cover content analyses of open-ended responses to e-survey questions, texts from tourism Web sites, and newspaper articles. Advantages and disadvantages of the proposed research technique, its contribution to tourism studies, and the place of the approach within the quantitative paradigm of content analysis are also discussed.

**Keywords:** *CATA software; CATPAC; content analysis; destination image; tourism research; WORDER*

Content analysis is a well-established research methodology commonly used in social sciences to analyze communications (Holsti 1969). Over the past two decades, content-analysis research has remarkably benefited from the exponentially increasing volume of electronic data, including articles in general media databases, communications in virtual communities, and textual and pictorial materials from Web sites (Neuendorf 2002; Rainer and Hall 2003; Romano et al. 2003; Wickham and Woods 2005). The escalating employment of e-surveys by social scientists (Sills and Song 2002) also contributed to the availability of electronic data. Immense volumes of easily accessible textual material, speed and simplicity of the data-collection process, lack of complications associated with human subjects, and advances in development of various computer programs to support textual data analysis are factors that stimulate the use of content-analysis research in social sciences (Macnamara 2003; Miles and Weitzman 1994; Romano et al. 2003).

The methodology of content analysis has been developing since the early 1920s in such areas of scientific inquiry as political science, psychology, and communications; it was also adopted in tourism research, though

to a lesser extent. In the authors' view, content analysis used in tourism studies is, generally, less sophisticated than in other disciplines; it is especially true for research that falls within the quantitative paradigm of content analysis. Quantitative methods are considered more conducive to making statistical inferences, comparisons, and hypothesis testing; however, the analyses are repeatedly limited to simple word-frequency counts. Mehmetoglu and Dann (2003) noted that tourism researchers are reluctant to use computer-assisted content analysis, although they referred to a more qualitatively oriented class of content-analysis methods like semiotic analysis of content. While computers can reduce quite effectively the tedium of data preparation and the time necessary for handling large volumes of textual data, the absence of clearly outlined ways to discern categories from textual data in computer-assisted content-analysis projects and methods for subsequent dimensional reduction of word-frequency data has slowed the adoption of computer-assisted content analysis in tourism research.

This article proposes a methodological approach to analyzing multiple files of textual data typical in tourism studies in a transparent, replicable, and effective way. The approach proceeds with data preparation, identification

of key variables, obtaining the word-frequency matrix, and subsequent dimensional reduction of word-frequency data. Obtaining a matrix of word frequencies from multiple units of qualitative data allows more sophisticated statistical analyses of data and, ultimately, hypothesis testing. The approach uses an efficient combination of two computer programs, CATPAC and WORDER; however, the methodology is not dependent on this particular software tandem. Other programs that perform the same functions can be used, and the choice, as always, is with the researcher. The objective of the article is to show that the proposed methodology is firmly grounded in the theory and practices of content analysis and is both simple and efficient enough to facilitate statistical data analysis in tourism studies.

The article is organized as follows. The Literature Review section discusses two general classes of content analysis, quantitative and qualitative, with an emphasis on the former as more relevant to the proposed approach. It also provides a brief overview of computer-aided text analysis (CATA) software and discusses its general limitations and usage in tourism-related studies. The approach itself is described in the Method section. It combines the two CATA programs, CATPAC and WORDER, to effectively deal with data preparation, identification of image variables and processing of a large number of textual files of similar types, and use dimensional reduction of word-frequency data, for example, by means of factor analysis. Three examples of approach implementation in destination-image studies are given in the Application Examples section; each example focuses on certain aspects of the approach. The Theoretical Considerations section discusses the approach in the context of content-analysis methodology, particularly taxonomy proposed by Roberts (2000), as well as issues regarding its transparency, replicability, sampling and generalizability; enrichment of the data with contextual variables; and use of factor analysis for dimensional reduction of word-frequency data. The article concludes with the relevance of the proposed methodological approach to tourism research—its strengths and limitations.

## Literature Review

### Content Analysis

Content analysis is a nonobtrusive research methodology used to study a wide range of textual data—for example, various types of media messages, interview transcripts, discussion boards in virtual communities, and/or travel diaries. It is “a technique which aims at

describing, with optimum objectivity, precision, and generality, what is said on a given subject in a given place at a given time” (Lasswell, Lerner, and Pool 1952, p. 34). Berelson (1952, p. 18) summarized content analysis as a “research technique for the objective, systematic, and quantitative description of the manifest content of communication.” More recently, Weber (1990, p. 9) described content analysis as “a research method that uses a set of procedures to make valid inferences from text.” Content analysis examines textual data for patterns and structures, singles out the key features to which researchers want to pay attention, develops categories, and aggregates them into perceptible constructs in order to seize text meaning (Gray and Densten 1998; Shoemaker and Reese 1996). Content analysis is capable of capturing a richer sense of concepts within the data due to its qualitative basis and, at the same time, can be subjected to quantitative data-analysis techniques (Insch and Moore 1997). There are two general classes of epistemologies employed for content analysis in social sciences: qualitative and quantitative. The former term refers to nonstatistical and exploratory methods, which involve inductive reasoning (Berg 1995), while the latter term refers to methods that are capable of providing statistical inferences from text populations.

A central idea of quantitative content analysis is that “many words of text can be classified into much fewer content categories” (Weber 1990, p. 7). The methodology of extracting content categories from the text, counting their occurrences in the sampled text blocks, and analyzing associations between categories using the frequency matrix was developed by the mid-20th century, primarily by a group of Harvard researchers, and is often referred to as contingency analysis (Pool 1959; Roberts 2000). George (1959), one of the pioneers of content analysis, was critical of the use of contingency analysis, saying that this method was not sensitive enough to the intended meaning of the author. Indeed, contingency analysis assumes that “what an author says is what he means” (Pool 1959, p. 4) and cannot take into account such text features as, for example, figures of speech or irony. George’s opinion is supported by Shoemaker and Reese (1996, p. 32) who argue that the process of reducing large volumes of text to quantitative data “does not provide a complete picture of meaning and contextual codes, since texts may contain many other forms of emphasis besides sheer repetition.” Newbold, Boyd-Barrett, and Van Den Bulck (2002, p. 80) agree that “there is no simple relationship between media texts and their impact, and it would be too simplistic to base decisions in this regard on mere figures obtained from a statistical content analysis.” Moreover, quantitative content

analysis does not always account for source credibility, political or social context of messages being examined, and audience characteristics such as age, sex, or education (Macnamara 2003). However, despite its limitations, quantitative content analysis has long been employed in social studies due to its clear methodological reasoning based on the assumption that the most frequent theme in the text is the most important, as well as to the ability to incorporate such scientific methods as “*a priori* design, reliability, validity, generalizability, replicability, and hypothesis testing” (Neuendorf 2002, p. 10).

From a philosophical perspective, quantitative tradition of content analysis is based on the positivist premise that “there is something like an objective reality (social facts) ‘out there’ that can be observed, measured, analyzed and thus understood” (Newbold, Boyd-Barrett, and Van Den Bulck 2002, p. 59); therefore, decontextualization of the textual material and selection of the outsider variables for analysis of social phenomena are the main issues in quantitative paradigm. In contrast, the qualitative epistemologies share the view that “reality” is a social and cultural creation, which can only be interpreted, approximated but not fully apprehended; thus, in qualitative tradition, the focus is on complexity, context, and detail (Denzin and Lincoln 1994). Qualitative tradition heavily relies on the researcher’s reading of the content and includes such approaches as rhetorical, narrative, semiotic, and discourse analyses to textual data that cannot easily be summarized (Neuendorf 2002; Newbold, Boyd-Barrett, and Van Den Bulck 2002). Because it must necessarily consider multiple interpretational perspectives, the qualitative approach is time consuming and rarely involves large samples of data. It has been also pointed out that in qualitative data analysis, causality cannot be established without high levels of subjectivity (Mehmetoglu and Dann 2003), and qualitative studies have also been criticized as “impossible to do with scientific reliability” (Macnamara 2003, p. 6).

However, the complete separation of the two traditions is not always possible, given the diversity of approaches to content analysis and wide range of its applications. For example, the grounded-theory methodology of content analysis developed by Glaser and Strauss (1967) can be viewed as a “compromise” of inductive and deductive analyses specific to qualitative and quantitative research traditions, respectively (Newbold, Boyd-Barrett, and Van Den Bulck 2002). Modern media scholars such as Hansen et al. (1998), Curran (2002), Gauntlett (2002), and Newbold, Boyd-Barrett, and Van Den Bulck (2002), among others, support analyzing latent as well as manifest content of texts and tend to view qualitative and quantitative content

analysis as complementary parts of a continuum of methods that can be applied to capture the meaning and impact of texts. As Hansen et al. (1998, p. 91) formulated, “content analysis is and should be enriched by the theoretical framework offered by other more qualitative approaches, while bringing to these a methodological rigor, prescriptions for use, and systematicity rarely found in many of the more qualitative approaches.”

There are two main traditions in the quantitative content-analysis research delineated by Weber (1983): substitution model and correlational model. In the substitution tradition, text is analyzed with a priori established categories that are understood as “a group of words with *similar* meaning and/or connotations” (Weber 1983, p. 140). For example, the words *ice*, *snow*, and *igloo* all represent the same idea of cold and, thus, can be united under one category of “cold” (Hogenraad, McKenzie, and Peladeau 2003). Various categories are organized into dictionaries, which are used for making necessary substitutions in the text and for obtaining category frequency counts. These frequencies are organized in a matrix, and associations/correlations between categories can be calculated. Over the history of content analysis, several well-known dictionaries were developed—for example, the *Harvard IV Psychological Dictionaries* (Kelly and Stone 1975) and the *Lasswell Value Dictionary* (Lasswell and Namenwirth 1968). The correlational model, on the opposite, discerns categories from the text analyzed. In this tradition, categories are “groups of words with *different* meaning or connotations that taken together refer to some theme or issue” (Weber 1983, p. 140). These themes are extracted from the matrix of word frequencies by means of factor analysis or other data-reduction technique. The approach was first developed in the late 1950s to early 1960s almost simultaneously in areas of computerized data search, linguistics, and political science (Iker 1974); subsequently, the latent semantic analysis theory provided more theoretical grounding for dimensional reduction of word-frequency data (Landauer and Dumais 1997; Simon and Xenon 2004). One can follow the debate on the theoretical soundness and comparative advantages of substitution and correlational traditions in Weber (1983), Muskens (1985), and Weber (1986).

The place that interpretation takes in quantitative content-analysis methodologies has been considered by Roberts (2000) in his 2x3 taxonomy of quantitative content-analysis approaches. In his classification, one dimension, structural, distinguishes among the ways of obtaining a numerical data matrix of themes or categories frequencies from the text, and the other dimension, interpretational, reflects the perspective from which results

are interpreted. Along the structural dimension of Roberts's taxonomy, there are thematic, semantic, and network text analyses. The thematic approach is rooted in contingency analysis and involves counting themes (categories or key words) belonging to a certain theoretical construct within text blocks. In the semantic text analysis, textual data are separated into specified semantic units—for example, subject-action-object triplets—and every unit is associated with a certain numerical sequence (Franzosi 1997). Last, in the network analysis, text is presented as a network of interrelated themes, and theme linkages are measured by specially generated variables (Salisbury 2001). A quantitative content analysis always produces a two-dimensional data matrix suitable for further statistical analysis.

The interpretational dimension of Roberts's (2000) taxonomy differentiates between the two types of text interpretation: representational and instrumental. "When a researcher understands texts representationally, they are used to identify their sources' intended meanings. When a researcher understands texts instrumentally, they are interpreted in terms of the researcher's theory" (Roberts 2000, p. 262). Roberts illustrated the necessity of this distinction by using an example from Namenwirth and Weber (1987, p. 237), who noted that sometimes, the sources of the texts "are unfamiliar with many fundamental properties of their own culture and . . . to recover culture's properties and rules, we cannot ask culture's participants to answer these questions. Instead, we must rely on outsiders as investigators." Having followed the discussion by George (1959), Pool (1959), Osgood (1959), and Shapiro (1997) regarding representational versus instrumental interpretational dimensions, Roberts (2000) concluded that in many instances, text analysis involves both representational and instrumental perspectives since the researcher can interpret instrumentally the thematic categories that were obtained representationally from the text.

## CATA Software

The large volumes of digital textual data available and the repetitiveness of the task make the computer a natural and powerful choice for content analysis. CATA software is routinely used for storage, search, and retrieval of textual data. It can also assist in theme identification and coding, a time-consuming, prone-to-error "bottleneck" of the content-analysis process (Romano et al. 2003) that is amplified when a large number of cases have to be processed (Macnamara 2003; Wickham and Woods 2005). However, the question of whether automated content analysis is too simplistic and unreliable for

sophisticated interpretation of texts is still under discussion and is inherently connected to the preference for quantitative or qualitative epistemologies for the content-analysis project at hand (Neuendorf 2002; Neuman 1997; Newbold, Boyd-Barrett, and Van Den Bulck 2002). The debate primarily centers on the issue of "manifest versus latent content" (Duriau and Reger 2004; Woodroom 1984), with the concern that computerized measurement of content invariably misses such latent aspects of the text as figures of speech, irony, tone, colloquialisms, and so on (Morris 1994). Yet Duriau and Reger (2004) rightfully noted that human coders also exhibit low reliability for latent content, and besides, the significance of the latent content might be overestimated in certain areas. For example, business texts that appear in corporate documents are mainly written for clarity; thus, analysis can be justifiably restricted to manifest content only. Still, validity is an issue when dealing with metaphors, homonyms, and other aspects of language (Carley 1997; Morris 1994).

Development of CATA software followed the evolution of content analysis and transitory preferences for quantitative and qualitative methodologies in content-analysis research. In the past 20 years, the number of CATA programs has dramatically increased with the revived interest in qualitative methods and with the integration of qualitative and quantitative approaches; several attempts to describe and classify the wealth of CATA software with respect to its functions have been made (Duriau and Reger 2004; Popping 1997; Lowe n.d.; Neuendorf 2002). Alexa and Zuell (2000) in their review of 16 selected CATA software products noted that the demarcation between quantitatively and qualitatively oriented CATA tools is difficult to define due to the absence of a distinct line separating quantitative and qualitative content analyses. Overall, quantitatively oriented programs generally support automatic coding of text using dictionaries, while qualitatively oriented programs better carry out operations associated with theme identification and category coding (Alexa and Zuell 2000; Mehmetoglu and Dann 2003).

Lowe (n.d.) categorizes software for content analysis into three major groups according on its intended function: dictionary-based programs, development environments, and annotation aids. The first group of programs performs content analyses, based on either custom-made or preinstalled dictionaries. *Dictionary*, in this context, is often understood as mapping a set of words (e.g., *large*, *big*, and *huge*) to one word (e.g., *big*); the word *big* in this case is the label for the whole category. Software from this group (Diction, General Inquirer, and VBPro,

among others) often includes such functions as word-frequency analysis, category-frequency counts, cluster analysis, and visualization, which Lowe calls “basic handful.” Content-analysis software from this group is often considered as more quantitatively oriented. The second set of CATA software (DIMAP and Profiler Plus, among others) contains development environments—that is, programs that assist in construction of dictionaries, grammars, and other tools for text analysis. Programs from this group are often oriented toward a specific application area (e.g., linguistics). The third group of programs, annotation aids, is more of “an electronic version of the set of marginal notes, cross-references and notepad jottings that a researcher will generate when analyzing a set of texts by hand” (Lowe n.d., ¶ 1). This group includes such packages as ATLAS-ti, Nudist, and NVivo, among others, and is often considered software for qualitative content analysis.

Studies have noted that the usage of CATA is hindered by the lack of functionality of any given software product in certain areas, but at the same time, the functional possibilities offered by various software programs are not used to their full advantage (Alexa and Zuell 2000; Miles and Weitzman 1994; Zuell and Landmann 2004). The analysis by Alexa and Zuell (2000) concluded that all reviewed CATA products had their strengths and weaknesses and might not support certain operations associated with content analysis in an efficient and user-friendly manner. This opinion was seconded by Romano et al. (2003, p. 216): “Many [CATA] tools are excellent for specific research functions in the social sciences, case studies, and ethnographies; however, they are not designed to deal with analysis of Internet-based Q[ualitative] D[ata] for business purposes.” CATA tools are also subject to biases embedded in design, and some have poor interface and efficiency (Rouse and Dick 1994; Weitzman 1999; Zuell and Landmann 2004). The lack of convenience and efficiency is due to an enormously wide range of possible content-analysis applications, which makes it impractical to create a program that can support all conceivable operations for all types of content analysis. Researchers (Alexa and Zuell 2000; Miles and Weitzman 1996; Tesch 1990) “have argued, convincingly, that no one [CATA] program supports the entire qualitative research life cycle, rather there are categories of software designed to support specific functions within the process” (Romano et al. 2003, p. 216). Alexa and Zuell (2000, p. 318) contended that this lack of support is not a problem “if it were possible to use two or more different software packages for a single text analysis project in a seamless and user-friendly way.” There have already been attempts for integrative use of several

software products; for example, Romano et al. (2003) suggested a computer-assisted methodology to analyze Web-based customer comments, and Wickham and Woods (2005) proposed the integration of different content-analysis software for dissertational research.

## Content Analysis in Destination-Image Research

While there have been some notable examples of using CATA in tourism-related studies and destination-image research, in particular, Mehmetoglu and Dann (2003), noted that, with few exceptions, tourism researchers have been reluctant to rely on CATA for content analysis. This assessment is supported by a record of destination-image studies published from 2000 to 2007, which reported use of content-analysis techniques. Destination-image studies provide a good example because the number of articles in this field have been rapidly increasing since 1970s, and this research area is currently experiencing a growing proportion of studies using qualitative data for image assessment (Echtner and Ritchie 1991; Gallarza, Saura, and Garcia 2002; Pike 2002).

In the meta-analysis of 154 studies on destination-image research published from 2000 to 2007 (Stepchenkova and Mills forthcoming), 53 articles reported use of qualitative approaches at some stage of research: focus groups for questionnaire development, open-ended questions in surveys, collection of textual and pictorial materials from media sources and the Internet, and so on. However, only 6 studies reported use of CATA programs for data analysis. Andsager and Drzewiecka (2002) studied representations of destinations as expressed by college students in a writing form based on the guidebook materials. These researchers used CATA software, although without reporting its name, that identified co-occurrences of key words within the cases, and these key words were cluster analyzed to determine relationships between the most frequent words used to describe destinations. Govers, Go, and Kumar (2007a, 2007b) used CATPAC, a software that is based on the principle of artificial neural network, and its perceptual mapping tool, ThoughtView, to measure destination images of Middle Eastern destinations conveyed as narratives by members of three prominent travel Web sites, with further clustering of the images. Ryan and Cave (2005) studied images of Australian cities and had to analyze large volumes of data obtained through qualitative interviews. These researchers used both CATPAC and TextSmart to analyze the data and construct the perceptual maps of the city images. Two more studies were conducted using the approach outlined in this article; they are given as application examples in the next sections (Stepchenkova & Morrison 2006, 2008).

In some cases, it is not apparent to the reader whether any CATA software was used at all. Images of Lofonten Islands based on qualitative answers to seven survey questions were studied by Jacobsen and Dann (2003). The authors obtained distribution of nouns, adjectives, and verbs for each of the questions and classified them into categories such as nature/landscape, weather/climate, built environment, local people and their activities, and tourists and their activities. The authors did not report whether any CATA software aided their analysis. The study on Brand Singapore by Henderson (2007) conducted interviews with 400 key stakeholders. The data analysis was outsourced to a prominent brand-consultancy firm, and no mention of software used in the analysis was made. Ooi (2004) studied branding of Denmark using a dialogic approach: the researcher had emerged in the field and conducted interviews with officers of different parties in the tourism industry. The author gave a detailed account of how he was set to obtain quality data; however, technical details on how the wealth of the data was analyzed to formulate recommendations for the Brand Denmark were not reported in the article.

One reason for reluctance to use CATA, in the authors' opinion, is the above-discussed lack of software functionality, which precludes using a single specific software product for the entire cycle of the content-analysis project. Thus, the aim of this article is to propose and illustrate a combination of two CATA programs, CATPAC and WORDER, for analysis of textual data emblematic of destination-image research. A number of previous studies employed sorting and categorization techniques to identify the frequencies of certain concepts, words, or people in destination promotional materials and treated the most frequent ones as variables, or dimensions, of the destination-image construct (Andsager and Drzewiecka 2002; Dann 1988, 1996; MacKay and Fesenmaier 1997; Ryan and Cave 2005). These words are often referred to as key words, or image variables. The set of image variables can contain nouns, verbs, and descriptors because nouns are used to focus attention on attractions (e.g., museums), verbs describe actions or tourism types (e.g., rafting), and descriptors create atmosphere (e.g., exciting; Echtner 2002). For instance, if the research aims to find affective images that potential travelers have about a particular destination, the key words to look for would be descriptors like *beautiful*, *friendly*, *ancient*, and so on.

## Method

The methodological approach described in this article is based on a premise that for every question of interest,

there is a universe of textual data (text population) from which a representative sample can be drawn. Applying an appropriate methodology, a researcher is able to get a response from any textual unit in the sample on every variable of interest and measure it to obtain a numerical matrix of frequencies for further statistical analysis (Neuendorf 2002; Roberts 2000). The approach can handle a large number of textual files at once: the software selected is able to process simultaneously multiple textual files of similar type providing separate per-unit output from a single run. The approach aids in the identification of variables of interest (key words), which are grounded in the data (Glaser and Strauss 1967), and provides a convenient interface for specifying these key words and their variants for counting. It assists in "data smoothing," a tedious, error-prone task that often precedes the actual content analysis. Combination of two software programs, CATPAC (Woelfel 1998) and WORDER (Kirilenko 2007), efficiently addresses these issues.

CATPAC has been long employed in social sciences for analysis of political speeches, focus-group interviews, and tourism-related research. The program produces a variety of outputs: word counts, frequency rankings, cluster diagrams, and interactive neural cluster analysis. Its add-on function ThoughtView can generate two- and three-dimensional concept maps based on the results of the analyses. CATPAC features are comparatively tabulated against other software programs in Neuendorf (2002) and described by Lowe (n.d.). The program has a clustering function; however, one has difficulties processing files of substantial size with CATPAC (Woelfel 1998, p. 25). CATPAC allows processing just one file at a time and does not assist with data smoothing. The proposed approach uses one function of CATPAC, namely, word-frequency ranking, which is the basis for determining the variables of interest, or key words, for subsequent counting in multiple textual files.

WORDER was developed to automate counting of specified words and their variants in a large number of textual files. During one run, WORDER is capable of parsing up to 1,000 textual files looking for up to 1,000 words and counting their occurrences in every data file. These words (key words) are specified for counting by means of an input table (or a dictionary) constructed by the researcher. The other input for WORDER is a list of names of all data files that need to be processed. The result of WORDER analysis is a numerical matrix of key-word frequencies, which can be easily transported to general statistical analysis packages like SPSS. If the original data are obtained as responses to open-ended questions collected in online surveys and saved in a spreadsheet-type file, a convenient feature of WORDER can split these responses into separate files.

**Table 1**  
**WORDER Input Table**

monastery	cloister	convent	abbey	
redsquare	Red Square			
museum	museums			
unsafe	not safe	not feel safe	afraid	risky

Data smoothing should be performed prior to any computer-assisted content analysis; however, the necessary changes should concern only the key words (Schmidt 2001). Such issues as misspellings, synonyms, multiword concepts, and singular or plural forms of key words can be dealt using WORDER in the same manner. Key words and their variants are placed in the input table (dictionary) row by row (see Table 1). Guided by the dictionary, WORDER replaces variants with the corresponding key word and counts the number of times the key word occurs in the textual file. However, if the key words contain possible homographs, the only way to determine the intended meaning is to scan the original data for all occurrences of the word (Insch and Moore 1997). Negatives are also difficult to deal with, because the negation and the actual word can be separated by a large number of other words (e.g., “I don’t think that I would feel safe there”). One way to deal with negative concepts would be “to reverse negative statements to commonality with positives” (Ryan and Cave 2005, p. 146). For a detailed technical description of WORDER, visit the program’s Web site ([www.kirilenko.org/worder](http://www.kirilenko.org/worder)).

When key words are known a priori, WORDER can be used on its own (see Example 1). When key words have yet to be identified, a combination of CATPAC and WORDER is proposed. In broad strokes, the method works as follows. Textual data are pooled together and analyzed by CATPAC; words that do not add to meaning are placed in the Exclude file (option provided by CATPAC) and ignored. The output of the CATPAC procedure is the list of words ranked in order of their frequencies. Guided by theory, the researcher identifies the most relevant key words and uses them to construct the dictionary for WORDER. For example, if key words contain the highly frequent word *church*, the researcher might wish that such words as *monastery* and *abbey* also be counted under the key word *church*. The convenient, though not defining feature of the approach is its iterative character. WORDER can be used at the stage of key-word identification to provide, if necessary, desirable substitutions in the data, so key words similar in meaning would be mapped to a chosen variable. This iterative feature allows words with lesser frequencies come forward from the data and, thus, refine the dictionary. Using the custom-made dictionary, WORDER counts every occurrence of every key word in every textual file. The

matrix of key-word frequencies obtained by WORDER is a subject of further statistical analysis. Both CATPAC and WORDER complement each other and, when used together, broaden the possibilities for researchers working with textual data. Ultimately, the approach allows (1) data preparation, (2) identification of key words in textual data using CATPAC, (3) convenient construction of dictionaries, (4) counting the occurrences of key words in every textual file with WORDER and obtaining a matrix of key-word frequencies, and (5) subsequent dimensional reduction of key-word-frequency data.

The proposed approach is particularly suitable if, first, the nature of the research allows one “to intuit a gestalt image of repetitive concepts” from reading the text (Ryan and Cave 2005, p. 146), meaning that while in the course of the research, key words might require interpretation beyond their primary meaning (e.g., from the “favorable-unfavorable” perspective), at the stage of image-variable selection, the meaning of the key words can be taken at face value. And, second, the proposed approach is particularly suitable when a large number of textual files of a similar type have to be analyzed. It should be noted, however, that the researchers do not claim that other CATA programs cannot do what the proposed software tandem does. The frequency-ranking feature, which aids in identification of key words, is included in many CATA programs, even if the feature of processing multiple files simultaneously is met much less often. The purpose is to show how the proposed approach can enrich the analysis of textual data in terms of transparency, speed, and statistical possibilities. As was noted by Neuendorf (2002) and Hogenraad, McKenzie, and Peladeau (2003), the nature of the text, task at hand, and researcher’s preferences and expertise should dictate the choice of software product. Moreover, no CATA programs or their combination can provide completely automated content analysis. Human contribution is a crucial factor; that is, contextualization and interpretation of the key concepts are still in the hands of the investigator (Neuendorf 2002; Mehmetoglu and Dann 2003).

## Application Examples

The application examples demonstrate the proposed approach in three destination-image studies. Destination-image research has important marketing implications for destination competitiveness, management, and development (Ritchie and Crouch 2003) and has made substantial theoretical advancements since 1970s, when it first was proposed by Hunt (1971). Analyses of destination images provide important insights into consumer travel

behavior (Chon 1991); thus, several aspects of the destination-image construct (i.e., cognitive, affective, and behavioral; Gartner 1993) were identified to study its influence on travel choices. While the attribute-based aspects of destination image have received ample attention in empirical research, images of a more holistic nature have not been sufficiently studied (Pike 2002). Example 1 illustrates how to use the approach when key words are known a priori. Example 2 shows how key words are identified and counted by using the CATPAC-WORDER approach and then clustered into more holistic concepts by means of factor analysis in a general purpose statistical package like SPSS. Example 3 illustrates the applicability of the approach when a large number of textual files need to be processed and a stronger interpretational aspect is present in the research.

### **Example 1: Induced Images—Comparative Travel Offer**

The first study where WORDER was used dealt with the induced image of Russia and popularity of Russia's recreational regions among U.S. and Russian tour operators who target English-speaking travelers (Stepchenkova and Morrison 2006). One of the research objectives was to identify destinations within Russia mentioned most frequently by U.S. and Russian tour operators. Russia-related texts in English from 79 U.S. and 84 Russian tour-operator Web sites were collected. The U.S. sample included the Web sites of the members of United States Tour Operators Association, American Society of Travel Agents, cruise lines, Internet travel guides, and "independents." Web sites in the Russian sample were obtained through search engines, as well as from the Web sites of such organizations as the *Moscow Times*, Moscow International Travel and Tourism Exhibition, official Moscow city Web site, Yellow pages, and so on. Collected textual files were regarded as a representative sample from a population of all U.S. and Russian Web sites promoting trips to Russia.

Independently, a master list of 344 tourist destinations within Russia was compiled based on the standard division of the country into 13 recreational regions (Goscomstat 2000). To generate the master list, the researchers used the reports of the Russian Federal State Statistics Service (Goscomstat), the official Web site of the Federal Tourism Agency of the Russian Federation, Web sites of administrative entities of the Russian Federation, and UNESCO, in addition to their own expertise. Since a great number of destination names were spelled differently in English (e.g., Saint-Petersburg and Sankt-Peterburg) or had equally used

names (e.g., Zagorsk and Sergiyev Posad), a dictionary for WORDER included alternative spellings of all destinations from the master list.

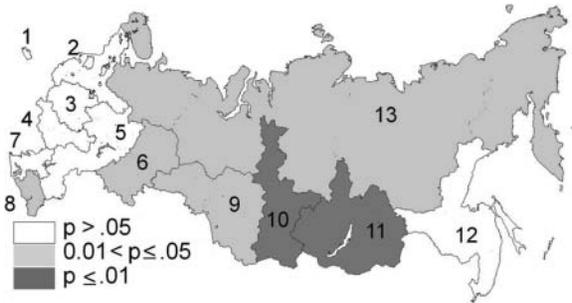
Destinations specified in the dictionary were counted in every file using WORDER. It was natural to assume that not all the destinations that had been mentioned in the sampled Web sites were actually included into the master list for counting. To estimate the richness of the destination pool and to see how many destinations might have "slipped through" undetected, the Chao and jackknife statistical methods of extrapolated richness were used (Chao 1987; Oksanen n.d.). Chao's extrapolation of destination richness gave 331 destinations; first- and second-order jackknife methods gave 346 and 357 destinations, respectively. Overall, the specified pool of 344 destinations was considered to be a very good scope.

To get a broader picture of the U.S. and Russian travel offers, destinations belonging to the same recreational region were aggregated by adding up the frequencies of the original destinations; thus, 13 integrated regional variables were computed for both samples. It was noted from the data that tour operators sometimes offered trips to Russia as a part of multicountry tours. Therefore, one more regional variable, namely, "foreign," was computed. *T*-tests were conducted to compare frequencies of the regional integrated variables for the U.S. and Russian samples: significant differences were registered for regions representing Northern, Siberian, Urals, and Caucasus parts of Russia, with Russian Web sites promoting these regions more heavily. However, destinations that belonged to the foreign group and offered trips to Russia as a part of joint tours to Finland, China, or Baltic states were more actively promoted by U.S. operators. The visual representation of the *t*-test results is given in Figure 1. For the entire analysis, complete results, and their detailed discussion, refer to the original article.

### **Example 2: Holistic Images—Extending Echtner and Ritchie (1993)**

The proposed methodology was used in a study of Russia's holistic destination images (Stepchenkova and Morrison 2008). The overall research design followed the conceptual framework suggested by Echtner and Ritchie (1993); however, content analysis of responses to open-ended questions was conducted using the CATPAC-WORDER approach. To get insights into Russia's stereotypical holistic images, 317 textual responses to the question, "What images or characteristics come to mind when you think of Russia as a travel destination?" were analyzed. A list of 72 most frequent key words (all

**Figure 1**  
Regional Variables: T-Test Comparisons



Note: 1 = Western; 2 = Northwestern; 3 = Central; 4 = Southern; 5 = Povolzhje; 6 = Urals; 7 = Azov-Black Sea; 8 = Caucasus; 9 = Ob'-Altai; 10 = Yeniseisky; 11 = Baikal; 12 = Far Eastern; 13 = Northern.

frequencies are 5 or higher) was generated using CAT-PAC. Some words (e.g., *history/historic/historical* or *large/big*) were grouped together for counting under the most frequent name, in these cases, *history* and *large*. This procedure resulted in the list of 45 key words that were regarded as Russia's stereotypical image variables. The frequencies of these image variables were counted in every textual response using WORDER. Table 2 contains the overall frequencies of Russia's stereotypical image variables.

The next step was to reduce the number of stereotypical image variables to a smaller number of image concepts by means of principal components analysis with varimax rotation. The data matrix, generated by WORDER, had 317 cases and 45 variables, which gave a solid case to variable ratio of 7.04 (Kline 1994). The correlation matrix was found to be factorable with highly significant result on the Bartlett's test and with the Kaiser-Meyer-Olkin statistic of sampling adequacy of .529. It was decided to look for stable word combinations, which might include as few as two words, since textual responses to survey questions were generally short (e.g., "Cold. Beautiful churches."). Therefore, the number of factors was not specified, and the option "eigenvalues larger than 1" was chosen. Three weak items with low coefficients in the diagonal of the anti-image matrix ( $< .40$ ), low communalities ( $< .50$ ), and those that did not load higher than .35 on any factor were eliminated (Kline 1994). Without them, the factorability of the matrix increased up to .537, and the remaining 41 variables produced 17 factors that explained 67% of the total variance. The results of the rotated solution are given in Table 3.

Guided by the produced solution, the factors were checked against the original data to ensure that word

combinations containing descriptive items, such as *cold*, *beautiful*, *poor*, *old*, *large*, *great*, *vast*, *friendly*, and *different*, were not used in a negative context, which would affect factor interpretation. Factor 16, "great food," was eliminated as the result of this check as well as low reliability alpha. Another concern was that some stable word combinations produced by factor analysis did not account for large differences in frequencies. In Factor 9, the frequencies of words *old* and *buildings* were 25 and 39, respectively. It meant that at least 14 occurrences of the word *buildings* were used in combinations with words other than *old*. Therefore, factors that contained words with large differences in frequencies were checked against the original data as well. For example, it was discovered that the word *poor* was also used by respondents with concepts *lodgings*, *hotels*, and *accommodations*, which were not included into the image-variable set produced by CATPAC. Finally, some image factors were combined together since they described the same image concepts. For example, Factors 4 and 8 made one holistic image of "orthodox churches with onion-shaped domes." The final results of Russia's stereotypical holistic images are given in Table 4.

### Example 3: Images of China and Russia in the U.S. General Media

One of the objectives of the study by Stepchenkova, Chen, and Morrison (2007) was to comparatively examine images of China and Russia as they were projected by U.S. newspaper sources. Destination image is inherently connected to the country image (Mossberg and Kleppe 2005), and organic data sources like general newspapers are often seen as objective sources of image formation (Gartner 1993). Image issues related to political, epidemiological, and criminal situation, which are likely to be covered by general newspapers, are important factors in making destination choices (Crompton 1977). Both China and Russia are large countries with distinctive cultures, abundant heritage resources, beautiful scenery, and art that shared a communist history and planned economies. Both countries actively promote cultural tourism, luxury cruises, and transcontinental railroad travel from Moscow to Beijing and, from the perspective of the U.S. pleasure traveler, can be considered long-haul destinations. It was reasoned that due to these similarities, China and Russia should share, at least to some degree, the same potential target audiences and compete for them in the U.S. pleasure-travel market to the Asian region.

The study selected general media articles about China ( $n = 540$ ) and Russia ( $n = 540$ ) from the LexisNexis

**Table 2**  
**Russia's Stereotypical Image Variables**

Key Word	Freq.	Key Word	Freq.	Key Word	Freq.	Key Word	Freq.
Cold	69	Kremlin	24	Food	12	Orthodox	7
Beautiful	55	Palaces	23	Culture	12	Open	7
People	54	Weather	19	Friendly	12	Vodka	6
History	45	Museums	19	Domes	10	Exotic	6
Buildings	39	Churches	19	Countryside	10	Sites	6
Poor	38	Cities	18	Snow	9	Volga	5
Architecture	37	Large	15	Hermitage	9	River	5
Red Square	36	Interesting	13	Music	9	Spaces	5
Saint-Petersburg	34	Onion*	13	Winter	9	Ballet	5
Moscow	30	Art	13	Dark	8		
Country	28	Great	12	Different	8		
Old	25	Vast	12	Places	7		

\* Onion-shaped domes are an architectural feature of great many Russian churches.

**Table 3**  
**Russia's Stereotypical Images: Factor Analysis**

Factor	Key Word	Loading	Alpha, Variance	Factor	Key Word	Loading	Alpha, Variance
1	Hermitage	.785	0.552	10	People	.825	0.415
	Palaces	.687	4.961		Friendly	.692	3.688
	Museums	.633					
2	River	.930	0.887	11	Sites	.830	0.418
	Volga	.922	4.780		History	.805	3.630
3	Open	.794	0.586	12	Weather	.788	0.468
	Spaces	.742	4.776		Cold	.782	3.537
	Snow	.535					
4	Onion	.884	0.814	13	Moscow	.758	0.500
	Domes	.873	4.602		Petersburg	.536	3.521
5	Cities	.680	0.445	14	Country	.801	0.424
	Large	.663	4.443		Vast	.556	3.483
	Places	.645					
	Poor	.416					
6	Ballet	.804	0.590	15	Culture	.798	0.516
	Music	.775	3.944		Different	.795	3.480
7	Beautiful	.714	0.399	16	Food	.727	0.239
	Countryside	.639	3.876		Great	.576	3.397
	Architecture	.565					
8	Churches	.805	0.512	17	Kremlin	.706	0.321
	Orthodox	.740	3.803		Red Square	.596	3.394
					Art	.334	
9	Buildings	.793	0.494	18	Vodka		
	Old	.718	3.764				

**Table 4**  
**Russia's Stereotypical Holistic Images**

No.	Stereotypical Holistic Images	No.	Stereotypical Holistic Images
1	Cold weather, snow	9	Orthodox churches with onion-shaped domes
2	Beautiful architecture and old buildings	10	Big cities, interesting old cities
3	Poor people, country, lodgings, and food choices	11	Great culture, different culture
4	Historic sites and places	12	Beautiful music, ballet, art
5	Moscow, Red Square, and Kremlin	13	Friendly/unfriendly people
6	St. Petersburg, Hermitage, palaces, and museums	14	Volga River
7	Vast country with lots of open spaces	15	Vodka
8	Beautiful countryside		

**Table 5**  
**China and Russia 2002 to 2004: Image Themes**

China				Russia			
No.	Factor	Variance Explained	Cronbach's Alpha	No.	Factor	Variance Explained	Cronbach's Alpha
1	Economic Growth	5.91	.587	1	YuKOS	4.94	.810
2	Industry	4.72	.726	2	Iraq	4.73	.776
3	World Trade Organization	4.46	.712	3	Presidential Elections	4.11	.747
4	Global Market	4.05	.623	4	Law	3.85	.595
5	Taiwan	3.97	.726	5	Natural Monopolies	3.69	.690
6	Technology Transfer	3.94	.745	6	Chechnya	3.55	.667
7	Government	3.54	.784	7	Soviet Past	3.52	.720
8	Labor Market	3.18	.672	8	Iran	3.29	.730
9	Cultural Communications	3.16	.636	9	Adoption of Russian Children	3.20	.673
10	SARS	3.14	.699	10	NATO	2.84	.512
11	Human Rights	3.12	.940	11	Russia-China Relations	2.82	.593
12	Communist China	2.94	.583	12	Power Sector Reform	2.78	.635
13	Asian Politics	2.91	.423	13	Sports	2.44	.489
14	Security Concerns	2.48	.663	14	U.S.-Russia Space Cooperation	2.32	.725
15	Educational Exchange	2.15	.779	15	Beslan	1.87	.678

database for the period of 2002 to 2004 and treated them as Chinese and Russian samples of textual data. Articles for any given month that had the words *China/Chinese* or *Russia/Russian* in the headlines were identified, alternating four groups of U.S. newspaper sources: Northeast, Southeast, Western, and Midwest. For each of the two samples, 15 articles from each month were selected using a systematic random-sampling algorithm with a randomized base number (Lohr 1998) to distribute the chosen articles evenly through time. The entire body of selected *Chinese* or *Russian* texts was first analyzed with CATPAC to identify up to 100 of the most frequent key words used with reference to each country. These key words were regarded as the country's image variables in the U.S. general media. Second, WORDER counted the identified variables in every textual file in each of the

two samples. Factor analysis was further employed to identify the main image themes with regard to China and Russia as conveyed by the U.S. press (see Table 5).

Principal components analysis with direct oblimin rotation was used to identify 15 factors for each country. The oblique rotation was preferred over the orthogonal one since the oblique rotation allows factors to covary (Kline 1994), which is important in text analysis where the same word can be used in multiple language contexts. With a number of words loaded highly on several factors, the direct oblimin rotation produces the most simple and interpretable factor structure (Kline 1994). The number of factors (15) was chosen on the basis of the desirable variable to factor ratio and to make the factors more global. Weak variables were eliminated, and the retained variables (China = 83, Russia = 70) produced an

interpretable solution that explained 57.2% and 58.1% of variance in the original matrices.

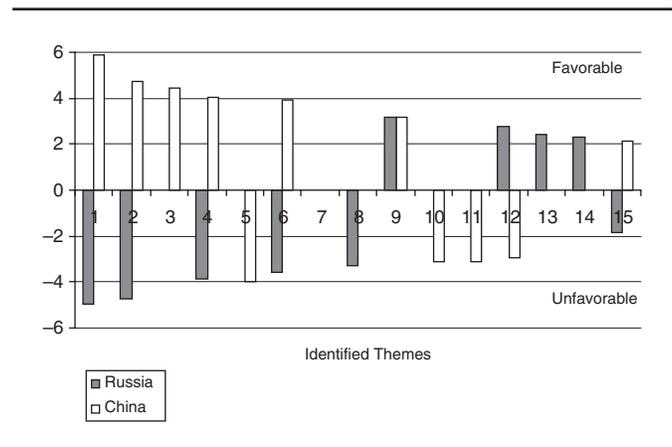
Out of the 15 factors identified for China, 6 related to economic issues: Economic Growth, Industry, World Trade Organization, Global Market, Technology Transfer, and Labor Market. Cultural Communications and Educational Exchange were the 2 factors that reflected the social aspect of China's image. The aspect related to the country's internal affairs was captured by the Government and Taiwan factors. The Asian Politics and Security Concerns factors highlighted China's influence on the political stability in the Asian region and showed the U.S. concern for security in Asia. Human Rights, Communist China, and SARS were the remaining 3 factors.

Out of the 15 Russian factors, 4 related to the role that Russia plays in foreign politics with regard to Iraq, Iran, NATO, and Russia-China Relations. The Soviet Past factor reflected the influence that Russia has on the world's political arena as a successor of the Soviet Union. Economic issues were represented by 3 factors: YuKOS, Natural Monopolies, and Power Sector Reform. The Law factor emphasized the selective application of Russian law, and the Chechnya factor encompassed themes related to the war in Chechnya. Russian internal affairs were represented by the Presidential Elections factor. The remaining factors were the U.S.-Russia Space Cooperation, Sports, Adoption of Russian Children, and Beslan (referred to the act of terrorist activity in September 2004 in the city of Beslan).

Each factor was assigned one of the three "favorability" values, based on whether the coverage of the topic in the U.S. media was predominantly negative (-1), neutral (0), or positive (+1). The interpretation of favorability was done by the researchers and involved two main considerations: (1) the attitude toward the issue in the U.S. media and (2) the issue itself. For example, the Iraq factor for the Russian sample was interpreted as negative because the coverage of Russia's position on the war in Iraq was mostly critical in the U.S. media. However, the Beslan tragedy was interpreted as negative despite the very compassionate coverage in the U.S. media, as the tourism industry is generally hurt by terrorist activity.

Thus, China was given positive ratings for such factors as Economic Growth, Industry, World Trade Organization, Global Market, Technology Transfer, Cultural Communications, and Educational Exchange. Zero ratings were received for the topics related to China's Government, Labor Market, Asian Politics, and Security Concerns. Negative ratings were assigned to the factors of Taiwan, SARS, Human Rights, and Communist China. In the Russian sample, the positive

**Figure 2**  
**Images of China and Russia: Favorability Comparisons**



issues were Russian Children in the United States, Power Sector Reform, Sports, and U.S.-Russia Space Cooperation. Zero favorability ratings were assigned to such factors as Presidential Elections, Natural Monopolies, Soviet Past, NATO, and Russia-China Relations. Negative ratings were received by the image themes of YuKOS, Iraq, Chechnya, Iran, Law, and Beslan. The importance of the factors was understood as the amount of variance that they explained in the factor solution. The results of favorability comparison are displayed in Figure 2. As can be seen, the most important image themes for China in the U.S. press were generally positive, while for Russia, they were mostly negative.

## Theoretical Considerations

Neuendorf (2002), as well as Zuell and Landmann (2004), noted that most software programs are "black boxes," meaning that they do not reveal the details of their measurements, which reduces replicability of content analysis not only with different CATA programs but also with the same program. In contrast, the CATPAC-WORDER approach is extremely transparent and verifiable. Identification of key words, aided by CATPAC, is based on their compounded frequencies in the pooled data. Then, WORDER counts identified key words in every textual file, and aggregation of key words into latent constructs of a more holistic nature can be done using general statistical packages where all calculations are replicable. Examples 2 and 3 demonstrate usage of the correlational model of content analysis when categories (or image themes) were uncovered by means of

factor analysis (Iker 1974), while Example 1 lies within the substitution tradition where decisions on the categories (destinations in that case) are made independently of the actual data. Finally, the interpretation of latent variables, or themes, can be done using qualitative methods, thus enriching the analysis. Interestingly, while writing this article, the authors came across an article by Simon and Xenon (2004), which advocated the use of a similar, though not identical, methodological approach for analysis of political texts; their approach was grounded in the correlational model of quantitative content analysis and the theory of latent semantic analysis (Landauer and Dumais 1997) that supports data-driven category selection.

One important aspect of the scientific method in content analysis is a priori design. Neuendorf (2002, p. 11) strongly argued for a deductive approach where “all decisions on variables, their measurement, and coding rules must be made before the observation begins.” A primarily deductive content analysis aided by WORDER is illustrated in Example 1. Bias is more easily introduced into research when key words are defined from the text—that is, when the inductive approach to variable identification is used (Macnamara 2003). As Wickham and Woods (2005) noted, in inductive qualitative content analysis, new categories are sometimes added without a logical argument for their creation. However, when an inductive approach is used, researchers study concepts and categories that are grounded in the data rather than those imposed from the outside (Glaser and Strauss 1967; Mehmetoglu and Dann 2003; Zuell and Landmann 2004). Nevertheless, at some point prior to actual counting, the pool of the variables should be firmly defined. In the authors’ view, the choice between deductive and inductive approaches depends on the type of study and research question and often is a balance between the two.

To be able to generalize the results in quantitative content analysis, the selection of textual units should be representative of the textual population as a whole. “To be quantitative, a text analysis must both address a social scientific question of a well-defined text population, and provide an answer to the question having a known probability of inaccurately reflecting aspects of the text population” (Roberts 2000, p. 260). While the simplest approach for selecting texts for analysis would be a census since it provides the greatest possible representation, this is impractical in most cases. Generalizability of content-analysis results is inherently connected to the research design in general, especially to identification of the textual population and selection of a representative sample. For instance, in Example 3, a combination of stratified and systematic random

sampling of the database articles was employed to assure representativeness of the sample and generalizability of results, while in Example 2, the validity of Russia’s stereotypical images among American pleasure travelers depended on representativeness of the obtained sample of survey respondents.

Within the 2x3 taxonomy for quantitative data analyses proposed by Roberts (2000), the CATPAC-WORDER approach is structurally thematic since it counts the occurrences of specified key words in sampled text blocks, and numerical associations between key words can be calculated. From the interpretability standpoint, the approach is both representational and instrumental. Computer software takes word meaning at face value; however, selection of key words, as well as the interpretational aspect of the results, is always motivated by some theoretical perspective (in our examples, the destination-image construct), and this perspective should be made explicit from the start to evaluate the validity of the conclusions (Roberts 2000). In Example 1, the approach was purely representative since WORDER counted destinations within Russia, and no interpretation as to the “meaning” of these destinations was necessary. In Examples 2 and 3, the choice of key words required more interpretation on the part of the researcher, and the instrumental aspect of the analysis was even stronger as researchers had to make a decision how to interpret the identified factors, or themes. Overall, the proposed approach is flexible enough to validate the factors, or latent variables, obtained quantitatively using qualitative methods to substantiate the area of focus (Gray and Densten 1998; Patton 1990), as was demonstrated in Examples 2 and 3.

The numerical data matrices obtained from the textual data by using the CATPAC-WORDER approach can be enhanced by contextual variables (e.g., data source, type, audience, time, etc.). The Russia-U.S. and Russia-China distinctions in Examples 1 and 3, respectively, are instances of such contextual variables. Moreover, the data can be “embellished with secondary variables that measure the source’s positive or negative sentiment regarding each theme” (Roberts 2000, p. 263). The psychological concept of “attitude” brings to content analysis an evaluative dimension, such as “favorable-unfavorable” (Krippendorff 1980; Insch and Moore 1997), as was demonstrated in Example 3. To provide such an enhancement, the proposed approach can be combined with more traditional research techniques—for example, assigning a favorability index to key words, or themes, based on experts’ qualitative evaluations (Example 3). The favorability analysis of affective destination images of Russia with further hypothesis testing was conducted by

Stepchenkova and Morrison (2008) using the CATPAC-WORDER method but is not given here because of space constraints.

## Conclusion

Content analysis is widely used in many disciplines to analyze various forms of communications, above all, those that utilize textual data. A growing number of tourism studies employ qualitative data (interviews, open-ended questions, promotional brochures, Web-based content, etc.) and, subsequently, content-analysis techniques to discern meaning from this wealth of textual material. In the past decades, many software packages, both quantitatively and qualitatively oriented, have been developed to assist researchers in content-analysis studies; nevertheless, CATA software is not often used in tourism research despite its obvious benefits. In the authors' view, there are two main reasons for this lack of enthusiasm: first, a quantitative paradigm of content analysis is still sometimes viewed as simple word counting, which is not conducive to uncovering deeper themes and issues of texts; and second, no single software package is currently able to provide the full spectrum of functions that might be needed in various content-analysis projects. Thus, this article offers a transparent and effective approach to facilitate content analysis of textual data typical in tourism-related studies; it also demonstrates that the proposed approach is firmly grounded in the theory and practices of content-analysis methodology.

The approach assists in data preparation and identification of key words by employing concepts grounded in the data. It can handle both theory-driven variables defined a priori (Example 1) and data-driven key words (Examples 2 and 3). The approach aids in obtaining a key-word-frequency matrix and allows the clustering of identified key words into more holistic themes (Examples 2 and 3). It facilitates statistical comparisons (Example 1) and, ultimately, hypothesis testing. The proposed approach combines two software products, CATPAC and WORDER, and is particularly useful when a large number of textual files representing a certain text population needs to be analyzed. The approach is not dependent on this particular software tandem; other CATA software products that can handle the same operations can also be used. While the approach was initially intended to broaden the scope of destination-image analysis into theoretically developed but not adequately researched aspects of the destination-image construct, such as holistic, organic, and induced images, its application scope can be extended into other areas as well. The approach is easy and efficient enough to be used by

managers and practitioners to, say, quickly examine customer complaints submitted online or to get insights into issues of interest discussed in virtual communities that their companies maintain.

As with any methodology, the appropriateness of the CATPAC-WORDER approach for a particular research question has to be assessed prior to using it, in order to avoid a mechanistic counting of words poorly related to a research topic. Three limitations of the proposed approach should be noted. First, its applicability depends on the type of variables of interest: key words should be conducive to identification by CATA software; that is, their meaning should be taken at face value. As was demonstrated in the application examples, studies on destination image often fall into this class of research. Second, the approach suggests a combination of two CATA products, CATPAC and WORDER, which possibly requires a steeper learning curve than where only one software program is used. However, as was already mentioned, researchers can use CATA software that they are familiar with, provided that it has the same analytical functions. Third, to the authors' knowledge, the approach has not been tested beyond the research cited in this article and a study by Chen and Lehto (2007), who analyzed brand images of spa resorts using textual content from spa resorts' Web sites. While several tests were successfully conducted by the authors to verify the functionality and accuracy of WORDER, more testing is desirable.

## References

- Alexa, M., and C. Zuell (2000). "Text Analysis Software: Commonalities, Differences and Limitations: The Results of a Review." *Quality & Quantity*, 34: 299-321.
- Andsager, J. L., and J. A. Drzewiecka (2002). "Desirability of Differences in Destinations." *Annals of Tourism Research*, 29 (2): 401-21.
- Berelson, B. (1952). *Content Analysis in Communication Research*. Glencoe, IL: Free Press.
- Berg, B. L. (1995). *Qualitative Research Methods for the Social Sciences*. Boston: Allyn & Bacon.
- Carley, K. M. (1997). "Extracting Team Mental Models through Textual Analysis." *Journal of Organizational Behavior*, 18: 533-59.
- Chao, A. (1987). "Estimating the Population Size for Capture-Recapture Data with Unequal Catchability." *Biometrics*, 43: 783-91.
- Chen, Y., and X. Y. Lehto (2007). "Brand Personality of Spa Destinations (Resorts) on the Web." In Proceedings of the 12th Annual Graduate Education and Graduate Students Research Conference on Hospitality and Tourism, Houston, Texas, pp. 911-19.
- Chon, K. (1991). "Tourism Destination Image: Marketing Implications." *Tourism Management*, 12: 68-72.
- Crompton, J. L. (1977). "A System Model of the Tourist's Destination Selection Process with Particular Reference to the Role of Image and Perceived Constraints." PhD dissertation, Texas A&M University, College Station.
- Curran, J. (2002). *Media and Power*. London: Routledge.

- Dann, G. (1988). "Images of Cyprus Projected by Tour Operators." *Problems of Tourism*, 11 (3): 43-70.
- (1996). "Tourists' Images of a Destination: An Alternative Analysis." *Journal of Travel and Tourism Marketing*, 5 (1-2): 41-55.
- Denzin, N. K., and Y. S. Lincoln (1994). "Introduction: Entering the Field of Qualitative Research." In *Handbook of Qualitative Research*, edited by N. K. Denzin and Y. S. Lincoln. Thousand Oaks, CA: Sage, pp. 1-17.
- Duriau, V. J., and R. K. Reger (2004). "Choice of Text Analysis Software in Organization Research: Insight from a Multi-Dimensional Scaling (MDS) Analysis." Paper presented at the Seventh International Conference on the Textual Data Statistical Analysis, Louvain-la-Neuve, Belgium.
- Echtner, C. M. (2002). "The Content of Third World Tourism Marketing: A 4A Approach." *International Journal of Tourism Research*, 4: 413-34.
- Echtner, C. M., and J. R. B. Ritchie (1991). "The Meaning and Measurement of Destination Image." *Journal of Tourism Studies*, 2 (2): 2-12.
- (1993). "The Measurement of Destination Image: An Empirical Assessment." *Journal of Travel Research*, 31 (4): 3-13.
- Franzosi, R. (1997). "Labor Unrest in the Italian Service Sector: An Application of Semantic Grammar." In *Text Analysis for the Social Sciences: Methods for Drawing Statistical Inferences from Texts and Transcripts*, edited by C. W. Roberts. Mahwah, NJ: Lawrence Erlbaum, pp. 79-100.
- Gallarza, M. G., I. G. Saura, and H. C. Garcia (2002). "Destination Image: Towards a Conceptual Framework." *Annals of Tourism Research*, 29 (1): 56-78.
- Gartner, W. C. (1993). "Image Formation Process." *Journal of Travel & Tourism Marketing*, 2 (3): 191-216.
- Gauntlett, D. (2002). *Media, Gender & Identity*. London: Routledge.
- George, A. (1959). "Quantitative and Qualitative Approaches to Content Analysis." In *Sociological Methodology*, edited by A. Raftery. Oxford: Basil Blackwell, pp. 135-44.
- Glaser, B. G., and A. L. Strauss (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine.
- Goscomstat (2000). *Tourism in Figures. Annual Report* [in Russian]. Moscow, Russia: The Russian Federal State Statistics Service.
- Govers, R., F. Go, and K. Kumar (2007a). "Virtual Destination Image a New Measurement Approach." *Annals of Tourism Research*, 34 (4): 977-97.
- (2007b). "Promoting Tourism Destination Image." *Journal of Travel Research*, 46 (1): 15-23.
- Gray, J. H., and I. L. Densten (1998). "Integrating Quantitative and Qualitative Analysis using Latent and Manifest Variables." *Quality & Quantity*, 32: 419-31.
- Hansen, A., S. Cottle, R. Negrine, and C. Newbold (1998). *Mass Communication Research Methods*. London: Macmillan.
- Henderson, J. C. (2007). "Uniquely Singapore? A Case Study in Destination Branding." *Journal of Vacation Marketing*, 13 (3): 261-74.
- Hogenraad, R., D. P. McKenzie, and N. Peladeau (2003). "Force and Influence in Content Analysis: The Production of New Social Knowledge." *Quality & Quantity*, 37: 221-38.
- Holsti, O. R. (1969). *Content Analysis for the Social Sciences and Humanities*. Reading, MA: Addison-Wesley.
- Hunt, J. D. (1971). "Image: A Factor in Tourism." PhD dissertation, Colorado State University, Fort Collins.
- Iker, H. P. (1974). "An Historical Note on the Use of Word-Frequency Contingencies in Content Analysis." *Computers and the Humanities*, 8: 93-98.
- Insch, G. S., and J. E. Moore (1997). "Content Analysis in Leadership Research: Examples, Procedures, and Suggestions for Future Use." *Leadership Quarterly*, 8 (1): 1-25.
- Jacobsen, J. K. S., and M. S. G. Dann (2003). "Images of the Lofoten Islands." *Scandinavian Journal of Hospitality and Tourism*, 3 (1): 25-47.
- Kelly, E. F., and P. J. Stone (1975). *Computer Recognition of English Word Senses*. Amsterdam: North-Holland.
- Kirilenko, A. P. (2007). WORDER 2.2. [Computer software]. <http://kirilenko.org/worder>.
- Kline, P. (1994). *An Easy Guide to Factor Analysis*. New York: Routledge.
- Krippendorff, K. (1980). *Content Analysis: An Introduction to Its Methodology*. Newbury Park, CA: Sage.
- Landauer, T. K., and S. Dumais (1997). "A Solution to Plato's Problem: The Latent Semantic Analysis Theory of the Acquisition, Induction, and Representation of Knowledge." *Psychological Review*, 104: 211-40.
- Lasswell, H. D., D. Lerner, and I. de S. Pool (1952). *The Comparative Study of Symbols*. Stanford, CA: Stanford University Press.
- Lasswell, H. D., and J. Z. Namenwirth (1968). *The Lasswell Value Dictionary*. Vols. 1-3. New Haven, CT: Yale University Press.
- Lohr, S. L. (1998). *Sampling: Design and Analysis*. 1st ed. Pacific Groves, CA: Duxbury.
- Lowe, W. (n.d.). "Software for Content Analysis: A Review." <http://people.iq.harvard.edu/~wlowe/Publications/rev.pdf> (accessed on April 25, 2008).
- MacKay, K. L., and D. R. Fesenmaier (1997). "Pictorial Element of Destination in Image Formation." *Annals of Tourism Research*, 24: 537-65.
- Macnamara, J. R. (2003). "Media Content Analysis—Its Uses, Benefits and Best Practice Methodology." <http://www.carmaapac.com/downloads/Media%20Content%20Analysis%20Research%20Paper.pdf> (accessed on April 25, 2008).
- Mehmetoglu, M., and G. M. S. Dann (2003). "ATLAS/ti and Content/Semiotic Analysis in Tourism Research." *Tourism Analysis*, 8: 1-13.
- Miles, M. B., and E. A. Weitzman (1994). "Choosing Computer Programs for Qualitative Data Analysis." In *Qualitative Data Analysis: An Expanded Source Book*, edited by M. B. Miles and M. A. Huberman. London: Sage, pp. 311-17.
- (1996). "The State of Qualitative Data Analysis Software: What Do We Need?" *Current Sociology*, 44 (3): 206-24.
- Morris, R. (1994). "Computerized Content Analysis in Management Research: A Demonstration of Advantages and Limitations." *Journal of Management*, 20: 903-31.
- Mossberg, L., and I. A. Kleppe (2005). "Country and Destination Image—Different or Similar Image Concepts?" *Service Industries Journal*, 25 (4): 493-503.
- Muskens, G. (1985). "Mathematical Analysis of Content." *Quality & Quantity*, 19: 99-103.
- Namenwirth, J. Z., and R. P. Weber (1987). *Dynamics of Culture*. Winchester, MA: Allen and Unwin.
- Neuendorf, K. A. (2002). *The Content Analysis Guidebook*. Thousand Oaks, CA: Sage.
- Neuman, W. L. (1997). *Social Research Methods: Qualitative and Quantitative Approaches*. Needham Heights, MA: Allyn & Bacon.
- Newbold, C., O. Boyd-Barrett, and H. Van Den Bulck (2002). *The Media Book*. London: Arnold (Hodder Headline).
- Oksanen, J. (n.d.). "Extrapolated Species Richness in a Species Pool." <http://www.stat.ucl.ac.be/ISdidactique/Rhelp/library/vegan/html/specpool.html> (accessed on May 13, 2008).

- Ooi, C.-S. (2004). "Poetics and Politics of Destination Branding." *Scandinavian Journal of Hospitality and Tourism*, 4 (2): 107-28.
- Osgood, C. E. (1959). "The Representational Model and Relevant Research Methods." In *Trends in Content Analysis*, edited by I. de S. Pool. Champaign: University of Illinois Press.
- Patton, M. Q. (1990). *Qualitative Evaluation and Research Methods*. Newbury Park, CA: Sage.
- Pike, S. (2002). "Destination Image Analysis: A Review of 142 Papers from 1973 to 2000." *Tourism Management*, 23 (5): 541-49.
- Pool, I. de S. (Ed.) (1959). *Trends in Content Analysis*. Champaign: University of Illinois Press.
- Popping, R. (1997). "Computer Programs for the Analysis of Texts and Transcripts." In *Text Analysis for the Social Sciences: Methods for Drawing Statistical Inferences from Texts and Transcripts*, edited by C. W. Roberts. Mahwah, NJ: Lawrence Erlbaum, pp. 79-100.
- Rainer, A., and T. Hall (2003). "A Quantitative and Qualitative Analysis of Factors Affecting Software Processing." *Journal of Systems and Software*, 66 (1): 7-13.
- Ritchie, J. R. B., and G. I. Crouch (2003). *The Competitive Destination: A Sustainable Tourism Perspective*. Wallingford, UK: CABI Publishers.
- Roberts, C. W. (2000). "A Conceptual Framework for Quantitative Text Analysis." *Quality & Quantity*, 34: 259-74.
- Romano, N. C., Jr., C. Donovan, H. Chen, and J. F. Nunamaker Jr. (2003). "A Methodology for Analyzing Web-Based Qualitative Data." *Journal of Management Information Systems*, 19 (4): 213-46.
- Rouse, A., and M. Dick (1994). "The Use of NUDIST, a Computerized Analytical Tool, to Support Qualitative Information Systems Research." *Information Technology & People*, 7 (3): 50-62.
- Ryan, C., and J. Cave (2005). "Structuring Destination Image: A Qualitative Approach." *Journal of Travel Research*, 44 (2): 143-50.
- Salisbury, J. G. T. (2001). "Using Neural Networks to Assess Corporate Image." In *Applications of Computer Content Analysis*, edited by M. D. West. Norwood, NJ: Ablex, pp. 65-86.
- Schmidt, M. (2001). "Using an ANN-Approach for Analyzing Focus Groups." *Qualitative Market Research: An International Journal*, 4 (2): 100-12.
- Shapiro, G. (1997). "The Future of Coders: Human Judgments in a World of Sophisticated Software." In *Text Analysis for the Social Sciences: Methods of Drawing Statistical Inferences from Texts and Transcripts*, edited by C. W. Roberts. Mahwah, NJ: Lawrence Erlbaum, pp. 235-38.
- Shoemaker, P. J., and S. D. Reese (1996). *Mediating the Message: Theories of Influences on Mass Media Content*. White Plains, NY: Longman.
- Sills, S. J., and C. Song (2002). "Innovations in Survey Research: An Application of Web-based Surveys." *Social Science Computer Review*, 20 (1): 22-30.
- Simon, A. F., and M. Xenon (2004). "Dimensional Reduction of Word-Frequency Data as a Substitute for Intersubjective Content Analysis." *Political Analysis*, 12: 63-75.
- Stepchenkova, S., Y. Chen, and A. M. Morrison (2007). "China and Russia: Organic Destination Images in U.S. Media." *China Tourism Research*, 3 (1): 55-72.
- Stepchenkova, S., and J. E. Mills (forthcoming). "Projecting Destination Image Research into the 21st Century: A Trend Analysis." *Journal of Hospitality Marketing and Management*.
- Stepchenkova, S., and A. M. Morrison (2006). "The Destination Image of Russia: From the Online Induced Perspective." *Tourism Management*, 27 (5): 943-56.
- (2008). "Revisiting Echtner & Ritchie: Russia's Destination Image among American Pleasure Travelers." *Tourism Management*, 29 (3): 548-60.
- Tesch, R. (1990). *Qualitative Research: Analysis Types and Software Tools*. New York: Falmer Press.
- Weber, R. P. (1983). "Measurement Models of Content Analysis." *Quality & Quantity*, 17: 127-49.
- (1986). "Correlational Models of Content: Reply to Muskens." *Quality & Quantity*, 20 (2-3): 273-75.
- (1990). *Basic Content Analysis*. 2nd ed. Newbury Park, CA: Sage.
- Weitzman, E. A. (1999). "Analyzing Qualitative Data with Computer Software." *Health Services Research*, 34 (5): 1241-63.
- Wickham, M., and M. Woods (2005). "Reflecting on the Strategic Use of CAQDAS to Manage and Report on the Qualitative Research Process." *The Qualitative Report*, 10 (4): 687-702.
- Woelfel, J. (1998). *CATPAC: Users' Guide*. New York: RAH Press.
- Woodroom, E. (1984). "'Mainstreaming' Content Analysis in the Social Science: Methodological Advantages, Obstacles, and Solutions." *Social Science Research*, 13 (1): 1-19.
- Zuell, C., and J. Landmann (2004). "Computer-Assisted Content Analysis without Dictionary?" Paper presented at the Sixth International Conference on Logic and Methodology (RC33)—Recent Developments and Applications in Social Research Methodology, Amsterdam, the Netherlands.

**Svetlana Stepchenkova** is a doctoral candidate in the hospitality and tourism management program at Purdue University, West Lafayette, IN.

**Andrei P. Kirilenko** is an associate professor in the Department of Earth System Science and Policy at the University of North Dakota, Grand Forks.

**Alastair M. Morrison** is a distinguished professor at the College of Consumer and Family Sciences at Purdue University, West Lafayette, IN.