A Novel Design Science Approach for Integrating Chinese User-Generated Content in Non-Chinese Market Intelligence

Completed Research Paper

Aaron W. Baur
ESCP Europe Business School Berlin
Heubnerweg 8-10, 14059 Berlin
abaur@escpeurope.eu

Janna Lipenkova
Anacode GmbH
Kurfürstendamm 76, 10709 Berlin
janna.lipenkova@anacode.de

Julian Bühler
ESCP Europe Business School Berlin
Heubnerweg 8-10, 14059 Berlin
jbuehler@escpeurope.eu

Markus Bick
ESCP Europe Business School Berlin
Heubnerweg 8-10, 14059 Berlin
mbick@escpeurope.eu

Abstract

Market research has long relied on reactive means of data gathering, such as questionnaires or focus groups. With the wide-spread use of social media, millions of comments about customer opinions and feedback regarding products and brands are available. However, before using this 'wisdom of the crowd' as a source for marketing research, several challenges have to be tackled: the sheer volume of posts, their unstructured format, and the dozens of different languages used on the internet. All of them make automated usage of this data challenging. In this paper, we draw on dashboard design principles and follow a design science research approach to develop a framework for search, integration, and analysis of cross-language user-generated content. With 'MarketMiner', we implement the framework in the automotive industry by analyzing Chinese auto forums. The results are promising in that MarketMiner can dramatically improve utilization of foreign-language social media content for market intelligence purposes.

Keywords: Social Media Analytics, Text Mining, User-Generated Content (UGC), Market Research, Design Science, Design Science Research, Information Systems Design, Business Systems Engineering, Information Systems Research, Information System Design Online Reviews, Decision Support Systems, Business Intelligence, Management Support Systems, Dashboard Design

Introduction

Consider working as a European market researcher for a Western automotive manufacturer. As China constantly becomes more and more decisive for the success of your employer, it is your task to find out how Chinese customers think about your car models, their components, design, flaws, and which feelings they associate with them. Traditional market research methods such as customer surveys, customer interviews, or focus groups are typically applied for the purpose of answering these questions (Parment 2014). However, these methods are tedious, complex and time-consuming (Kim et al. 2010). The survey as the traditional 'silver bullet' of social science therefore loses backing amongst scientific and commercial (market) researchers (Bruhn 2010). To save costs and to move their craft into the realities of the 21st century, researchers increasingly substitute classical customer surveys and gather their data with non-reactive means, i.e., through observation and listening rather than questioning (Haenlein and Kaplan 2012).

Analyzing user-generated content (UGC) in social media may be the most prominent member of this 'non-reactive' market research means. The growth of Web 2.0 has generated large amounts of data, actively created by Internet users in forums, product review sites and other social media channels (Kaplan and Haenlein 2010). This richness of customer ideas, opinions and feedback opens up new opportunities to listen to and benefit from the 'wisdom of the crowd' (Berthon et al. 2012). Since the early 2000's, companies have gathered and analyzed customer reviews (Berry and Otley 2004; Bryman 2012). But before considering this 'social media analytics' or 'market intelligence' a valuable new component of market research and product development, several issues arise. Especially the well-known '4 Vs' of Big Data, i.e., volume, variety, velocity and veracity (Zikopoulos et al. 2012), pose challenges to an efficient and effective use of this information in academia and practice.

First, the pure dimension of growth of UGC makes exhaustive manual search, analysis, and comprehension impossible (Choudhury et al. 2011). Automated solutions are increasingly indispensable. Second, the customer feedback valuable for market research is generally presented in a variety of different formats, e.g., semi-structured textual content (XML data from RSS feeds), unstructured textual content (text data from forums, online groups, social networking sites) or unstructured visual content (photo and video images from social multimedia sites). Gathering, pre-processing and integration across sources is therefore non-trivial (Gelman and Wu 2011; Jones 2011). Third, as users write online content mostly in their mother tongue, customer feedback which is valuable for multinational corporations exits in hundreds of different languages. Search platforms and analytics tools, however, are normally designed for a specific grammatical and semantic structure and are limited in their capability to process multi-language queries (Kawamura 2010). Thus, automated literal or conceptual translations are needed.

These pain points were identified through literature review and eighteen practitioner interviews. Additionally, the significance of opportunities and relevance of challenges, especially for the business-world, are underpinned by the renowned Marketing Science Institute (MSI), which biennially issues its 'Research Priorities' (Keller 2014). The study captures "the areas of most interest and importance to MSI member companies." The two highest-ranked priorities *Understanding Customers and the Customer Experience* and *Developing Marketing Analytics for a Data-Rich Environment* explicitly call for the development of "new analytical methods to gain greater insights from unstructured data (e.g., social media)."

These business intelligence and analytics (BI&A) challenges, rooted primarily in marketing, can only be faced with a concerted research effort (Appleford et al. 2014). Disciplines such as Computer Science, Information Systems, Economics, Psychology, Media and Communication Sciences as well as Computational Linguistics must join forces, leading to human-centered, powerful IT systems with advanced text analytics capabilities (Baur et al. 2014b; Consalvo and Ess 2011).

Our main objective in this paper is to propose and develop a system that provides an infrastructure capable of accessing, integrating, analyzing and visualizing Chinese-language user-generated content from various social media sites. It is built on modular, multi-layered framework architecture and offers advanced functionalities for content integration, concept analytics, and visualization. The analysis results of this integrated customer-research tool, which we call *MarketMiner*, are displayed in English and are easily usable for non-Chinese speakers. The tool offers an effective data collection and analysis support on large volumes of user-generated content, obtained from a variety of publicly accessible Chinese-language forums.

This study contributes to academia and practice in the fields of market research and product development. As a first field of application for *MarketMiner*, we chose the automotive industry. The system can be of interest and value to automobile-related professionals, e.g., executives from strategic product management, market research, marketing and sales, or strategic management. Additionally, computer and information systems researchers as well as social scientists in general will also profit from the system. Its utility, quality, and efficacy as well as its appropriateness to solve a whole class of identified problems are confirmed through real-life case studies and expert evaluations.

We follow the logic of Chen et al. (2012, p. 1166) and define business intelligence and analytics as a unified term embracing "the techniques, technologies, systems, practices, methodologies, and applications that analyze critical business data to help an enterprise better understand its business and market and make timely business decisions". Besides underlying data processing and analytical technologies, BI&A comprises practices and methodologies specifically relevant to high-impact applications. As we specifically look at BI&A that takes UGC as its main data source, good examples of such applications include e-commerce (Doan et al. 2011; Zwass 2010), market intelligence (di Gangi et al. 2010), science and technology (Hand 2010), politics (Karpf 2009; Wattal et al. 2010), health care (Gao et al. 2010), and public safety (Dang et al. 2014). Since the focus of this research is the (commercial) automotive industry with respective customer feedback and sentiment, we deem the term *market intelligence* as most appropriate for describing the area of application. Therefore, this term will be used in the remainder of this paper.

This study is conducted following the design science research (DSR) guidelines of Hevner et al. (2004), Peffers et al. (2007) and Gregor and Hevner (2013). Hence, the system *design-build-evaluate* cycle is followed. We begin with a short analysis of prior work related to the analysis of multi-language user-generated content and its areas of application. Dashboard design principles (Few 2006) as the implementational basis are also shortly presented. We then report on the 'design search' process and a detailed description of the developed IT artifact, representing the system 'build' part of the cycle. We continue with a systematic and rigorous evaluation of the IT artifact using two different methods, expert evaluation and case studies, detailing the system 'evaluate' part of the cycle. A discussion of implications and contributions to research and practice follows before the paper concludes.

Related Work

This section grounds the research project in the available literature pertinent to the research question. Our research contributes an instantiation that integrates content from several Chinese online sources, aggregates and analyzes it, and bridges the language barrier to provide English-language output. This output aims to support better corporate decision making. Therefore, work in the areas of management support systems, unstructured data, data integration, and cross-language support is highlighted here.

Management Support Systems (MSS)

Partially, our research draws from previous studies in a broad class of systems whose fundamental purpose is the support of managerial actions and decision making. We follow the broad terminology of Scott Morton (1984) and label this class as management support systems (*MSS*). According to Clark et al. (2007), these systems are known as decision support systems (*DSS*, Alter 2004), executive information systems (*EIS*, Walls et al. 1992; Watson and Frolick 1993), knowledge management systems (*KMS*, Alavi and Leidner 2001; Bick et al. 2012; Lee and Choi 2003; Schultze and Leidner 2002), and business intelligence (*BI*, Anderson-Lehman et al. 2004; Rouibah and Ould-ali 2002). The common goal of all systems is to reduce uncertainty in the decision making process. While these systems have their roots in the 'offline-age', the paradigm-shift (Chang et al. 2014) to the online world has widened the technical and managerial requirements.

Unstructured data wave with the shift from Web 1.0 to Web 2.0

The bi-directionality of the internet is one of the decisive differentiations compared to previous technologies of mass communication (Dellarocas 2003). However, the 'Web 1.0' (Chen et al. 2012) consisted mainly of professionally published content, which followed structured, homogeneous patterns, i.e., very similar to traditional media like newspapers, TV, radio ("one-to-many communication", Stevens 1981, 2013). Zhan et al. (2009) and Netzer et al. (2012) contend that much of the academic research in the past has focused on developing tools to deal with structured content created by webmasters or news portals, which is far easier to process. Exemplary applications include the mining of scientific databases (Dang et al. 2011), news and commentary providers in the IT domain (Marshall et al. 2004), or official statistics (Avrahami et al. 2006). In all cases, the content in the analyzed resources followed a specified, consistent structure.

However, with the advent of 'Web 2.0' (O'Reilly 2005), i.e., the wide-spread publishing of mostly non-professional content, the bi-directionality was actually realized. Hence, companies have realized the enormous potential of this data pool. They seized the opportunity to overcome the conventional one-way marketing and instead establish a two-way conversation between business and customers (Lusch et al. 2010). They started efforts for the automated gathering and analyzing of online media content in manifold ways.

Mining user opinions about products, services, the respective brands and emerging trends in a commercial context is known as market intelligence (e.g., Lee and Bradlow 2011). It encompasses using this large volume of data for enhanced, data-driven decision-making in several functional departments of the firm. For example, in marketing and public relations, 'social media monitoring' is used to watch sentiments about brands and competitors or to track the success of a recent advertising campaign.

Integrating user-generated content across multiple sites

The review of work about content integration across different social media sites confirms its very limited automation support for users. The process of accessing, aggregating and visualizing the unstructured content is still in its infancy. Gelman and Wu (2011) study the real estate information service Zillow.com and combine structured and unstructured information sources. Jones (2011) apply emergent user-generated content in a crisis situation with already available, situation-relevant information. He argues for the development of technologies that generate a more powerful exploitation of social media content in crisis response. Becker et al. (2010) propose an approach to identify major events of internet users by integrating the rich contexts of both textual and non-textual features. They establish document similarity metrics to enable online clustering of media to events and evaluate these techniques on datasets of event images from Flickr. Lau et al. (2011) design an instantiation of novel computational text mining models and integrate them into a semantic language model for the detection of untruthful reviews. The models are evaluated based on a real-world dataset collected from amazon.com. Interestingly, their sample data in the automotive product category shows 1.49% spam, i.e., fake or non-reviews.

These studies provide considerable value, but fall short of addressing the challenge of integrating and merging UGC across several sources into one application or dashboard.

Information retrieval across multiple languages

Of the approximately 7,000 languages spoken today, about 150 (2%) are used online (Kornai 2013). English is about to lose its lead as the most popular language on the internet to Chinese (Miniwatts Marketing Group 2013). This underpins the necessity to research information retrieval systems that cross the language boundaries to make user-generated content universally available. Important movements are the inclusion of cross-language tracks at *TREC*[1] in the US, the set-up of *The CLEF Initiative*[2] in Europe and *NTCIR*[3] in Japan.

One project in this field of research is the study by Yin et al. (2013). The authors follow the DSR guidelines of van Aken (2004) and propose an ontology-based linguistic model to identify basic appraisal expressions in Chinese product reviews and introduce feature-opinion pair (FOP) identification. Based on comparative experiments, they show that their approach is efficient in obtaining a more accurate result compared to state-of-the-art algorithms. Additionally, the unstructured product reviews are converted into structured and machine-sensible expressions, which can then be further processed in MSS. Dang et al. (2014) develop an artifact called *Dark Web Forum Portal* that supports the gathering and display of social media content concerning national security. It is capable of integrating user-generated content from

---

[1] Text REtrieval Conference (sic!), http://trec.nist.gov/

[2] Conference and Labs of the Evaluation Forum, formerly known as Cross-Language Evaluation Forum, http://www.clef-initiative.eu/

[3] NII Testbeds and Community for Information access Research, http://research.nii.ac.jp/ntcir/index-en.html

different sources and across multiple languages. However, the tool is mainly for look-up and translation, the analytics capabilities are very limited, e.g., there is no aggregation functionality offered. Zhou et al. (2005) design and evaluate a Multilingual Information Retrieval (MLIR) web portal in the business domain in English, Chinese, Japanese, Spanish, and German. Similarly, Qin et al. (2006) construct a domain-specific bilingual (English-Chinese) lexicon and integrate cross-language information retrieval techniques in web portal developments. Talvensaari et al. (2007) apply graded relevance assessments to advance cross-language information retrieval to locate content of interest, based on certain corpus-data. Chen and Bao (2009) study the different services offered by Google Language Tools (GLT). Vulić et al. (2015) set up a cross-lingual information retrieval (CLIR) for the English-Dutch language pair. Here, they use Wikipedia as the dataset, again much more structured than general social media content. Abusalah et al. (2005) do a short review of CLIR and associated translation approaches. Finally, Ahmed and Nurnberger (2012) provide an overview of the literature of interactive CLIR tools. The tools they study are very tedious to handle, and need extensive manual work and high user concentration when checking the translations.

In summary, most papers focus on a certain isolated aspect or module. Only a very small portion of this work convincingly studies the entire chain of cross-language data collection, analysis, aggregation and visual output. We could identify a clear research gap when it comes to the development of multi-language, automated, real-time translation/concept-mapping components for user-generated content that cover the whole process.

Finally, the information output of the tool needs to be properly visualized to be easily accessible and comprehendible to decision-makers. Important design principles, such as visual perception, color form, spatial position (Ware 2000), and motion need to feed into the design search and development phase of useful artifacts.

Research Method

We used a design science approach to conduct this study. Design science research implicates the design, development and evaluation of innovative artifacts (constructs, models, methods, and instantiations) to enable organizations to accomplish information-related tasks (Hevner et al. 2004; Kuechler and Vaishnavi 2012; Peffers et al. 2007). The development and evaluation of an artifact can be considered ‘proof-of-concept’ or ‘proof-by-demonstration’ (Moody and Shanks 2003; Nunamaker et al. 1990). The artifact in this study is a prototype tool to support effective and efficient market intelligence based on Chinese-language social media content. It is implemented following well-accepted dashboard design principles (Few 2006; Hu et al. 2012; Santiago Rivera and Shanks 2015).

The research procedure followed in this study consist of three phases, each of them described according to the performed activities, applied techniques, and achieved output (Fig. 1). The methods were selected as to assure maximum research rigor (Nicolai and Seidl 2010).

The research approach begins with phase 1) *problem identification*, where we specified the research problem, provided practical relevance and justified the value of the artifact. By the term *relevance*, we follow Nicolai and Seidl[4] (2010) and refer mainly to *instrumental relevance:* we are building an artifact that practitioners can use to solve a problem in their organization. The structured literature review was conducted adhering to the principles set forth by Webster and Watson (2002). Hence, the available literature in the field of multi-language information retrieval, UGC and MSS was identified and screened. For this purpose, both scientific sources (EBSCO Business Source Complete, ACM Digital Library, Web of Science, IEEE Xplore, Science Direct) and more practice-oriented sources (trade magazines, industry reports, IT/technology magazines, reviews, blogs, market research publications, company Web sites) were accessed. This literature review revealed that interrelations and implications of advanced access to foreign-language user- generated content are not well researched; this underpins the relevance of the research subject. Hence, we chose an explorative method (Sonnenberg and vom Brocke 2012a; 2012b).

---

[4] In their empirically derived taxonomy, they differentiate three forms of practical relevance, i.e., instrumental relevance, conceptual relevance and legitimative relevance. Instrumental relevance is comprised of schemes, technological rules/recipes and forecasts.

Based on the literature review, an interview guideline was developed, adhering to common qualitative research standards (Kvale and Brinkmann 2015). Eighteen semi-structured, on-premise practitioner interviews (*field research*, Abraham et al. 2014) confirmed the actual existence of the problem in practice, the lack of suitable technologies and the business need for a solution. The selection of the interview partners and the respective eleven case companies was based on a diligent, purposive, theory-driven sampling strategy (Miles and Huberman 1994). This ensured the inclusion of most aspects of the underlying theories in the evidence gathered from informants (Kvale and Brinkmann 2015). It also enables comparisons as well as theoretical and literal replication (Yin 2003). Inclusion criteria were ‘automotive company’ or ‘professionals service firm working for the automotive industry’ with ‘business in China’ and they were carried out until a theoretic saturation had been reached (Miles et al. 2013). In the literature, multiple case studies enhance validity (Eisenhardt 1989, Eisenhardt and Graebner 2007), and the convergence of statements and observations increases confidence in the findings (Eisenhardt 1989). The interviews lasted between 45 and 96 minutes. Multiple interview partners were questioned at most companies to avoid a respondent bias and to do a first triangulation of findings. The inclusion of managers from different hierarchical levels, working for firms of several vertical levels of the industry, i.e., OEMs, suppliers and service providers, should assure the formation of a holistic picture and mitigate the possibility of missing important insights (see the case description in Appendix A). The interviews were transcribed, coded and analyzed. Finally, one of the researchers met with two native Chinese managers of a social media agency specialized in UGC analysis for automotive clients. They independently formulated scenarios of the use cases that should be solved by the artifact. After the development, the scenarios were evaluated by these social media analytics experts.

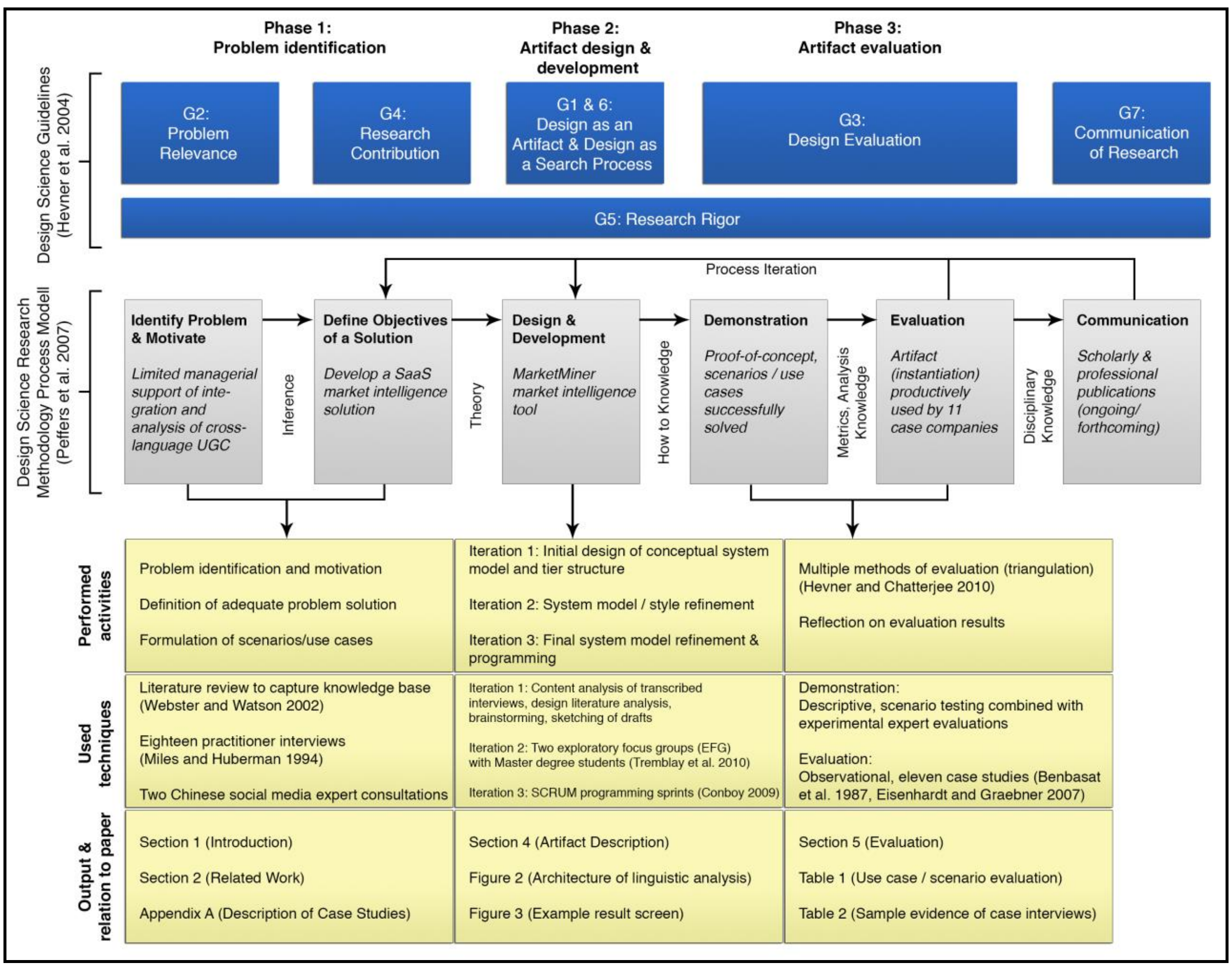

Figure 1. Applied design science research model, based on (Hevner et al. 2004; Neff et al. 2014; Peffers et al. 2007; Stindt et al. 2014)

Phase 2), *artifact design and development*, followed. The first iteration included another content analysis of the transcribed interviews (field research), a review of relevant literature about previous, similar attempts and respective software designs (desk research), several team brainstorming sessions, and a multitude of sketches using *Mind Objects*, a tool for visual requirements engineering (http://www.mind-objects.com). The sketches were then discussed with a projector and flip charts in two separate exploratory focus groups (EFG), each comprising six Master students enrolled in management-related degrees as participants. As demonstrated by Moody and Shanks (2003), graduate students are deemed suitable for acting as surrogates for the actual target users of management-related software tools. The students were recruited with the help of university professors, participation was voluntary; there were no credit points earned or grades influenced. A senior researcher acted as moderator and a junior researcher as observer (as suggested by Hevner and Chatterjee 2010; Tremblay et al. 2010). Information systems as socio-technical systems need to cater explicitly for easy user comprehension and high user experience (Winograd 1996). At this point of the design process, therefore, issues of style, visualization and information representation were extensively discussed (Simon 1996). Several style sketches of the artifact were creatively varied, presented and assessed. The student's feedback then fed into the final iteration and the actual programming of the artifact, which was done with a multidisciplinary development team using agile, incremental programming principles (*Scrum*, Conboy 2009).

After the artifact was developed, we continued with phase 3), *artifact evaluation*. Design science guidelines call for a rigorous demonstration of the utility, quality and efficacy of the developed artifact (Gregor and Hevner 2013). As we are striving for instrumental relevance, specifically the utility and degree of problem-solving is of importance. In this paper, the artifact was evaluated using a threefold approach: qualitative expert evaluations, scenario analysis, and multiple case studies. This triad triangulates the findings and strengthens the evidence base. The case studies involved the same companies and interviewees as in phase (1). Details of the evaluation are described in the respective sections below.

Artifact Description

After the problem had been identified and structured in phase 1), we continued with phase 2), *artifact design and development*.

Design Search Process

Design is fundamental in IS (Glass 2000; Winograd 1998). When designing this artifact, we combined a multidisciplinary team of scholars from computer science, information systems, China studies, linguistics and design. The team goal was explicitly set to solve the identified managerial issue, i.e.,

> *How can a market intelligence tool support the gathering, analysis and visualization of Chinese-language user-generated content for Western users?*

With an innovative solution, a business problem was to be solved, with implications for society as a whole. Through its usage, the solution should ensure a maximum level of competitive advantage (Wade and Hulland 2004). In the course of the iterative design process of the framework, we had to decide which industry to start with. Three constituting considerations lead to the automotive industry as our pilot business application, which we named *MarketMiner*.

First, automobiles are very complex products (Abrahams et al. 2013). They therefore yield abundant possibilities to develop a rich ontology with many product attributes and industry aspects. If the system works with such a multifaceted product, other industries like Fast Moving Consumer Goods (FMCG) or Tourism are likely to be feasible as well. Second, generally Chinese customers see the automobile as a status symbol and assign a high relative value to it (Sha et al. 2013). This leads to a massive amount of automobile-related online user-generated content. On the one hand, this yields a sufficient data base for valid analyses; on the other hand it increases the need for an automated analysis approach. And third, mainly from a practical impact perspective, China can be considered the key market for Western automotive manufacturers. It is the second largest and fastest growing market, with 20 mn cars and 4 mn trucks sold in 2014 (STATISTA 2015). This makes this artifact highly relevant for practical purposes.

Hence, we started desk research and case study analyses. Well-established design literature such as Few (2006) or Ware (2000) brought important impetus on how to focus on clear data visualization, a human-centered design, and high usability. Several team-internal iterations then lead to sketches, prototypes, and a pilot version of the tool. Details of this process have been highlighted in phase 2) of the research method section above.

Artifact Development Process

The framework is composed of a classical three-tier architecture, i.e., database tier, business logic tier, and presentation tier. Each layer is optimized with specific functions to address gathering, integration, analysis, and visualization of cross-lingual user-generated content. Specifically, three core functions, including crawling, concept-mapping and effective visualization, are implemented to address the challenges discussed in the previous sections. In the following, we describe each tier and the major functions it offers.

Database tier

The database tier is responsible for data collection and preprocessing.

Data collection is performed with a focused crawler which runs on a selection of major Chinese automotive forums. These websites were selected with the help of two social media experts working for market research agencies that deliver social media analytics/market intelligence to Western automotive OEMs in China. The selection was based both on quantitative and on qualitative criteria: to enhance the capacity of the product to handle data specific to the automotive sector, we targeted resources with highly detailed user opinions. Thus, the prototype of *MarketMiner* works with a collection of ca. 2 mn customer opinions with an average length of 107 words. Additionally to the text content, each opinion is accompanied by a set of metadata that are also used in the final evaluation. These metadata include general user information such as his location, contexts of use of the product, purchase price etc.

Data preprocessing includes cleaning, normalization and linguistic preprocessing. Cleaning mainly pertains to duplicate elimination. Normalization is performed both on the metadata and the text data; it involves the unified representation of named entities (product/brand names, locations etc.) as well as a number of text normalizations (numbers, punctuation, emoticons etc.). The final preprocessing stage - linguistic preprocessing – includes sentence and word segmentation as well as part-of-speech (POS) tagging.

Business logic tier

The business logic tier analyzes the so-far unstructured collection of metadata and text data and transfers it into business-relevant concepts and relations, i.e., it performs concept-mapping and matching of foreign language terminology to the respective counterparts in English. In our case, the target representations are pairs of features and evaluations/emotions, optionally complemented by contextual information.

This tier consists of two components: the lexical resources as a static component and the linguistic analysis as the processing component. The general structure is shown in Fig. 2. In the following, we briefly describe the structure of the resources as well as the analysis steps. For a more detailed description, see Lipenkova (2015).

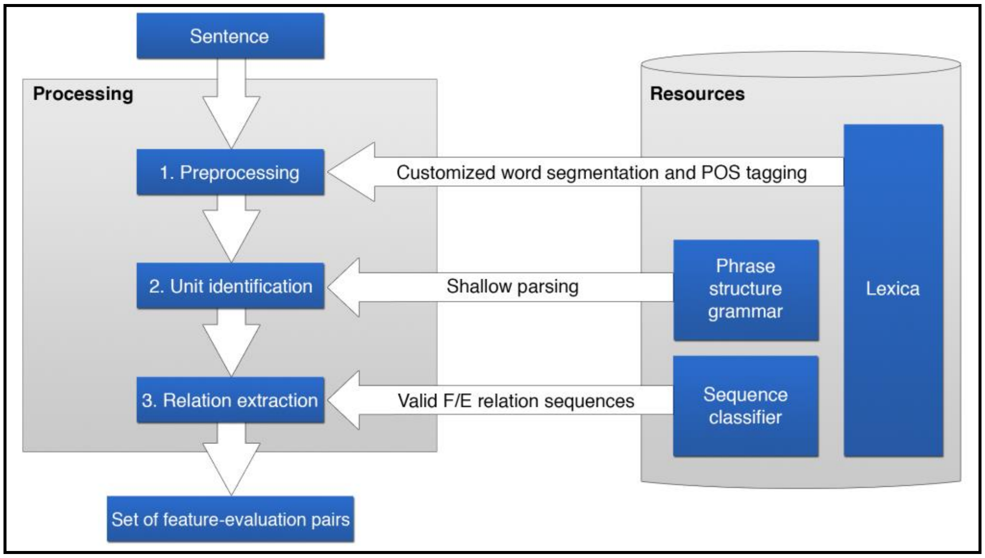

Figure 2. Architecture of the linguistic analysis component

1) Lexical resources

The goal of the lexical resources is to store and organize the relevant concepts as well as the linguistic expressions (mostly words) that describe these concepts. We use two lexical resources: a domain-independent lexicon which captures the general expression of sentiment and emotion in Chinese, as well as a domain-dependent ontology which captures automotive concepts and words (Bera et al. 2011).

The lexicon contains three categories of items:

- *Polarity words* are adjectives that imply a positive or negative evaluation, e.g., 好 'good', 棒 'great' 糟透 'horrible'. Conceptually, these words are mapped to a positive/negative scale with integer values between -3 and 3.
- *Emotion words* are nouns, adjectives and verbs that denote one of the basic emotions - Anger, Disgust, Fear, Joy, Sadness, Surprise (Ekman 1992). For example, the three words 烦恼 'anger', 生气 'angry' and 烦 'to make angry' fall under the emotion category Anger. Additionally to the emotion category, these words are also assigned an integer value on a scale from 1 to 3 which describes the intensity of the signified emotion.
- *Intensity modifiers* are adverbs that modify the intensity of emotion or polarity words. They can be degree adverbs, such as 很 'very' and 有点儿 'a bit', as well as negation words, such as 不 'not' and 无 'without'. In our model, intensity modifiers become meaningful only once they are combined with a polarity or emotion expression. They are associated with a real value which is multiplied by or added to the integer value of the modified expression.

In contrast to the lexicon with its fairly flat structure, the ontology has a hierarchical architecture. It is developed in the OWL/RDF framework using Stanford University's Protégé open-source ontology editor. OWL, the Web Ontology Language, can be regarded as the most widely used visual ontology language (Fliedl et al. 2010). The ontology includes a fine-grained conceptual classification, pertaining mainly to feature specifications. Each concept is seen as a 'container' for linguistic expressions which may denote different facets of the concept. For example, the feature *VisualAppearance* contains the expression 外观 ('visual appearance'), but also related evaluation concepts, such as 丑 ('ugly') and 好看 ('beautiful').

Four major feature types can be distinguished in the ontology:

1. Components of the product, e.g., Seats, Door, Chassis

2. "Physical" dimensions of evaluation, e.g., Size, Attrition, Performance
3. Subjective dimensions of evaluation (individual perceptions), e.g., Elegance, Enjoyment, Taste
4. Typical actions performed and evaluated during product use, e.g., Stopping, Braking, Steering

These features can be used independently or in combination with each other. For example, Performance can be used as a standalone feature to denote the overall performance of the car. On the other hand, it can be combined with a more specific component, e.g., the performance of the air conditioning.

2) Analysis process

The analysis is performed on sentences as basic units of linguistic utterances. The process is split into three steps, each corresponding to a level of linguistic representation:

1. Lexical preprocessing
2. Phrasal analysis
3. Sentence analysis

Lexical preprocessing involves the lookup of the words in the lexical resources and the assignment of custom tags.

The phrasal analysis identifies syntactic units ('chunks', Abney 1991) that describe features and evaluations. These units may consist of single words, e.g., 前脸 ('front side') and 漂亮 ('beautiful'). They can also have a more complex structure and meaning, e.g., 前脸的外观 ('the look of the front side'), 很时尚漂亮 ('very fashionable and beautiful'). The analysis of chunks is based on syntactic rules that are formed from the tags assigned during lexical preprocessing.

Sentence analysis extracts relations between chunks, specifically relations between evaluations and target features. It uses a Support Vector Machine (SVM) classifier (Joachims 1998) to decide whether the word sequence between a feature and an evaluation chunk establishes or disrupts a feature/evaluation relation.

Finally, rules of semantic compositionality are used to interpret the resulting relations and to form the set of conceptual feature/evaluation pairs for the sentence.

Presentation tier

For information systems, presenting decision-relevant information in a clear and logical manner can enhance user acceptance, task performance, and overall system assessment (Shneiderman 1996). For visualizing the output, we use an English-language graphical user interface (HTML5 with JavaScript) with customizable visualization possibilities. The information visualization techniques of Ware (2000) and Few (2006) were the guiding principles.

The main goal of *MarketMiner* is to show how features of the product and of related services are evaluated by the users. The user interface builds on a continuum from general to specific information: the software allows the user to view evaluations and perceptions for different levels of granularity. Thus, queries might be targeted at very broad topics, e.g., the general perception of a brand/model etc., but also at fine-grained details such as the evaluation of specific aspects of some component (e.g., the precision of the steering wheel). The selection of the query parameters happens in a menu that is identical to the relevant parts of the ontology; thus, we achieve a direct alignment between the linguistic analysis in the backend and the conceptual organization of the frontend.

The results of the query can be visualized according to individual user preferences. By default, we use boxplots, which provide an intuitive representation of multiple properties of the data set (Schmidt and Hollensen 2006). Figure 3 shows the results for an example query for the feature VisualAppearance of BMW/3-Series. In the leftmost graph, we see the evaluation of the general feature; the remaining graphs show the evaluations for subfeatures of VisualAppearance, such as Elegance, Temperament, Fashion, etc. All evaluations fit onto a scale from -4 to 4 and the two colors in the boxes visually mark the distribution of the evaluations around the median value.

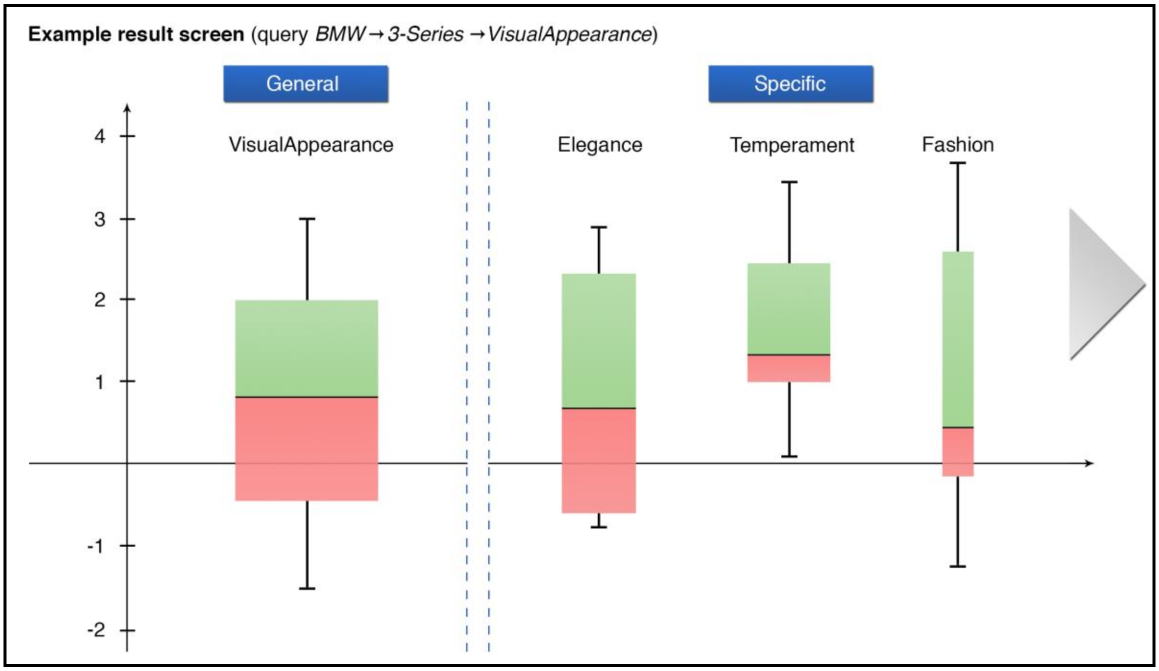

Figure 3. General and specific results for the feature VisualAppearance of BMW 3-Series[5]

Other possible representations are line charts series for the visualization of changes over time, pie charts for customer segment distributions, radar charts for competitor analysis as well as geographical maps for the distribution of the locations at which the evaluations are made (Huang et al. 2005). Visualizing textual information derived from large bodies of text remains challenging, since text cannot easily be described by numbers (Keim 2002). In subsequent releases, however, complementary presentation formats including multidimensional and text overlay methods (Keim 2002; Wise 1999) may be explored.

Additionally to the aggregated representation of the evaluation results, the user has the possibility to access the original texts. For each aggregation, *MarketMiner* allows to display the original text segments in which the evaluations were found. Additionally, we provide an API to Google Translate (https://cloud.google.com/translate/docs?csw=1#Translation), which allows posts to be translated into English at real-time. The post is sent to the Google Translator server, returning the translated text within seconds.

Evaluation

After the design and build phases of the previous sections, we now continue with describing the final phase 3), *artifact evaluation*. As another crucial part of design science research, guidelines demand a systematic and rigorous evaluation of the new artifact (Hevner et al. 2004) to answer the question "How well does it work?" (March and Smith 1995). We have explored and specified the business need, i.e., support of cross-language integration, analysis and visualization of unstructured user-generated content, in phase 1) and have replied to this need with the development of the artifact *MarketMiner* in phase 2). We now describe whether the artifact successfully addresses this open problem, as the primary goal of design science research is to meet the identified business need and to create utility.

To the best of our knowledge, *MarketMiner* is the first SaaS (Software as a service) tool that integrates industry-specific user-generated content from multiple online forums in Chinese and delivers the analysis results in English. As there is no comparable system available, a classical, quantitative benchmark comparison with formulated hypotheses and specified tasks to either confirm or refuse these hypotheses was not feasible.

[5] To save page space, the lower bounds of the figure ends at -2. In reality, the range of the tool is +4 to -4.

Instead, we applied three other forms of proof: For demonstration, a descriptive, scenario testing was combined with experimental, qualitative expert evaluations. After the proof-of-concept was successfully furnished, an observational, case study approach was used for evaluation purposes (Leukel et al. 2014). This is in line with Gregor and Hevner (2013, p. 351) who assure that [...] some degee (sic!) of flexibility may be allowed in judging the degree of evaluation that is needed when new DSR contributions are made [...]". Therefore, a diligent, qualitative test and evaluation of the utility, quality and efficacy of the artifact *MarketMiner*, i.e., its problem-solving capacity, was possible (Nunamaker et al. 1990). With these complementary evaluation methods, we could cross-check and triangulate our findings and prove that the designed instantiation delivers value to the business community.

For a first demonstration and proof-of-concept (Peffers et al. 2007), we applied scenario testing/expert evaluation: Two non-Chinese experts of a firm specialized in studies involving comparisons of business intelligence & analytics software were given two tasks: On the one hand, *solve the use cases formulated by the Chinese social media experts in phase* 1). On the other hand, *qualitatively evaluate your impression of the tool*, with the expressed focus on ease-of-use and efficacy. The results of the use case/scenario evaluation are displayed in Table 1.

Number	Use case description	Use case solvable	Ease of use & efficacy
1	Create a tag cloud of most mentioned features within the last month.	●	●
2	Show user sentiment towards *material quality* of seats of BMW 3 model.	●	●
3	Show most active forum user between January 1 and December 31, 2014.	●	◕
4	Compare aggregated user sentiment towards *engine agility* of Mercedes S-class and BMW series 7.	●	●
5	Show number and median length of forum entries regarding VW.	●	●
6	Translate several forum entries from Chinese to English.	●	●
7	Set email alert if number of daily negative forum entries regarding VW Sagitar *rear shafts* exceeds a number individually set by the user.	●	◕
8	Show development of number of postings after launch of new Audi A1 Sportback in 2013.	●	●
9	Customize dashboard style to your taste with drag & drop.	◑	◕
10	Combine and aggregate user forums, news portals, microblogging sites.	○	-
11	How is the *control panel* of VW Golf evaluated in terms of clarity, design, and usability?	●	●
12	How is relative brand sentiment of BMW versus Mercedes over time?	●	●
13	How does brand CSR impression differ by gender and location of users?	◑	◔
14	Which car model is rated best in utility for commuters?	●	●
15	How do users evaluate re-sale value of BMW series 1, 3, 5 and 6?	●	◕
16	Which features are evaluated most positively of the Audi car models?	●	●
Use case solvable: *Fully* ● *Partially* ◑ *No* ○ ***Ease of use & efficacy:*** *High* ● *Medium* ◕ *Low* ◔			

Table 1. Use cases/scenarios and respective evaluations

The experts could solve most of the pre-formulated scenarios. They mutually agreed that not only proof-of-concept to the functioning and usefulness was provided, but also a significant value-add to the profound issue of cross-language UGC analysis was given. Their feedback was used for another refining and programming iteration, before continuing with the case studies.

We granted our eighteen interview partners of eleven firms a four-week access to the tool. They were not given any standardized evaluation form, but rather were asked to use the tool 'productively' in their daily work. As not all of the interviewees actually do hands-on analyses, each case company nominated a lead-user, acting as primary contact point for the colleagues and the researchers. After this period, the lead users were asked to email their feedback in terms of utility, quality and efficacy to the research team. This was followed up with a Cisco Webex conference, held with each lead user to discuss their experience. Through screen sharing, some of the users presented their usage 'live' and also made suggestions for bug fixes, visual changes, or new functionality. This way, we got a detailed view of participants' usage evaluation. We analyzed the feedback and condensed it to categories. The most representative pieces of evidence per case are displayed in Table 2.

Group	Case Company	Utility	Quality	Efficacy
Automotive OEM	O1	*"MM has helped us getting a grasp on our Chinese customers"*	*"We cross-checked with Chinese colleagues, who confirmed the accuracy."*	*"I had several 'aha' effects when analyzing the data with MM, giving me insights in areas I never thought of."*
	O2	*"Not only getting insights to customers' sentiment but also drawing a strategic 'so-what' would elevate tool utility further."* • Expressed demand for implementation for truck business	• Stressed importance of more visualization possibilities	*"We employ of course native Chinese market researchers. But even they liked it, as it saved them much time and hassle of an alternative, manual approach."*
	O3	"China is [in many respects] our number one priority. This can help the HQ to support the local market research activities."	*"Surprising was the fact that even for our luxury class models there was abundant data available."*	*"The team saw many areas of application, such as QM, product management, competitor analysis and market research in general."*
Automotive Supplier	S1	*"As our parts are quite specialized, the available data for some products proved a bit limited."*	*"We [had] no cause for distrust in the results."*	*"[…] don't use much technology for market research yet. But this could convince our CMO & CTO to start investing in such things."*
	S2	*"For gathering customer feedback and initial ideas for product improvements, we will routinely use MarketMiner in the future."*	• Level of aggregation was deemed sufficient for most analyses	*"Especially for non-obvious product flaws – which are hard to replicate at the dealer – people complain about in blogs and forums. Access to that data supports our quality management efforts. Our [goals] were reached."*
Market Research Agency	M1	*"[…] helpful in many aspects. Hopefully more brands are added soon."*	*"For the automotive domain, there is nothing comparable."*	*"Efficacy is crucial, we evaluate this in detail for every tool we implement. Our preliminary feeling is: We need this cross-language power."*
	M2	*"Very helpful. Too bad the tool does not offer its current functionality plus the standard social media monitoring stuff."*	*"In these four weeks our test account worked, we used it nearly daily for project work for our automotive clients."*	• Stressed the key role MarketMiner could play for their clients, in light of the ever increasing significance of the Chinese economy
	M3	*"Especially our test-wise longitudinal studies were insightful"*	*"Quality is hard to check, just as with our social media monitoring tools."*	*"In terms of innovation research, we got several relevant ideas for our clients."*

Strategy Consultancy	C1	*"The [tool] proved valuable for several projects."*	*"Accuracy checks with other tools at our Chinese offices proved quite positive."*	*"[...] revealed many key aspects that our clients were not aware of."*
	C2	*"Advanced technology is crucial [for us]. It gives us the competitive edge we need."*	*"Quality was not the decisive point for us."*	*"MarketMiner provided us with infos we would not have had an alternative source for at all."*
	C3	*"This is useful in many scenarios, be it brand monitoring, or QM"*	*"Especially colleagues with backgrounds in AI [were surprised about] precision of the tool."*	*"For this pilot test, our self-set demands were for the most part satisfied."*

Table 2. Sample qualitative evidence from case interviews

Taken together, the gathered empirical evidence suggests that *MarketMiner* had a high utility, was easy to use and successful in solving the identified problem, i.e., supporting the retrieval and analysis of UGC market intelligence across linguistic and cultural barriers.

Discussion

The active participation of non-professionals in the Web 2.0 social media environment has led to an exponential growth of user-generated content. Gaining valuable insights into costumers' opinions, needs and attitudes is of utmost importance for market researchers and company executives. Results of our qualitative interview suggest that tools like *MarketMiner* can help to get closer to customers, to receive their feedback and hence to create better and more successful products. According to our sample, some of the interviewees indeed could get first looks at new markets and target groups as they got "a grasp on [our] Chinese customers" or figured out "key aspects" they or their customers "were not aware of".

Eventually, this increased level of insights also leads to lower costs and a higher market success of both providers and clients. Fully developed market intelligence tools can overcome major challenges, e.g., the large amount of data created every day or the variety of formats and structures of social media content, and thus boost efficiency. Our case interview results support this by emphasizing the competitive advantage of "advanced technology" which can be "crucial" while traditional data warehousing technologies might fail. This is especially true for analyses of high quality which seems to be less precise with traditional "social media monitoring tools" currently used by some of our interview partners.

Finally, our respondents were able to safe time while conducting their daily market research, especially in terms of language barriers. Even though English is the most widespread business language, the majority of internet users do not use English as their means of (online) communication. Understanding and sense-making of that foreign-language content is therefore difficult. For *MarketMiner* in particular, the conversion from Chinese to English is seen as useful because of the ability to report local activities on the Chinese market to foreign headquarters and "to support the local market research activities". But even for native Chinese, the tool was effective and "saved them much time and hassle". Still, "accuracy" was and always has to be fulfilled by the instruments in order to gain customer's confidence, e.g., in precarious scenarios of "non-obvious product flaws".

In order to effectively and efficiently employ user-generated content for market intelligence, a framework is needed for cross-platform integration, storage, analysis, and display (Baur et al. 2014a). To pursue this end, we contribute to the extant body of knowledge by advancing the search and analysis support of cross-language social media content across different Web 2.0 formats.

This study follows the design-build-evaluate guideline of design science research set forth by Hevner et al. (2004). Specifically, *MarketMiner* represents the IT artifact instantiated using advanced crawling, integration, translation, and analysis technologies. Evaluation of the framework through expert appraisal, use case/scenario analysis as well as productive use in the form of eleven case studies showed overarching support of it being useful to solve the identified problem, having a high data analysis quality and a high efficacy in reaching the stated goals.

The study creates several important implications for academics and practitioners interested in analyzing social media content. First, our results demonstrate that the use of an integrated portal is an effective and efficient mean to aggregate large volumes of product feedback, customer opinions, and market developments. Without such a system, market researchers need to browse an array of sites and familiarize themselves with various different structural formats. Second, translation support is vital in today's multitude of languages used online. Toward that end, our framework delivers both concept-mapping for a transference of foreign language terminology to the respective counterparts in English as well as real-time support to translate the underlying customer posts from Chinese to English via Google Translate. Third, the large amount of user-generated content available in the Chinese internet calls for the combination of complete and incremental crawling or spidering. Therefore, our artifact includes a daily routine and automatically updates the data repository with any new forum entries of the last 24 hours. This timeliness provides opportunities for detecting upcoming threats (such as product quality issues) or monitoring the effectiveness and impact of marketing campaigns.

However, the framework and the design science approach also bear certain limitations which should be addressed in the future. First, a hypothesis-driven, quantitative evaluation of the developed artifact was not possible, due to the lack of an appropriate benchmark. This may be healed in subsequent studies by the adoption of standardized survey forms to evaluate ease of use, utility, task performance, etc. with a large user base. These additional proof-of-use and proof-of-value analyses (Nunamaker and Briggs 2011) along with valid statistical and longitudinal measures would further strengthen the validity of the preliminary findings presented in this paper (Recker and Rosemann 2010). The development of systems to support decision making is an iterative process with active user participation (Arnott 2004). More refinement of the system has to be done in frequent feedback loops with the addressed target group. In addition, the framework in its current form solely captures textual semi-structured and structured content from a limited number of sources. An expansion to include more diverse sources and a structured storage and clear display of multimedia content, such as pictures or video files, could be of great additional value. Finally, making another leap, the framework should be expanded to include a real decision support system, i.e., interpreting the results of the UGC and formulating strategic alternatives that unfold from them. This interpretation would yield much additional value to the users.

In IS, it is possible to generalize using qualitative research methods (Conboy et al. 2012; Lee and Baskerville 2003). To enable this, and to comment more profoundly on external validity of the framework, two research directions may prove fruitful: First, the transfer of the artifact to applications outside the automotive sector, e.g., consumer products or health care. In these industries, there is also ample Chinese UGC available. Hence, transferability of the findings could be evaluated after changing the industry-dependent components of the software. Second, the system can be implemented in other languages beyond the Chinese-English pair. Again, the modular framework could be equipped with other lexica and natural language processing algorithms to test its quality and utility for other cross-language applications. Here, other emerging market languages like Arabic, Hindi, or Russian may be of high interest.

With *MarketMiner* and the underlying framework, we have displayed significant novelty *and* utility as demanded from DSR research outputs (March and Storey 2008). We could contribute by reasoning, proof-of-concept, and proof-of-acceptance and use (Davis 2005). As both research and practice will significantly benefit, we can answer the call both for rigor and relevance in IS research (e.g., Österle et al. 2011; Straub and Ang 2011).

Previous knowledge has been advanced by developing a framework useful to many of today's challenges within a globalized business world. In many cases, only the trade with emerging economies keeps the manufacturers of products in the saturated Western markets in the profit zone. Getting 'closer' to the local customers, their needs, opinions and tastes is therefor of utmost importance. MarketMiner and future tools that build on the developed framework may be steps towards that direction.

Conclusions

In this paper, we followed a DSR approach and designed, developed, and evaluated an integrated framework for automated search, storage, analysis, and visualization of foreign-language user-generated data in the automotive domain. The artifact *MarketMiner* offers convenient, English-language access to the aggregated content of various Chinese automobile forums. It thus enables the effective and efficient analysis of customer product feedback, sentiment, preferences, and market trends. As this framework is generic, it can easily be altered to build solutions for other knowledge domains. Our expert evaluations, scenarios and case studies demonstrate that *MarketMiner* enables professionals to gain unprecedented awareness of a formerly mysterious foreign-language market.

To conclude, this paper contributes to the IS community because prior work has not adequately addressed the important issues that we have tackled here: how foreign-language user-generated content can be efficiently and effectively harnessed to improve market intelligence and widen the corporate knowledge-base.

Appendix A: Description of Case Studies[6]

Group	Case Company	Number of Interviews	Company Type	Interview Partner
Automotive OEM	O1	1	European premium OEM	Head of Communication Insights
	O2	3	European mass OEM	Head of Strategy
				Senior Manager Digital/ Brand Marketing
				Team Lead Communication Strategy/ Branding & Marketing Management
	O3	2	European premium OEM	Associate Brand Strategies, Market Research and Competitor
				Associate Product Management, Small Vehicles
Automotive Supplier	S1	1	European supplier of steering systems, top 50 of global suppliers	Vice President Personnel and Services
	S2	2	European supplier of driveline and chassis technology, top 20 of global suppliers	Head of Market Development, Market Intelligence & Sales Coordination
				Project Manager Car Powertrain Technology
Market Research Agency	M1	1	European branch of one of the global leaders in (market) research, insight, consultancy	Director Automotive
	M2	1	Leading agency for content marketing & social media strategy	Director Social Media Strategies
	M3	1	Specialized agency for innovation, product development, idea generation & idea management	Team Lead Social Media Research
Strategy Consultancy	C1	4	Global strategy consultancy (top 10)	Partner, EMEA Head of Digital Business Practice
				Partner, EMEA Head of Automotive Practice
				Partner, EMEA & Global Head of Strategic IT Practice, Global Head of Digital Business Practice
				Partner, Vice President Knowledge Services
	C2	1	Global strategy consultancy (top 10)	Manager Technology Practice
	C3	1	Global strategy consultancy (top 10)	Manager Digital Services
Sum	**11**	**18**		

[6] Industry rankings (e.g., “top 50”) are based on the fiscal year 2014

References

Abney, S. P. 1991. “Parsing by Chunks,” in *Principle-Based Parsing*: *Computation and Psycholinguistics*, R. C. Berwick, S. P. Abney and C. Tenny (eds.), Dordrecht: Kluwer Academic Publishers, pp. 257–278.

Abraham, R., Aier, S., and Winter, R. 2014. “Fail Early, Fail Often: Towards Coherent Feedback Loops in Design Science Research Evaluation,” in *Proceedings of the 35th International Conference on Information Systems*, Auckland, New Zealand.

Abrahams, A. S., Jiao, J., Fan, W., Wang, G. Alan, and Zhang, Z. 2013. “What's Buzzing in the Blizzard of Buzz? Automotive Component Isolation in Social Media Postings,” *Decision Support Systems* (55:4), pp. 871–882.

Abusalah, M., Tait, J., and Oakes, M. 2005. “Literature Review of Cross-Language Information Retrieval,” *World Academy of Science, Engineering and Technology* (4), pp. 175–177.

Ahmed, F., and Nurnberger, A. 2012. “Literature Review of Interactive Cross Language Information Retrieval Tools,” *International Arab Journal of Information Technology* (9:5), pp. 479–486.

Alavi, M., and Leidner, D. E. 2001. “Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues,” *MIS Quarterly* (25:1), pp. 107–136.

Alter, S. 2004. “A work system view of DSS in its fourth decade,” *Decision Support Systems* (38:3), pp. 319–327.

Anderson-Lehman, R., Watson, H. J., Wixom, B. H., and Hoffer, J. A. 2004. “Continental Airlines Flies High with Real-Time Business Intelligence,” *MIS Quarterly Executive* (3:4), pp. 1–30.

Appleford, S., Bottum, J. R., and Thatcher, J. B. 2014. “Understanding the Social Web,” *ACM SIGMIS Database* (45:1), pp. 29–37.

Arnott, D. 2004. “Decision Support Systems Evolution: Framework, Case Study and Research Agenda,” *European Journal of Information Systems* (13:4), pp. 247–259.

Avrahami, T. T., Yau, L., Si, L., and Callan, J. 2006. “The FedLemur Project: Federated Search in the Real World,” *Journal of the American Society for Information Science and Technology* (57:3), pp. 347–358.

Baur, A. W., Breitsprecher, M., and Bick, M. 2014a. “Catching Fire: Start-Ups in the Text Analytics Software Industry,” in *Proceedings of the 20th Americas Conference on Information Systems*, Savannah, GI.

Baur, A. W., Genova, A. C., Bühler, J., and Bick, M. 2014b. “Customer is King? A Framework to Shift from Cost- to Value-Based Pricing in Software as a Service: The Case of Business Intelligence Software,” in *Proceedings of the 13th IFIP WG 6.11 Conference on e-Business, e-Services, and e-Society*, Sanya, China.

Becker, H., Naaman, M., and Gravano, L. 2010. “Learning similarity metrics for event identification in social media,” in *Proceedings of the third ACM international conference on Web search and data mining*, B. D. Davison (ed.), New York, NY: ACM, pp. 291–300.

Bera, P., Burton-Jones, A., and Wand, Y. 2011. “Guidelines for designing visual ontologies to support knowledge identification,” *MIS Quarterly* (35:4), pp. 883–A11.

Berry, A. J., and Otley, D. 2004. “Case-Based Research in Accounting,” in *The real life guide to accounting research: A behind-the-scenes view of using qualitative research methods.*, C. Humphrey and B. Lee (eds.), Amsterdam, Boston: Elsevier, pp. 231–256.

Berthon, P. R., Pitt, L. F., Plangger, K., and Shapiro, D. 2012. “Marketing meets Web 2.0, social media, and creative consumers: Implications for international marketing strategy,” *Business Horizons* (55:3), pp. 261–271.

Bick, M., Hetmank, L., Kruse, P., Maier, R., Pawlowski, J., Peinl, R., Schoop, E., Seeber, I., and Thalmann, S. 2012. “Manifesto for a Standard on Meaningful Representations of Knowledge in Social Knowledge Management Environments,” in *Tagungsband Multikonferenz Wirtschaftsinformatik*, D. C. Mattfeld and S. Robra-Bissantz (eds.), Braunschweig, pp. 1–17.

Bruhn, M. 2010. *Unternehmens- und Marketingkommunikation*: *Handbuch für ein integriertes Kommunikationsmanagement*, München: Vahlen.

Bryman, A. 2012. *Social Research Methods*, Oxford, New York: Oxford University Press.

Chang, R. M., Kauffman, R. J., and Kwon, Y. 2014. “Understanding the paradigm shift to computational social science in the presence of big data,” *1. Business Applications of Web of Things 2. Social Media Use in Decision Making* (63:0), pp. 67–80.

Chen, H., Chiang, R. H. L., and Storey, V. C. 2012. "Business intelligence and analytics: From big data to big impact," *MIS Quarterly* (36:4), pp. 1165–1188.

Chen, J., and Bao, Y. 2009. "Cross–language search: The case of Google Language Tools," *First Monday* (14:3).

Choudhury, M. D., Counts, S., and Czerwinski, M. 2011. "Identifying relevant social media content: leveraging information diversity and user cognition," in *Proceedings of the 22nd ACM Conference on Hypertext and Hypermedia,* Eindhoven, Netherlands, New York, NY: ACM Press, pp. 161–170.

Clark, T. D. J., Jones, M. C., and Armstrong, C. P. 2007. "The dynamic structure of management support systems: Theory development, research focus, and direction.," *MIS Quarterly* (31:3), pp. 579–615.

Conboy, K. 2009. "Agility from First Principles: Reconstructing the Concept of Agility in Information Systems Development," *Information Systems Research* (20:3), pp. 329–354.

Conboy, K., Fitzgerald, G., and Mathiassen, L. 2012. "Qualitative methods research in information systems: motivations, themes, and contributions," *European Journal of Information Systems* (21:2), pp. 113–118.

Consalvo, M., and Ess, C. 2011. "Introduction: What is "Internet Studies"?" in *The Handbook of Internet Studies*, M. Consalvo and C. Ess (eds.), Oxford, UK: Wiley-Blackwell, pp. 1–8.

Dang, Y., Zhang, Y., Hu, P. J.-H., Brown, S. A., and Chen, H. 2011. "Knowledge mapping for rapidly evolving domains: A design science approach," *Decision Support Systems* (50:2), pp. 415–427.

Dang, Y., Zhang, Y., Hu, P. J.-H., Brown, S. A., Ku, Y., Wang, J.-H., and Chen, H. 2014. "An integrated framework for analyzing multilingual content in Web 2.0 social media," *Decision Support Systems* (61:0), pp. 126–135.

Davis, G. B. 2005. "Advising and Supervising," in *Research in Information Systems*: *A Handbook for Research Supervisors and Their Students*, D. E. Avison and J. Pries-Heje (eds.), Amsterdam, Boston, London: Elsevier/Butterworth-Heinemann, pp. 3–34.

Dellarocas, C. 2003. "The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms," *Management Science* (49:10), pp. 1407–1424.

di Gangi, P. M., Wasko, M. M., and Hooker, R. E. 2010. "Getting customers' ideas to work for you: Learning from Dell how to succeed with online user innovation communities," *MIS Quarterly Executive* (9:4), pp. 163–178.

Doan, A., Ramakrishnan, R., and Halevy, A. Y. 2011. "Crowdsourcing systems on the World-Wide Web," *Communications of the ACM* (54:4), p. 86.

Eisenhardt, K. M. 1989. "Building Theories from Case Study Research," *Academy of Management Review* (14:4), pp. 532–550.

Eisenhardt, K. M., and Graebner, M. E. 2007. "Theory building from cases: Opportunities and challenges," *Academy of Management Journal* (50:1), pp. 25–32.

Ekman, P. 1992. "An argument for basic emotions," *Cognition & Emotion* (6:3), pp. 169–200.

Few, S. 2006. *Information Dashboard Design*, North Sebastopol, CA, USA: O'Reilly.

Fliedl, G., Kop, C., and Vöhringer, J. 2010. "Guideline Based Evaluation and Verbalization of OWL Class and Property Labels," *Data & Knowledge Engineering* (69:4), pp. 331–342.

Gao, G., McCullough, J. S., Agarwal, R., and Jha, A. K. 2010. "Are doctors created equal? An investigation of online ratings by patients," in *Proceedings of the Workshop on Information Systems and Economics (WISE),* St. Louis, MI, pp. 1–6.

Gelman, I. A., and Wu, N. 2011. "Combining Structured and Unstructured Information Sources for a Study of Data Quality: A Case Study of Zillow.com," in *Proceedings of the 44th Hawaii International Conference on System Sciences (HICSS),* Kauai, HI, pp. 1–12.

Glass, R. L. 2000. "On design," *Journal of Systems and Software* (52:1), pp. 1–2.

Gregor, S., and Hevner, A. R. 2013. "Positioning and presenting design science research for maximum impact," *MIS Quarterly* (37:2).

Haenlein, M., and Kaplan, A. M. 2012. "The impact of unprofitable customer abandonment on current customers' exit, voice, and loyalty intentions: an empirical analysis," *Journal of Services Marketing* (26:6), pp. 458–470.

Hand, E. 2010. "Citizen Science: People power," *Nature* (466:7307), pp. 685–687.

Hevner, A. R., and Chatterjee, S. 2010. *Design research in information systems*: *Theory and practice*, New York, London: Springer.

Hevner, A. R., March, S. T., Park, J., and Ram, S. 2004. "Design science in information systems research," *MIS Quarterly* (28:1), pp. 75–105.

Hu, W., Almansoori, A., Kannan, P. K., Azarm, S., and Wang, Z. 2012. "Corporate dashboards for integrated business and engineering decisions in oil refineries: An agent-based approach," *Decision Support Systems* (52:3), pp. 729–741.

Huang, S., Ward, M. O., and Rundensteiner, E. A. 2005. "Exploration of Dimensionality Reduction for Text Visualization," in *Proceedings of the Third International Conference on Coordinated and Multiple Views in Exploratory Visualization*, Los Alamitos, CA, pp. 63–74.

Joachims, T. 1998. "Text Categorization with Suport Vector Machines: Learning with Many Relevant Features," in *Proceedings of the 10th European Conference on Machine Learning*, London, UK, UK: Springer-Verlag, pp. 137–142.

Jones, G. J. F. 2011. "Integrating social media with existing knowledge and information for crisis response.," in *The 3rd Workshop on Social Web Search and Mining (SWSM 2011): Analysis of User Generated Content Under Crisis at SIGIR 2011*.

Kaplan, A. M., and Haenlein, M. 2010. "Users of the world, unite! The challenges and opportunities of Social Media," *Business Horizons* (53:1), pp. 59–68.

Karpf, D. 2009. "Blogosphere Research: A Mixed-Methods Approach to Rapidly Changing Systems," *IEEE Intelligent Systems* (24:5), pp. 67–70.

Keim, D. A. 2002. "Information Visualization and Visual Data Mining," *IEEE Transactions on Visualization and Computer Graphics* (8:1), pp. 1–8.

Keller, K. L. 2014. *MSI 2014-2016 Research Priorities*. http://www.msi.org/uploads/files/MSI_RP14-16.pdf. Accessed 4 April 2015.

Kim, W., Jeong, O.-R., and Lee, S.-W. 2010. "On social Web sites," *Information Systems* (35:2), pp. 215–236.

Kornai, A. 2013. "Digital Language Death," *PloS one* (8:10), pp. 1–11.

Kuechler, W., and Vaishnavi, V. 2012. "A framework for theory development in design science research: Multiple perspectives," *Journal of the Association of Information Systems* (13:6), pp. 395–423.

Kvale, S., and Brinkmann, S. 2015. *InterViews*: *Learning the craft of qualitative research interviewing*, Thousand Oaks, CA: SAGE Publications.

Lau, R. Y. K., Liao, S. Y., Kwok, R. C.-W., Xu, K., Xia, Y., and Li, Y. 2011. "Text mining and probabilistic language modeling for online review spam detection," *ACM Transactions on Management Information Systems* (2:4), pp. 1–30.

Lee, A. S., and Baskerville, R. L. 2003. "Generalizing Generalizability in Information Systems Research," *Information Systems Research* (14:3), pp. 221–243.

Lee, H., and Choi, B. 2003. "Knowledge Management Enablers, Processes, and Organizational Performance: An Integrative View and Empirical Examination," *Journal of Management Information Systems* (20:1), pp. 179–228.

Lee, T. Y., and Bradlow, E. T. 2011. "Automated Marketing Research Using Online Customer Reviews," *Journal of Marketing Research (JMR)* (48:5), pp. 881–894.

Leukel, J., Müller, M., and Sugumaran, V. 2014. "The State of Design Science Research within the BISE Community: An Empirical Investigation," in *Proceedings of the 35th International Conference on Information Systems*, Auckland, New Zealand.

Lipenkova, J. 2015. "A system for fine-grained aspect-based sentiment analysis of Chinese," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics,* Beijing, China.

Lusch, R. F., Liu, Y., and Chen, Y. 2010. "The phase transition of markets and organizations: The new intelligence and entrepreneurial frontier," *IEEE Intelligent Systems* (25:1), pp. 71–75.

March, S. T., and Smith, G. F. 1995. "Design and natural science research on information technology," *Decision Support Systems* (15:4), pp. 251–266.

March, S. T., and Storey, V. C. 2008. "Design science in the information systems discipline: an introduction to the special issue on design science research," *MIS Quarterly* (32:4), pp. 725–730.

Marshall, B., McDonald, D., Chen, H., and Chung, W. 2004. "EBizPort: Collecting and analyzing business intelligence information," *Journal of the American Society for Information Science and Technology* (55:10), pp. 873–891.

Miles, M. B., and Huberman, A. M. 1994. *Qualitative Data Analysis*, Thousand Oaks, CA: SAGE Publications.

Miles, M. B., Huberman, A. M., and Saldaña, J. 2013. *Qualitative data analysis*: *A methods sourcebook*.

Miniwatts Marketing Group 2013. *Number of Internet Users by Language*: *Internet World Stats*.

Moody, D. L., and Shanks, G. G. 2003. "Improving the quality of data models: empirical validation of a quality management framework," *Information Systems* (28:6), pp. 619–650.

Neff, A. A., Hamel, F., Herz, T. P., Uebernickel, F., Brenner, W., and vom Brocke, J. 2014. "Developing a maturity model for service systems in heavy equipment manufacturing enterprises," *Information & Management* (51:7), pp. 895–911.

Netzer, O., Feldman, R., Goldenberg, J., and Fresko, M. 2012. "Mine Your Own Business: Market-Structure Surveillance Through Text Mining," *Marketing Science* (31:3), pp. 521–543.

Nicolai, A., and Seidl, D. 2010. "That's Relevant! Different Forms of Practical Relevance in Management Science," *Organization Studies* (31:9-10), pp. 1257–1285.

Nunamaker, J. F., and Briggs, R. O. 2011. "Toward a broader vision for Information Systems," *ACM Transactions on Management Information Systems* (2:4), pp. 1–12.

Nunamaker, J. F., Chen, M., and Purdin, T. D. M. 1990. "Systems Development in Information Systems Research," *Journal of Management Information Systems* (7:3), pp. 89–106.

O'Reilly, T. 2005. *What is Web 2.0? Design Patterns and Business Models for the Next Generation of Software*. http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/.

Österle, H., Becker, J., Frank, U., Hess, T., Karagiannis, D., Krcmar, H., Loos, P., Mertens, P., Oberweis, A., and Sinz, E. J. 2011. "Memorandum on design-oriented information systems research," *EJIS* (20:1), pp. 7–10.

Parment, A. 2014. *Auto brand*: *Building successful car brands for the future*, London, UK: Kogan Page.

Peffers, K. E., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. 2007. "A Design Science Research Methodology for Information Systems Research," *Journal of Management Information Systems* (24:3), pp. 45–77.

Qin, J., Zhou, Y., Chau, M., and Chen, H. 2006. "Multilingual Web retrieval: An experiment in English–Chinese business intelligence," *Journal of the American Society for Information Science and Technology* (57:5), pp. 671–683.

Recker, J. C., and Rosemann, M. 2010. "A measurement instrument for process modeling research: development, test and procedural model," *Scandinavian Journal of Information Systems* (22:2), pp. 3–30.

Rouibah, K., and Ould-ali, S. 2002. "PUZZLE: a concept and prototype for linking business intelligence to business strategy," *The Journal of Strategic Information Systems* (11:2), pp. 133–152.

Santiago Rivera, D., and Shanks, G. 2015. "A Dashboard to Support Management of Business Analytics Capabilities," *Journal of Decision Systems* (24:1), pp. 73–86.

Schmidt, M., and Hollensen, S. 2006. *Marketing research*: *An international approach*, Harlow, England, New York: Prentice Hall/Financial Times.

Schultze, U., and Leidner, D. E. 2002. "Studying Knowledge Management in Information Systems Research: Discourses and Theoretical Assumptions ," *MIS Quarterly* (26:3), pp. 213–242.

Scott Morton, M. S. 1984. "The State of the Art of Research in Management Support Systems," in *The Information Research Challenge*, F. W. McFarlan (ed.), Boston, MA: Harvard University Press, pp. 13–41.

Sha, S., Huang, T., and Gabardi, E. 2013. "Upward Mobility: The Future of China's Premium Car Markettr," McKinsey Automotive & Assembly Practice (ed.).

Shneiderman, B. 1996. "The eyes have it: a task by data type taxonomy for information visualizations," in *IEEE Symposium on Visual Languages*, pp. 336–343.

Simon, H. A. 1996. *The sciences of the artificial*, Cambridge, MA: MIT Press.

Sonnenberg, C., and vom Brocke, J. 2012a. "Evaluation Patterns for Design Science Research Artefacts," in *Practical Aspects of Design Science*, M. Helfert and B. Donnellan (eds.), Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 71–83.

Sonnenberg, C., and vom Brocke, J. 2012b. "Evaluations in the Science of the Artificial - Reconsidering the Build-Evaluate Pattern in Design Science Research," in *Design Science Research in Information Systems. Advance in Theory and Practice*, K. Peffers, M. Rothenberger and B. Kuechler (eds.), Las Vegas, NV: Springer, pp. 381–397.

STATISTA 2015. *Sale of Automobiles in China*. http://de.statista.com/statistik/daten/studie/215337/umfrage/autoabsatz-in-china/.

Stevens, C. H. 1981. *Many-To Many Communication*, Boston, MA: Center for Information Systems Research, Sloan School of Management, Massachusetts Institute of Technology.

Stevens, C. H. 2013. *Many-To Many Communication - Primary Source Edition:* Nabu Press.

Stindt, D., Nuss, C., Bensch, S., Dirr, M., and Tuma, A. 2014. "An Environmental Management Information System for Closing Knowledge Gaps in Corporate Sustainable Decision-Making," in *Proceedings of the 35th International Conference on Information Systems*, Auckland, New Zealand.

Straub, D., and Ang, S. 2011. "Editor's Comments: Rigor and Relevance in IS Research: Redefining the Debate and a Call for Future Research," *MIS Quarterly* (35:1), pp. iii–xi.

Talvensaari, T., Juhola, M., Laurikkala, J., and Järvelin, K. 2007. "Corpus-based cross-language information retrieval in retrieval of highly relevant documents," *Journal of the American Society for Information Science and Technology* (58:3), pp. 322–334.

Tremblay, M. C., Hevner, A. R., and Berndt, D. J. 2010. "Focus groups for artifact refinement and evaluation in design research," *Communications of the Association for Information Systems* (26:27), pp. 599–618.

van Aken, J. E. 2004. "Management Research Based on the Paradigm of the Design Sciences: The Quest for Field-Tested and Grounded Technological Rules," *Journal of Management Studies* (41:2), pp. 219–246.

Vulić, I., Smet, W. de, Tang, J., and Moens, M.-F. 2015. "Probabilistic topic modeling in multilingual settings: An overview of its methodology and applications," *Information Processing & Management* (51:1), pp. 111–147.

Wade, M., and Hulland, J. 2004. "The Resource-Based View and Information Systems Research: Review, Extension, and Suggestions for Future Research," *MIS Quarterly* (28:1), pp. 107–142.

Walls, J. G., Widmeyer, G. R., and El Sawy, Omar A. 1992. "Building an Information System Design Theory for Vigilant EIS," *Information Systems Research* (3:1), pp. 36–59.

Ware, C. 2000. *Information Visualization*, San Francisco, CA, USA: Morgan Kaufmann.

Watson, H. J., and Frolick, M. N. 1993. "Determining Information Requirements for an EIS," *MIS Quarterly* (17:3), pp. 255–269.

Wattal, S., Schuff, D., Mandviwalla, M., and Williams, C. B. 2010. "Web 2.0 and politics: The 2008 U.S. presidential election and an e-politics research agenda.," *MIS Quarterly* (34:4), pp. 669–688.

Webster, J., and Watson, R. T. 2002. "Analyzing the past to prepare for the future: Writing a literature review," *MIS Quarterly* (26:2), pp. 13–23.

Winograd, T. 1996. *Bringing design to software*, Reading, MA: Addison-Wesley.

Winograd, T. 1998. "The Design of Interaction," in *Beyond calculation: The next fifty years of computing*, P. J. Denning and R. M. Metcalfe (eds.), New York: Copernicus, pp. 149–162.

Wise, J. A. 1999. "The Ecological Approach to Text Visualization," *Journal of the American Society for Information Science* (50:13), pp. 1224–1233.

Yin, P., Wang, H., and Guo, K. 2013. "Feature–opinion pair identification of product reviews in Chinese: a domain ontology modeling method," *New Review of Hypermedia and Multimedia* (19:1), pp. 3–24.

Yin, R. K. 2003. *Case study research*: *Design and methods*, Beverly Hills, CA: Sage.

Zhan, J., Loh, H. T., and Liu, Y. 2009. "Gather customer concerns from online product reviews – A text summarization approach," *Expert Systems with Applications* (36:2, Part 1), pp. 2107–2115.

Zhou, Y., Qin, J., Chen, H., and Nunamaker, J. F. 2005. "Multilingual Web Retrieval: An Experiment on a Multilingual Business Intelligence Portal," in *38th Annual Hawaii International Conference on System Sciences*, Big Island, HI, USA.

Zikopoulos, P., Eaton, C., De Roos, D., Deutsch, T., and Lapis, G. 2012. *Understanding Big Data*: *Analytics for Enterprise Class Hadoop and Streaming Data*, New York, NY, USA: McGraw-Hill.

Zwass, V. 2010. "Co-Creation: Toward a Taxonomy and an Integrated Research Perspective," *International Journal of Electronic Commerce* (15:1), pp. 11–48.