

# Investigating the Impact of Corporate Social Responsibility on Firm's Short- and Long-Term Performance with Online Text Analytics

Heesung Lee

Center for Advanced Information  
Technology, Kyung Hee University,  
Seoul, Republic of Korea  
(obyung@gmail.com)

Yunseon Jin

Center for Advanced Information  
Technology, Kyung Hee University,  
Seoul, Republic of Korea  
(dudhrha@khu.ac.kr)

Ohbyung Kwon

School of Management,  
Kyung Hee University,  
Seoul, Republic of Korea  
(obkwon@khu.ac.kr)

.....

Despite expectations of short- or long-term positive effects of corporate social responsibility (CSR) on firm performance, the results of existing research into this relationship are inconsistent partly due to lack of clarity about subordinate CSR concepts. In this study, keywords related to CSR concepts are extracted from atypical sources, such as newspapers, using text mining techniques to examine the relationship between CSR and firm performance. The analysis is based on data from the New York Times, a major news publication, and Google Scholar. We used text analytics to process unstructured data collected from open online documents to explore the effects of CSR on short- and long-term firm performance. The results suggest that the CSR index computed using the proposed text - online media - analytics predicts long-term performance very well compared to short-term performance in the absence of any internal firm reports or CSR institute reports. Our study demonstrates the text analytics are useful for evaluating CSR performance with respect to convenience and cost effectiveness.

**Key Words** : Corporate Social Responsibility; Text Mining; Unstructured Data; Keyword Extraction; Multiple Regression; Firm Performance; Tobin's q

.....

Received : May 11, 2016    Revised : May 17, 2016    Accepted : May 19, 2016  
Publication Type : Regular Paper    Corresponding Author : Ohbyung Kwon

## 1. Introduction

Corporate social responsibility (CSR) is defined as a concept according to which companies voluntarily decide to contribute to the attainment of a better society and a cleaner environment (European Commission, 2001). CSR is important for business enterprises to improve sustainability

and increase long-term value. CSR reduces risk and increases competitiveness and profit; as a variable, it therefore has a positive relationship with firm performance (Margolis et al., 2011). Hence, many studies on CSR have demonstrated the positive impact of CSR on firm performance (Becker-Olsen et al., 2006).

Unfortunately, literature shows that there is a

disconformity in explaining the relationship between CSR and financial performance (Dixon-Fowler et al., 2013). According to Shareholder theory, the firm's CSR is regarded as a donation or showing friendliness from shareholders to stakeholders, which result in reduced firm performance especially in term of profitability. Some firms have a skeptical view of CSR, mainly due to the inadequacy of measures of the effects of CSR on firm performance (Costa and Menichini, 2013). This may be attributable to measurement problems associated with CSR performance and firm performance not included in the models used (Bouslah et al., 2016). Stakeholder theory suggested by Freeman (Freeman, 1984) addresses that firms should consider the interests of everyone who can substantially affect, or be affected by, the welfare of the firm. According to this viewpoint, CSR activity can be interpreted as means, not ends, to secure the shareholder's goal – his(her) company's profitability and competitiveness (Flammer, 2015). Likewise, theory of resource-based view of the firm stands for stakeholder theory in that the theory regards CSR activities and its outcomes as resource to achieve the firm's heterogeneity which results in competitive advantage (Porter and Kramer, 2006). Consequently, the contradictions in viewing the relationship between CSR performance and firm performance partially come from the theories that the researchers stand for. This implies that inter-theoretical approach will be potentially beneficial to better understand the contributions of CSR to firm performance.

In addition, literature also shows the inconsistency in measuring the firm's CSR performance, as well as firm performance. Measures of CSR performance are not identical to each other in the literature, rather the dimensionality of the concept of CSR performance (Rayton et al., 2015) and the diversity of the measures is high. One reason is the dexterity of the CSR performance: A firm addresses social problems either instrumental or moral rationale (Aguilera et al., 2007). Another reason is the diversity of the stakeholders. Corporate Social Performance (CSP) is primarily concerned with the management of firms' relationships with external stakeholders, such as consumers and local communities, and is strongly oriented to the external image and reputation of the organization (Brammer et al., 2007). Hence, CSR performance cannot be identical according to what the stakeholders are interested in. The next reason is that we have seen a number of new assessment standards for both qualitative and quantitative improvement of CSR performance, which reflect an increased interest in the environment and ethical management. Conventionally, Kinder, Lydenberg, and Domini (KLD) scores have been widely used to measure CSR performance. KLD uses qualitative data from surveys, financial statements, media reports, regulatory filings, and other sources to capture a firm's CSR activity in seven different categories (i.e., community, corporate governance, diversity, employee relations, environment, human rights, and product characteristics) (Harjoto et al., 2015). In addition, a lot of measures have been

proposed for CSR performance by the institutions and scholars other than KLD (Luo et al., 2015; Shahzad and Sharfman, 2015). For example, <Table 1> summarizes the various international organizations and NGOs and their CSR standards. However, the guidelines to measure CSR performance contain some limitations. First, there are cases where fairness and transparency are not guaranteed because of undisclosed criteria about the terms of the assessment and/or the audit. Second, assessment agencies may not be independent, which degrades reliability. Third, due to a flood of new CSR agencies and standards and the different weights of the indicators, reliable and consistent assessment is difficult (Aupperle et al., 1985; Barnett and Salomon, 2006). Next, assessment models have not truly reflected changing trends in industries or markets. Lastly, despite the availability of professional ratings on firm CSP dimensions, such information is too intricate to be directly understood and priced by general investors who are not certified industry experts and are often constrained by time and resources (Fombrun et al., 2000; Goyal et al., 2013; Surrice et al., 2010). Therefore, development of a consistent, trustworthy assessment tool using objective data, agreed-upon rating features or concepts, and timely reflections of changing trends is necessary.

In sum, CSR assessment has historically been an arduous task (Lindgreen and Swaen, 2010). Most data related to CSR performance are internal and difficult to access without voluntary disclosure or legal force (Dhaliwal et al., 2011). CSR

assessments can be fair, objective, and valid when data is provided by third-party sources, such as articles and papers that consider the footprint of corporate activities. Furthermore, data from the Internet is constantly available in enormous quantities and always changing, which facilitates timely assessment.

<Table 1> CSR Standards

Title	Organization	Features
AA1000	Account Ability 1999	Establishing healthy communication among stakeholders and improving explanations of the social ethical aspects of running a business
SA8000	Social Accountability International 1997	Third-party authentication of improvements in working conditions
SD21000	Association Française de Normalisation 2003	Introduces management strategies and strategic CSR to corporate executives
AS8003-2003	Standards Australia 2003	Supports establishment of effective CSR programs for corporations
GRI	Global Reporting Initiative 2000	Offers guidelines to present information about the economic, social, and environmental aspects of CSR
SIGMA	British Standards et al. 2003	Unifies existing standards and guidelines as a CSR management system

Recently, an infrastructure has been developed that includes big data provided by companies; this data is open and suitable for tracking, gathering, and analyzing trends in real time. Social media is

another especially powerful new marketing tool. Existing research results show that customer awareness as measured by advertising expenditure affects the relationship between CSR activity and firm value (Servaes and Tamayo, 2013), and the frequency of exposure of specific firms can be used as a proxy for marketing cost. All these are readily available sources of data. Open big data may potentially shed light on the relationship between CSR and firm performance using these new metrics. The advantage of applying open big data in constructing a CSR measurement model is that the method will improve the quality of CSR measurement and our ability to understand this concept. We demonstrate that incorporating all factors related to CSR from open big data can be done in an automated, timely, and cost-effective manner.

Hence, in this study, we propose a method of dealing with CSR assessment using atypical sources and focusing on the operational and financial usefulness of keywords in a big data which characterize the CSR performance and its relationship between firm performance in a sample of listed firms.

Our study makes several contributions to the ongoing debate about the impact of CSR on firm performance. First, we use open, unstructured data and text analytics to examine the impact of CSR on short- and long-term performance. Although extant literature (Deng et al., 2013) examines the impact of CSR on firm performance, the analysis most frequently focuses on the effect of target CSR on acquirer shareholder wealth. To the best

of our knowledge, this is the first empirical study to consider open big data as a means to develop a CSR measurement model, rather than performing content analysis which is very labor intensive and costly (Luo et al., 2015), and to verify the validity of the proposed model using a complete set of firm data, rather than using Kinder, Lydenberg, and Domini (KLD) scores which has been widely used but are biased due to incorporating a small set of firms (Harjoto et al., 2015). Evaluating CSR using text analytics complements the methods in previous research on the effects of CSR on firm performance; we view CSR evaluations by the market as an important determinant of long-term firm performance (Carroll and Shabana, 2010; Jensen, 2001). The results of our analysis suggest that the proposed text analytics method may be more effective for evaluating CSR in terms of cost and agility compared to external methods or self-reported evaluations. Thus, we examine the causality between CSR and firm value while simultaneously introducing a new line of research.

The paper proceeds as follows. Section 2 describes the theoretical background to the problem of evaluating the effects of CSR. Section 3 presents the overall process involved in text analytics-based CSR evaluations, and section 4 describes the data and provides the statistics for the variables of interest. In section 5, we discuss the main results of the empirical analysis, with robustness of test results and conclude the paper.

## 2. Methods

### 2.1 Overall Process

<Figure 1> shows the overall process necessary to establish a CSR document using atypical data available online and then evaluating each firm's CSR Corpora, which indicates a large and structured set of texts (Choi et al., 2015). The first step is to collect data related to a firm's CSR available online and establish the CSR Corpora. The second is to reprocess the CSR Corpora, extract terms from it, and then build its vocabulary. The third is to compute the polarity value of the words in the CSR Vocabulary. The fourth is to add synonyms to the CSR Vocabulary to expand it further. The fifth is to assess the firm's CSR activity using the extended CSR Vocabulary by collecting content related to the firm online. Appendix A provides more details related to the overall process.

- [1] Get CSR Corpora from accessible and downloadable online news and articles
- [2] Extract terms from CSR Corpora and then build CSR Vocabulary
- [3] Distinguish polarity of words in CSR Vocabulary
- [4] Add similar words to CSR Vocabulary
- [5] Assess CSR activity with CSR Vocabulary

<Figure 1> Overall process

### 2.2 CSR Vocabulary and Polarity

The CSR Vocabulary is built based on the CSR Corpora created from online information sources. To obtain CSR concepts and features without relying on existing CSR standards or assessment models, we collected documents containing the terms "CSR" and "Corporate Social Responsibility" from online news sites. In total, 484 articles were collected in 2015 from the New York Times using a Python web crawler program. In total, 71 scholarly articles about CSR and firm performance were collected from Google Scholar from 1975 to 2014. The CSR Corpora were then created using data extracted from the two sources.

In total, 17,311 terms were extracted including verbs, adjectives, and nouns from the CSR Corpora using a program with related APIs. Among them, pronouns, numerals, numbers, prepositions, and other symbols were excluded in addition to stop words such as "said," "go," and "even". Finally, we identify 14,817 terms as valid words.

For each valid word extracted during this preparatory stage, the frequency ratio was calculated (the frequency of the word / the number of valid words except those in the stop words list) compared to default business news articles in the New York Times. We then removed words with no significant difference from the ratio of CSR-related frequency (the frequency of CSR words / the number of valid words except those in the stop words list). Then we dropped the words of which ratio is under 0.0076 according to 20/80

rules. In other words, 0.076 indicates the ratio of the top 20% frequently appeared word (See Table 2), As a result, we identified CSR Vocabulary from the CSR Corpora.

The words in the CSR Corpora have three vector values: stem word, ratio, and polarity. One of the stem words, for example, “universe,” has derivatives such as “universal”. The stem words may have polarity (-1: negative, 1: positive) according to the information in the SentiWordNet, which is one of the most widely used dictionaries for sentiment analysis (Yu et al., 2013; Kim and Jeong, 2013). For example, the polarity of the words “sustain” and “organic” are positive, while that of “illegal” and “illness” are negative. This process yielded a final CSR vocabulary of 426 words, of which 76 were negative and the remaining 350 were positive. Lastly, we improved word variety by analyzing similar words, yielding the 504 CSR Vocabularies illustrated in <Table 2>, which contain 103 negative and 401 positive words.

To evaluate firms’ CSR scores using CSR Vocabulary, we extracted word frequencies from the articles and sentences that include the name of the firm, including sentences indicating a specific firm, such as, “the company.” When negative indicators such as “~don’t think” or “not,” “no,” and “never,” appear, the polarity of the word is inverted.

We calculated the CSR scores for each firm and then obtained the correlation between a firm’s evaluated CSR score, present and future performance (ROE, ROA, ROI, Tobin’s q, and

sales), and competency variables. We then sorted the Vocabularies according to the variables present, future, and competency.

Finally, we filtered the vocabularies to exclude words that do not apparently belong to the CSR category and modified related Vocabularies based on this sorting. Simultaneously, we collected Vocabularies that were apparently included in the CSR category separately.

<Table 2> CSR Vocabulary Examples

Source	Word	Ratio	Polarity (-1 or 1)
Scholarly	Social	0.3381	1
Scholarly	Respons	0.204	1
Scholarly	Relationship	0.0797	1
Scholarly	Value	0.0755	1
News	Good	0.0751	1
Scholarly	CSR	0.0709	1
--- omitted ---			
Scholarly	Difficult	0.0081	-1
News	Holt	0.008	1
News	Poverti	0.0079	1
News	Danger	0.0078	-1
News	Pollut	0.0077	-1
News	Afford	0.0076	1
--- omitted ---			

### 3. Results

#### 3.1 Procedure

We used two comparison targets to determine the performance of the proposed CSR Index

computation. The comparison targets are as follows:

Method 1: Firm performance estimation based on a CSR Index published by a CSR evaluation institute. We used evaluation data published by the rating agency CSRHub (<http://csrtoday.co.kr/news/article.html?no=392>), because CSRHub is the world's largest and most comprehensive databases of CSR and sustainability-related information. CSRHub provides access to corporate social responsibility and sustainability ratings on 15,571+ companies from 134 industries in 132 countries.

Method 2: Firm performance estimation using the frequency of appearance in business articles.

Method 3: Firm performance estimation using CSR Index scores obtained from the proposed method in this study.

We used the FnGuide to obtain performance data for firms listed in the KOSPI. The performance indices include each company's present value in terms of sales, market capitalization, reputation, and Tobin's  $q$  to represent the company's future value and growth potential. The experiment was conducted on 233 listed firms whose financial data are published extensively online.

### 3.2 Assessment Using CSR Vocabulary

In the first step, "firm social responsibility" and target company names were chosen as query terms to determine the degree of CSR of KOSPI-listed companies (e.g., "firm social responsibility" Samsung Electronics). Company names were based

on the names used in the KOSPI listing, but Korean-English transfer was allowed in consideration of the number of search queries. Based on each query, we selected URLs for news articles posted on NAVER from January 1, 2013 to December 31, 2013. We extracted 4,496 news articles on 773 companies.

In the second step, we converted the CSR Vocabulary list selected through the New York Times to Korean to use it in the analysis of KOSPI-listed companies. For converted words without a stem form, a stemming process was conducted. Additionally, we processed news articles for tokenizing and stemming to ensure program compatibility with Korean CSR Corpora and news articles. For example, the Korean word 'nanhae' obtained from "difficult" should not be confused with 'gi-nanhae', which means "last year" in English. Lastly, we examined the tokens of extracted news articles included in the Korean CSR Corpora list. We then added them to the list considering the polarity given to each token. In this case, polarity aggregation was done independently within each article. We did not allow for duplicate words within a single article.

### 3.3 Assessment Using CSR Vocabulary

CSR indices by industry obtained using the method proposed are used in this study. The result suggests that CSR indices for the paper, machinery, medical supplies, gas and electricity, steel, and metal industries are relatively high, whereas those for the textile, apparel, construction,

distribution, and medical precision industries are relatively low. The paper industry is closely associated with carbon dioxide emission and global warming; thus, CSR activities related to environmental preservation are important to the success of firms in this industry (Amran et al., 2014; Tuominen et al., 2008). The pharmaceutical and medical precision industries have legal limitations due to ethical issues (Carroll, 2000), so companies must adopt high levels of social responsibility, which is why their CSR indices are high. The textile and apparel industries are labor-intensive, so they must process products using raw materials imported from developing countries. These industries also make great efforts toward CSR because workers' welfare is one of the major issues (Blowfield and Frynas, 2005). Thus, we show that the proposed CSR Index appropriately reflects the characteristics of many industries.

We also obtained the relationships among present value (sales), market capitalization, future firm value (Tobin's q, TQ), the CSRHub's CSR index (CSRHub), the number of URLs calculated from analysis of atypical data for a specific period, and the CSR score computed by the CSR Index proposed in this study (CSR, CSR2). Among those, sales, market capitalization and TQ are used as dependent variables, respectively. CSRHub, Number of URLs, CSR and CSR2 are considered as independent variables. A multiple regression analysis was conducted to determine the exact relationships among them.

The results of the analysis using the same

independent variables against sales, market capitalization, and TQ are as follows. We conducted the multiple regression analysis using the R-lm library. As seen in <Table 3>, the results show that values for all four variables measuring CSR levels are not statistically significant in terms of sales and market capitalization. However, the CSR value of CSRHub does not affect TQ, which indicates future value, though this study's proposed CSR Index does affect TQ (at the .05 significance level). Therefore, CSR activity levels do not have a major impact on a firm's short-term performance, but they do affect the firm's future value. This result is in line with that from previous studies (Carroll and Shabana, 2010; Jensen, 2001). Interestingly, the CSR2 Index, which was created as an alternative to the CSR Index, shows no significant effect. Considering the adjusted R-squared value, we conclude that Model 3 is the most explanatory.

<Table 3> MLEs for Each Model

Variable	Model 1	Model 2	Model 3
Dependent variable	Sales	Market capitalization	TQ
(Intercept)	8.339*	28.677***	-2.091
CSRHub	-.011	.032	.010
Number of URLs	-.004	.002	-.004
CSR	-17.916	-5.751	44.341**
CSR2	-.005	-.082	-.080
Adjusted R-squared	.011	.035	.114

Note) \* p<0.1, \*\* p<0.05, \*\*\* p<0.01



We then conducted the following regression analysis to compare methods 1 and 3, which are the CSR indices provided by other rating agencies, such as CSRHub, and the CSR Index computed in this study. Therefore, we established a new model, Model 4, by considering TQ as dependent variable, and CSRHub and CSR as independent variables. The results in <Table 4> indicate that the results using the CSR Index are statistically significant, while those using the CSRHub are not statistically significant, indicating that the CSR computation method presented in this study has more explanatory power for TQ.

<Table 4> MLE for Model 4

Variable	Model 4
Dependent variable	TQ
(Intercept)	-.652
CSRHub	.007
CSR	32.292**
Adjusted R-squared	.083

Note) \* p<0.1, \*\* p<0.05, \*\*\* p<0.01

We also conducted an ANOVA analysis to determine how the two CSR Index variables and the number of URLs, both of which are included in the optimal model, affect the full and reduced

models. The full model presented above is as follows:

- Full Model:  $TQ \sim CSR + \text{Number of URLs} + \text{Number of employees} + \text{Current liability} + \text{Current assets} + \text{Sales} + \text{Net profit} + \text{Market capitalization}$

By excluding the CSR Index and the number of URLs from the full model, we created the following reduced model:

- Reduced Model:  $TQ \sim \text{Number of employees} + \text{Current liability} + \text{Current assets} + \text{Sales} + \text{Net profit} + \text{Market capitalization}$

<Table 5> summarizes the ANOVA results. The full model is more statistically significant than the reduced model, indicating that the CSR Index and number of URLs proposed in this study are statistically significant.

## 4. Discussion

### 4.1 Academic Implications

This study has the following implications for academia. First, we developed a CSR Index to evaluate corporate social responsibility using

<Table 5> ANOVA Test for Full and Reduced Models

Model	Df	RSS	Df	Sum of Sq	F	Pr(>F)
Full Model	224	109.96				
Reduced Model	226	113.01	-2	-3.0517	3.1082	0.04661 **

Note) \* p<0.1, \*\* p<0.05, \*\*\* p<0.01

text-mining techniques from atypical sources available online, enabling a timely assessment of CSR at any time. This is more time- and cost-efficient compared to self-reporting or other assessment methods used by professionals. Additionally, if online data provides an objective representation of a company's footprint, more objective results can be obtained than with the self-reporting method (Fombrun et al., 2000; Surrica et al., 2010). In addition, companies can determine the effects of their policies or marketing strategies related to social responsibility using the CSR Index techniques described here. In sum, a company's CSR assessment using data analysis from atypical sources can be an alternative to CSR assessment by a rating agency.

Second, the results of this study suggest that our CSR Index can be used to analyze the effects of CSR on firm performance alongside other performance-related information, such as sales and Tobin's *q*. In particular, the CSR Index developed in this study showed no statistically significant relationship between short-term corporate performance (i.e., sales and market capitalization) and CSR. Instead, the results for Tobin's *q*, which predicts a firm's future performance, were statistically significant. In other words, CSR activities have little effect on short-term performance, but this empirical study demonstrated that CSR increases corporate value in the long term. This discrepancy may be one of the causes of the contradictory results in previous studies regarding the effects of CSR on firm performance (Cheng et al., 2014).

Previous work established that CSR assessment has a positive effect on long-term performance rather than short-term performance (Carroll and Shabana, 2010; Jensen, 2001). CSR activities may not directly result in improved management performance (Cochrane, 1991), but they may enhance satisfaction and company pride for many stakeholders, such as employees, leading to better long-term performance (Romer, 1986; White, 2006). They may also provide a good impression for acquiring firms' shareholders during M&A and have a positive influence on long-term stock returns (Deng et al., 2013). From another perspective, if CSR is modified based on firm strategy, it can affect profit (Anderson and Bieniaszewska, 2005). However, this is the first study to reveal the positive causal relationship between CSR performance and long-term performance through a big data analysis.

Third, the CSR Index developed in this study explained Tobin's *q*, a measure of future firm performance, more significantly than other assessments, such as those from other CSR ratings agencies. Because CSR is one factor that can affect long-term rather than short-term performance (Carroll and Shabana, 2010; Jensen, 2001), CSR activity may be associated with the maturity of the corporate environment (Baumgartner and Ebner, 2010). In other words, the notion of a CSR-business fit pertains here (Du et al., 2010). In developed countries in recent decades, firms have realized the link between being socially active and creating profit, reporting the results of their CSR activities publicly in order to meet higher

stakeholder expectations (DeGeorge, 1995; Hardjono and Marrewijk, 2001).

Finally, while previous papers typically examine the correlation between a firm's CSR performance and its value (Carroll and Shabana, 2010; Jensen, 2001), our study investigates the causal link between these two constructs by explicitly controlling for potential biases associated with the endogenous nature of social performance, thereby identifying a clear channel through which CSR contributes to shareholder wealth.

#### 4.2 Practical Implications

This study has the following practical implications. First, we developed a CSR Index to evaluate corporate social responsibility using text-mining techniques from available atypical sources. This study demonstrates that a company can self-evaluate its CSR performance by using unstructured public data in combination with the proposed text mining method; this is true even for companies that are not supported by professional CSR consulting institutes. Adoption of big data analysis practices is useful to periodically monitor the opinions of stakeholders and predict the company's future value by estimating Tobin's  $q$  based on the values determined by our CSR Index. The big data analysis practices described here include crawling through environmental data generated on online news or social network sites and blogs. Our method is more useful for companies to conduct agile CSR evaluations that may influence profitability management rather than

just relying on the CSR review report periodically provided by the CSR institutes, which is very costly (Márquez and Fombrun, 2005).

Second, companies may make use of the proposed method to avoid concerns about the differences in CSR evaluation results due to the heterogeneity and subjectivity of the evaluating institutes (Campbell, 2007; Carroll and Shabana, 2010; Garriga and Melé, 2013; Pelozo and Shang, 2011). Despite the technical difficulties of analyzing unstructured documents (Harrison, 2006), public data reflects the footprint of the company's CSR efforts and how stakeholders actually have perceived those efforts (Allern, 2002; Bednarek and Caple, 2014). Moreover, companies must regard online news and SNS as crucial sources of evidence to demonstrate their CSR activities. In the literature, online news and SNS have become important to building the company image and reputation (Carroll and McCombs, 2003; Einwiller et al., 2010; Fombrun and Shanley, 1990), which may affect the company's future value (Bernes et al., 2005; Fombrun et al., 2000; Lai et al., 2010; Lii and Lee, 2012; Luo and Bhattacharya, 2006; Nan and Heo, 2007) and profitability (Du et al., 2010). Finally, the Index developed in this study provides companies with consistent results that can be used to compare CSR from McCombs year to year.

It is important to note that CSR Vocabulary does not need to be constructed from scratch. In this study, 504 CSR Vocabularies containing 103 negative and 401 positive words were generated using the CSR Index. In addition, this study also

provides the pseudocode for the proposed procedure. These will decrease the time and cost of performing data analytics for CSR performance evaluation.

### 4.3 Limitations

This study has a few limitations that point to several future areas of study. First, CSR assessment was performed using atypical data in order to develop a CSR Index. However, a sensitivity analysis would have enabled us to identify factors that affect the results of the CSR Index through CSR Vocabulary clustering. It would then be possible to prioritize where firms should concentrate their efforts to improve CSR performance in the future.

Second, to demonstrate the superiority of the CSR Index developed in this study, we compared it with CSRHub assessment. However, generalizability would be improved by comparing the proposed Index to a variety of results from various ratings agencies. In fact, it is difficult to access CSR assessment data for all listed companies in Korea because assessment results and other rating agencies' methods are protected as a core asset. We also could not use the KLD Research & Analytics, Inc. STATS database or other sources of data because most of the available data concerns global companies, of which Korean companies are only a part. CSRHub publishes only data pertaining to Korean listed companies publicly, so future research could include assessment results from other ratings agencies.

Third, the CSR Index Vocabulary is in English. When we evaluated KOSPI-listed companies, we used the NAVER representative Korean-English dictionary to translate the English vocabulary into Korean. However, this process could introduce subtle changes in meaning. Business articles related to CSR assessment do not have literary terms and are unambiguous, mostly fact-oriented, and objective. While this is unlikely to distort the assessment results, there may be some slight differences in interpretation.

## 5. Conclusions

In this study, we used text analytics to process unstructured data collected from open online documents to explore the effects of CSR on short- and long-term firm performance. The CSR Index computed using text analytics predicts long-term performance well compared to short-term performance in the absence of any internal firm reports or CSR institute reports. The positive relationship between an acquirer's CSR performance and its long-term financial performance is also evident when we capture CSR performance using an indicator for inclusion on the list of companies with good CSR, such as the FTSE4Good Index and CSRHub. Incorporating text analytics in a CSR evaluation provided better results than previous evaluation methods in terms of data processing costs, relevance, and agility. Firms and CSR evaluation institutes can make use of this Index to improve their CSR evaluation

methods. Overall, these results suggest that text analytics is a promising method for use in the CSR research community.

## References

- Aguilera, R. V., D. E. Rupp, C. A. Williams, and J. Ganapathi, "Putting the S Back in Corporate Social Responsibility: A Multilevel Theory of Social Change in Organizations," *Academy of Management Review*, Vol.32, No.3(2007), 836-863.
- Allem, S., "Journalistic and Commercial News Values," *Nordicom Review*, Vol.23, No.1-2 (2002), 137-152.
- Amran, A., S. P. Lee, and S. S. Devi, "The Influence of Governance Structure and Strategic Corporate Social Responsibility Toward Sustainability Reporting Quality," *Business Strategy and The Environment*, Vol.23, No.4(2014), 217-235.
- Anderson, C. L. and R. L. Bieniaszewska, "The Role of Corporate Social Responsibility in an Oil Company's Expansion into New Territories," *Corporate Social Responsibility and Environmental Management*, Vol.12, No.1(2005), 1-9.
- Aupperle, K. E., A. B. Carroll, and J. D. Hatfield, "An Empirical Examination of the Relationship Between Corporate Social Responsibility and Profitability," *Academy of management Journal*, Vol.28, No.2(1985), 446-463.
- Barnett, M. L. and R. M. Salomon, "Beyond Dichotomy: The Curvilinear Relationship Between Social Responsibility and Financial Performance," *Strategic Management Journal*, Vol.27, No.11 (2006), 1101-1122.
- Baumgartner, R. J. and D. Ebner, "Corporate Sustainability Strategies: Sustainability Profiles and Maturity Levels," *Sustainable Development*, Vol.18, No.2(2010), 76-89.
- Becker-Olsen, K. L., B. A. Cudmore, and R. P. Hill, "The Impact of Perceived Corporate Social Responsibility on Consumer Behavior," *Journal of Business Research*, Vol.59, No.1(2006), 46-53.
- Bednarek, M. and H. Caple, "Why Do News Values Matter? Towards a New Methodological Framework for Analysing News Discourse in Critical Discourse Analysis and Beyond," *Discourse & Society*, Vol.25, No.2(2014), 135-158.
- Berens, G., C. B. V. Riel, and G. H. V. Bruggen, "Corporate Associations and Consumer Product Responses: The Moderating Role of Corporate Brand Dominance," *Journal of Marketing*, Vol.69, No.3(2005), 35-48.
- Blowfield, M. and J. G. Frynas, "Editorial Setting New Agendas: Critical Perspectives on Corporate Social Responsibility in the Developing World," *International Affairs*, Vol.81, No.3(2005), 499-513.
- Bouslah, K., L. Kryzanowski, and B. M'Zali, "Social Performance and Firm Risk: Impact of the Financial Crisis," *Journal of Business Ethics*, (2016), 1-27.
- Brammer, S., A. Millington, and B. Rayton, "The Contribution of Corporate Social Responsibility to Organizational Commitment," *The International Journal of Human Resource Management*, Vol.18, No.10(2007), 1701-1719.
- Campbell, J. L., "Why Would Corporations

- Behave in Socially Responsible Ways? An Institutional Theory of Corporate Social Responsibility,” *Academy of Management Review*, Vol.32, No.3(2007), 946-967.
- Carroll, A. B. and K. M. Shabana, “The Business Case for Corporate Social Responsibility: A Review of Concepts, Research and Practice,” *International Journal of Management Reviews*, Vol.12, No.1(2010), 85-105.
- Carroll, A. B., “Ethical Challenges for Business in the New Millennium: Corporate Social Responsibility and Models of Management Morality,” *Business Ethics Quarterly* (2000), 33-42.
- Carroll, C. E. and M. McCombs, “Agenda-Setting Effects of Business News on the Public’s Images and Opinions about Major Corporations,” *Corporate Reputation Review*, Vol.6, No.1(2003), 36-46.
- Cheng, B., I. Ioannou, and G. Serafeim, “Corporate Social Responsibility and Access to Finance,” *Strategic Management Journal*, Vol.35, No.1(2014), 1-23.
- Choi, S., Jeon, J., Subrata, B. and O. Kwon, “An Efficient Estimation of Place Brand Image Power Based on Text Mining Technology,” *Journal of Intelligence and Information Systems*, Vol.21, No.2(2015), 113-129.
- Cochrane, J. H., “Production-based Asset Pricing and the Link Between Stock Returns and Economic Fluctuations,” *The Journal of Finance*, Vol.46, No.1(1991), 209-237.
- Costa, R. and T. Menichini, “A Multidimensional Approach for CSR Assessment: The Importance of the Stakeholder Perception,” *Expert Systems with Applications*, Vol.40, No.1(2013), 150-161.
- DeGeorge, R. T, *Business Ethics*, 4th Edition, Prentice-Hall: New York, 1995.
- Deng, X., J. K. Kang, and B. S. Low, “Corporate Social Responsibility and Stakeholder Value Maximization: Evidence from Mergers,” *Journal of Financial Economics*, Vol.110, No.1(2013), 87-109.
- Dhaliwal, D. S., O. Z. Li, A. Tsang, and Y. G. Yang, “Voluntary Nonfinancial Disclosure and the Cost of Equity Capital: The Initiation of Corporate Social Responsibility Reporting,” *The Accounting Review*, Vol.86, No.1(2011), 59-100.
- Dixon-Fowler, H. R., D. J. Slater, J. L. Johnson, A. E. Ellstrand, and A. M. Romi, “Beyond ‘Does it Pay to Be Green?’ A Meta-Analysis of Moderators of the CEP-CFP Relationship,” *Journal of Business Ethics*, Vol.112, No.2(2013), 353-366.
- Du, S., C. B. Bhattacharya, and S. Sen, “Maximizing Business Returns to Corporate Social Responsibility (CSR): The Role of CSR Communication,” *International Journal of Management Reviews*, Vol.12, No.1(2010), 8-19.
- Einwiller, S. A., C. E. Carroll, and K. Korn, “Under What Conditions Do the News Media Influence Corporate Reputation? The Roles of Media Dependency and Need for Orientation,” *Corporate Reputation Review*, Vol.12, No.4(2010), 299-315.
- European Commission, *Green Paper: Promoting a European Framework for Corporate Social Responsibility*, Office for Official Publications of the European Communities, Brussels: COM. 366, 2001.
- Flammer, C., “Does Corporate Social Responsibility

- Lead to Superior Financial Performance? A Regression Discontinuity Approach,” *Management Science*, Vol.61, No.11(2015), 2549-2568.
- Fombrun, C. J. and M. Shanley, “What’s in a Name? Reputation Building and Corporate Strategy,” *Academy of Management Journal*, Vol.33, No.2(1990), 233-258.
- Fombrun, C. J., N. A. Gardberg, and M. L. Barnett, “Opportunity Platforms and Safety Nets: Corporate Citizenship and Reputational Risk,” *Business and Society Review*, Vol.105, No.1(2000), 85-106.
- Freeman, R. E., *Strategic Management: A Stakeholder Approach*, Pitman, Boston, 1984.
- Garriga, E. and D. Melé, “Corporate Social Responsibility Theories: Mapping the Territory,” *In Citation Classics from the Journal of Business Ethics*, Springer Netherlands(2013), 69-96.
- Goyal, P., Z. Rahman, and A. A. Kazmi, “Corporate Sustainability Performance and Firm Performance Research: Literature Review and Future Research Agenda,” *Management Decision*, Vol.51, No.2(2013), 361-379.
- Hardjono, T. W. and M. V. Marrewijk, “The Social Dimensions of Business Excellence,” *Corporate Environmental Strategy*, Vol.8, No.3(2001), 223-233.
- Harjoto, M., I. Laksmana, and R. Lee, “Board Diversity and Corporate Social Responsibility,” *Journal of Business Ethics*, Vol.132, No.4(2015), 641-660.
- Harrison, J., *News*, London: Routledge, 2006.
- Jensen, M. C. and W. H. Meckling, “Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure,” *Journal of Financial Economics*, Vol.3, No.4(1976), 305-360.
- Jensen, M. C., “Value Maximization, Stakeholder Theory, and the Corporate Objective Function,” *Journal of Applied Corporate Finance*, Vol.14, No.3(2001), 8-21.
- Kim, Y. and S.R. Jeong, “Intelligent VOC Analyzing System Using Opinion Mining,” *Journal of Intelligence and Information Systems*, Vol.19, No.3(2013), 113-125.
- Lai, C. S., C. J. Chiu, C. F. Yang, and D. C. Pai, “The Effects of Corporate Social Responsibility on Brand Performance: The Mediating Effect of Industrial Brand Equity and Corporate Reputation,” *Journal of Business Ethics*, Vol.95, No.3(2010), 457-469.
- Lii, Y. S. and M. Lee, “Doing Right Leads to Doing Well: When the Type of CSR and Reputation Interact to Affect Consumer Evaluations of the Firm,” *Journal of Business Ethics*, Vol.105, No.1(2012), 69-81.
- Lindgreen, A. and V. Swaen, “Corporate Social Responsibility,” *International Journal of Management Reviews*, Vol.12, No.1(2010), 1-7.
- Luo, X. and C. B. Bhattacharya, “Corporate Social Responsibility, Customer Satisfaction, and Market value,” *Journal of Marketing*, Vol.70, No.4(2006), 1-18.
- Luo, X., H. Wang, S. Raithel, and Q. Zheng, “Corporate social Performance, Analyst Stock Recommendations, and Firm Future Returns,” *Strategic Management Journal*, Vol.36, No.1 (2015), 123-136.

- Margolis, J. D., H. A. Elfenbein, and J. P. Walsh, "Does It Pay to Be Good... and Does It Matter? A Meta-Analysis of the Relationship Between Corporate Social and Financial Performance," *SSRN Working Paper Series*, 2011.
- Márquez, A. and C. J. Fombrun, "Measuring Corporate Social Responsibility," *Corporate Reputation Review*, Vol.7, No.4(2005), 304-308.
- Nan, X. and K. Heo, "Consumer Responses to Corporate Social Responsibility (CSR) Initiatives: Examining the Role of Brand-Cause Fit in Cause-Related Marketing," *Journal of Advertising*, Vol.36, No.2(2007), 63-74.
- Pelozo, J. and J. Shang, "How Can Corporate Social Responsibility Activities Create Value for Stakeholders? A systematic review," *Journal of the Academy of Marketing Science*, Vol.39, No.1(2011), 117-135.
- Porter, M. E. and M. R. Kramer, "Strategy and Society: The Link Between Competitive Advantage and Corporate Social Responsibility," *Harvard Business Review*, Vol.84, No.12(2006), 78-92.
- Rayton, B. A., S. J. Brammer, and A. I. Millington, "Corporate Social Performance and the Psychological Contract," *Group & Organization Management*, Vol.40, No.3(2015), 353-377.
- Romer, P. M., "Increasing Returns and Long-Run Growth," *The Journal of Political Economy*, Vol.94, No.5(1986), 1002-1037.
- Servaes, H. and A. Tamayo, "The Impact of Corporate Social Responsibility on Firm Value: The Role of Customer Awareness," *Management Science*, Vol.59, No.5(2013), 1045-1061.
- Shahzad A. M, and M. P. Sharfman, "Corporate Social Performance and Financial Performance Sample-Selection Issues," *Business and Society*, 2015.
- Surroca, J., J. A. Tribó, and S. Waddock, "Corporate Responsibility and Financial Performance: The Role of Intangible Resources," *Strategic Management Journal*, Vol.31, No.5(2010), 463-490.
- Tuominen, P., T. Uski, I. Jussila, and U. Kotonen, "Organization Types and Corporate Social Responsibility Reporting in Finnish Forest Industry," *Social Responsibility Journal*, Vol.4, No.4(2008), 474-490.
- White, A. L., *Business Brief: Intangibles and CSR*, 2006.
- Yu, E., Kim, Y., Kim, N., and S.R. Jeong, (2013). "Predicting the Direction of the Stock Index by Using a Domain-Specific Sentiment Dictionary," *Journal of Intelligence and Information Systems*, Vol.19, No.1(2013), 95-110.



## Appendix A.

### Get CSR Corpora

Get (*news articles*) from online news website (e.g., *NY Times*) [keyword: CSR]  
 Get (*scholarly articles*) from academic website (e.g., Google scholar) [keyword: CSR and firm performance]  
 Get (*comparison news articles*) from online news website (e.g., *NY Times*) [keyword: general business]

### Construct CSR Vocabulary

Get (*text*) from CSR Corpora  
**For every** (*text*) in (*news articles, scholarly articles, and comparison news articles*) **then**,  
 Tokenize, remove stopwords, and do stemming  
     Get (*vocabulary, frequency, and sum of valid vocabularies*)  
**If** ( $(\text{scholarly articles-frequence} / \text{sum of valid vocabularies}) > (\text{comparison news articles-frequence} / \text{sum of valid vocabularies})$ ) **then**,  
     Append (*vocabulary*) to list of (*CSR vocabulary*) # tf-idf  
     **End if**  
**End for**  
**For every** (*vocabulary*) in (*CSR vocabulary*) **then**,  
     Polarity of (*vocabulary*) is determined by encoders or SentiWordNet  
     **If not** consent **then**,  
         Reconsider (*vocabulary*)  
         **If** consent **then**,  
             Go to \*  
         **Else** still don't consent  
             Exclude (*vocabulary*) from (*CSR vocabulary*)  
         **End if**  
     **Else** consent  
         **If** the meaning of (*vocabulary*) is positive **then**,  
             Get (*vocabulary-polarity*) ← 1  
         **Else** the meaning of (*vocabulary*) is negative  
             Get (*vocabulary-polarity*) ← -1  
         **End if**  
     **End if**  
**End for**

### Obtain CSR Score

**For every** (*firm*) with complete financial data in (*FnGuide*)  
     (*sum of polarity*) ← 0  
 Get (*articles*) from web search [keyword: CSR or corporate social responsibility, and the name of each firm]  
     **For every** (*vocabulary*) in (*CSR vocabulary*)  
         **If** (*vocabulary*) is in (*articles*) **then**,  
             **If** negative or negative nuance is in a sentence **then**,  
                 Inverse (*vocabulary-polarity*) and add it to (*sum of polarity*)  
             **End if**  
             Add (*vocabulary-polarity*) to (*sum of polarity*)  
         **End if**  
     Get (*sum of polarity*) and *the number of considered URLs* of each firm  
**End for**  
**End for**

국문요약

## 온라인 텍스트 분석을 통해 추정한 기업의 사회적책임 성과가 기업의 단기적 장기적 성과에 미치는 영향 분석

이희승\* · 진윤선\* · 권오병\*\*

그동안 기업의 사회적 책임(CSR)관련 활동의 결과가 기업 성과에 미치는 단기적 및 장기적 영향에 대한 다양한 연구가 진행되었지만 그 결과는 일관되지 못한다. 그 주된 원인은 기업의 사회적책임이라고 하는 개념의 불일치였다. 따라서 본 연구는 온라인 뉴스와 같은 비정형 공개 데이터로부터 기업의 사회적책임에 관련한 키워드를 텍스트 마이닝 기법을 사용하여 추출하고 그 개념에 대한 통계치와 기업 성과와의 관계성을 이해하려고 했다. 이를 위해 개념과 관련한 키워드는 뉴욕타임즈와 구글 스칼라에서 CSR이라고 하는 단어로 검색한 비정형 데이터로부터 인식하였다. 그런 다음 점점 대상이 되는 기업에 대한 글이 실려 있는 온라인 문서를 수집하여 기업의 사회적 책임과 기업 단기적 및 장기적 성과 사이의 인과관계를 분석하였다. 그 결과, 기업의 사회적 책임에 대한 전문적인 평가 보고서의 도움 없이도 본 연구에서 개발한 기업의 사회적 책임 인덱스만으로 기업의 단기적 성과에는 영향이 없지만 장기적 성과와는 통계적으로 유의하게 정비례관계가 있는 것이 밝혀졌다. 본 연구는 빅데이터 분석을 통해 효율적이고 의미 있는 기업의 사회적 책임 평가 방법을 개발한 첫 번째 시도라는 의미가 있다.

**주제어** : 기업의 사회적 책임, 텍스트 마이닝, 비정형 데이터, 키워드 추출법, 다중회귀분석, 기업 성과, 토빈Q

논문접수일 : 2016년 5월 11일    논문수정일 : 2016년 5월 17일    게재확정일 : 2016년 5월 19일  
원고유형 : 일반논문            교신저자 : 권오병

---

\* 경희대학교차세대정보기술연구센터  
\*\* 교신저자 : 권오병  
경희대학교 경영대학  
26 Kyungheedaero, Dongdaemungu, Seoul 130-701, Korea  
Tel: +82-2-961-2148, Fax: +82-2-961-0515, E-mail: obkwon@khu.ac.kr

## 저 자 소개



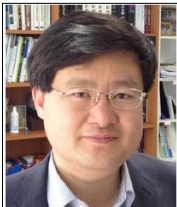
### 이희승

경희대학교에 재학 중이며 차세대정보기술연구센터의 연구조원으로 활동하고 있다. 빅데이터 분석에서의 최적화 이슈 관련 문제 해결을 주된 연구 과제로 수행하고 있으며, 책임경영 및 기업의 사회적책임 관련 연구도 추진하고 있다. 주요 관심분야는 빅데이터, 데이터마이닝, 텍스트마이닝 등이다.



### 진윤선

숙명여자대학교에서 e비즈니스학으로 석사학위를 취득하고, 현재 경희대학교에서 빅데이터경영 전공으로 박사과정에 재학 중이다. 2012년~2015년까지 웹발전연구소에서 선임연구원으로 재직한 바 있으며, 현재 개인 행복 증진을 위한 큐레이션 커머스용 연구를 주로 수행하고 있다. 주요 관심분야는 빅데이터, 공공데이터, 데이터마이닝, 텍스트마이닝 등이다.



### 권오병

현재 경희대학교 경영학과 교수로 재직 중이다. 1988년 서울대학교 경영학과 (경영학사), 1990년 한국과학기술원 경영과학과 (공학석사), 1995년 한국과학기술원 경영과학과 (공학박사)를 졸업하였다. 2001년~2002년에는 카네기멜론대학 전산학부에서 방문과학자로 근무한 바 있으며 2009년~2011년에는 샌디에고주립대학 경영정보학과의 겸직교수로 재직한 바 있다. 관심분야는 빅데이터분석, 사물인터넷, 의사결정지원시스템 등이다.