Article

Integrating Text Mining and Balanced Scorecard Techniques to Investigate the Association between CEO Message of Homepage Words and Financial Status: Emphasis on Hospitals

Hyung Jong Na[1], Kun Chang Lee [2,*] and Sung Tae Kim[3]

[1] Professor, School of Global Business Administration, Semyung University, Jecheon 27136, Republic of Korea; freshna77@semyung.ac.kr
[2] Professor, SKK Business School, Sungkyunkwan University, Seoul 03063, Republic of Korea
[3] PhD Student, School of Management, Kyung Hee University, Seoul 02447, Republic of Korea; goodthink365@naver.com
* Corresponding Author; Correspondence: kunchanglee@gmail.com

Abstract: (1) Background: The CEO message of hospital homepage contain various contents such as the hospital's future vision, promises with customers, upgraded services and public activities. The CEO's message of the homepage includes non-financial information as well as financial information of corporates.   Also, it provides useful information for not only company's goals and vision but also firm performance and strategies for the future. This study aims to investigate associations between CEO's message of hospitals homepages and financial status. We used the balanced scorecard frame to analyze what content on the hospital's homepage is related to the hospital's various financial ratios. (2) Methods: We adopt a text mining method to extract significantly repeated keywords from the CEO's message of hospital website. And we classify these keywords by a balanced scorecard frame. To examine the relationship between keywords of CEO's message of the hospital homepage and hospital's financial ratio, T-test is conducted for the difference in the TF-IDF (Term Frequency is Divided by Inverse Document Frequency) mean of the home page contents and its relationship with the views of the balanced scorecard framework. (3) Results: According to empircal results on 65 samples collected from local hospitals, there are some significant relationship between the qualitative content of the hospital's homepage and the quantitative financial ratio that indicates profitability, activity, leverage, liquidity, and transfer to essential business fund (EBF) income. (4) Conclusions: The introduction section of a homepage is most accessible to customers, containing the aims and ideals of hospitals and reflecting their values and visions [1]. In addition, in view of financial status, they can either emphasize financial strength or focus on other areas to mask weakness of financial information. This study reminds us of the importance of hospital website's disclosure, and it can be inferred from the financial status of the hospital. It also highlights the need for harmonization between quantitative data, financial statements, and qualitative data, CEO's messages. (5) Implications: To our best knowledge, this paper is the first research attempting to investigate the relation between text of hospital homepage and financial ratio of hospital through text mining technique and balanced scorecard frame. Hospitals take a crucial part in a country's welfare and healthcare backbone industry. Nevertheless, in many countries, hospital organization sectors tend to remain a source of critical fiscal deficits due to its ineffective and sloppy management. We expect that the result of this paper can provide hospital managers to useful information.

Keywords: homepage words; financial ratio; text-mining; balanced scorecard

1. Introduction

This study examines the relation between chief executive officer's (CEO) message of hospital homepage and hospital financial ratio. The CEO message of hospital homepage contain various contents such as the hospital's future vision, promises with customers, upgraded services and public activities.

CEO tries to convey various information about the firm in the CEO message. The CEO's message of the homepage includes non-financial information as well as financial information of corporates. Also, it provides useful information for not only company's goals and vision but also firm performance and strategies for the future [2].

We expect that some of the qualitatively expressed CEO's message of hospital homepage contents can be inferred that there is a significant relationship with the financial ratio. To analyze this, we use text mining technique to convert unstructured data, such as homepage text, into multiple keywords and classify them into meaningful groups based on a balanced scorecard framework [3].

The specific analysis process of this study is as follows. First, we use text mining technique to calculate the TF-IDF (Term Frequency is Divided by Inverse Document Frequency) of words from the 'About Us' or 'CEO Greeting' contents of the hospital homepage. The TF-IDF of words indicates how significantly they are repeated in the contents of the hospital homepage.

Next, we use a balanced scorecard framework to classify words extracted by text mining. The balanced scorecard consists of four perspectives: finance, customer, internal process, learning and growth. However, since hospitals are non-profit organizations that serve the people, in this study we replace financial perspectives with public views. And if keywords have nothing to do with any of these four points of view, we classify them as neutral words, adding a fifth point of view.

Third, the financial statements disclosed on each hospital's website are collected. The various financial ratios, such as hospital size, profitability, activity of fixed assets, leverage, liquidity, transferring earnings into essential business fund (EBF), are calculated using the data.

Finally, to investigate the relationship between homepage words and the financial ratio of hospitals, we conduct a t-test on the difference of the mean of TF-IDF of the words related to five perspectives. In the process, we analyze the difference of the mean of TF-IDF of the words between the high level and the low level quintiles within each perspective.

The results of this paper are as follows. Firstly, the result of public point of view shows that the current ratio has a significant difference between more and less frequently used words. There is a positive relationship between using words related to the public more frequently and increasing the current ratio. Secondly, the result of customer point of view indicates that the reserve fund translation ratio for essential business has a significant difference between more and less frequently used words. There is a positive relationship between using words related to the customer more frequently and improving the reserve fund translation ratio for essential business. Thirdly, the result of internal process point of view is that the size ratio has a significant difference between more and less frequently used words. There is a negative relationship between using words related to internal process more frequently and increasing the size ratio. In other words, using words related to internal process less frequently has a positive relationship with improving the size ratio. Fourthly, the result of learning and growth point of view shows that debt ratio, asset turnover ratio, fixed-asset turnover ratio, and reserve fund translation ratio for essential business have significant differences between more and less frequently used words. We found negative relationship between using words related to learning and growth more frequently and an increase in each ratio. Namely, using words related to learning and growth less frequently has a positive association with improving the debt ratio, fixed-asset turnover ratio, asset turnover ratio, and reserve fund translation ratio for essential business.

The contributions of this paper are as follows. First, this study uses unstructured text data of hospital homepages and conduct empirical tests. Utilizing unstructured data, we expand the range of the hospital and finance research field. Second, by conducting a t-test of the relationship between more and less frequently used words of hospital homepages, we find useful information. Some words have significant relationships to financial ratios.

The results of this study supply hospital managers and customers with an understanding of which words can be indicators for financial ratios.

This paper is organized as the following. Prior literatures related to hospital homepage and hospital finance information is addressed to develop hypotheses in section 2. Methodology adopted in this study is explained in section 3. New findings of our empirical analysis are suggested in section 4 with discussions and implications. This paper ends in section 5 with some concluding remarks.

2. Review of prior literature and hypothesis development

2.1 Vision statement

Most hospital homepage introductions mention the vision of the hospital. There has been much research done about vision statements. Quigley (1994) documented that the success of an organization depends on how well a leader can communicate his or her vision and inspire organization members [4]. Lucas (1998) explained that an effective vision statement should promote growth and development [5]. It also serves to inspire and encourages people to act toward achieving the purpose of the organization. Raynor (1998) found that many executives are convinced of the importance of mission or vision statements [6]. Berson et al. (2001) explained that organizational size relates to vision strength and moderate the association between vision strength and passive leadership style [7]. Gulati et al. (2016) stated that a vision provides an organization with a core goal as well as a bright future [8]. Further, vision is associated with performance outcomes [9].

2.2 Hospital homepage contents

Homepage contents of hospitals contain not only vision statements but other meaningful information. Prior studies related to hospital homepage contents explain information utility. Kim and Lee (1999) examined contents and operational situations of homepages [10]. To effectively provide customers with hospital information, they insist that hospital homepages should include various types of content. Lee and Ahn (2012) explained that the university hospital's website includes relatively more content than other hospitals [11]. They report that the university hospital's website relatively puts more effort into providing more information to visitors and satisfying customers. Jin et al (2011) investigated surgery using robotic content presented on the hospital's website [12]. Their results show that no hospital website commented on risks and that surgical robots overestimate profits. Lee (2013) reported that in order to improve the hospital's image, factors such as hospital homepage information and hospital management should be managed well as marketing methods [13].

2.3 Hospital financial performance

There are prior literatures related to hospital and financial performance. Nelson et al. (1992) documented hospital quality as related to financial performance of hospitals [14]. Zeller et al. (1997) outlined how measures of financial factors, such as capital structure, fixed-asset age, working capital liquidity, and fixed-asset efficiency can be beneficial to hospital boards, policy-makers, healthcare financial managers, and other relevant groups [15]. Watkins (2000) found that non-accounting information is highly significant in estimating bond grade [16]. Alexander et al. (2006) reported that the improvement of hospital quality increased hospital organizational performance [17]. Kaissi and Begun (2008) mentioned that establishment of strategic plans and management participation have positive associations with earlier financial performance [18]. Upadhyay et al. (2015) reported that hospital managers can improve profitability by reducing the length of cash conversion [19]. Dobrzykowski et al. (2016) explained that providing patients with professional service improves patient safety and hospital financial performance [20]. Wang et al. (2018) presented that health information technology expense, capital expense, and information technology operating expense have positive associations with return on assets [21].

2.4 Hospital research using text mining

Hospital research has yet to make full use of text mining. However, the following introduces research in the hospital field using text mining. Hahn et al. (2001) analyzed major patterns in medical documents and extracted knowledge using a natural language process [22]. Zhou et al. (2010) mentioned that text mining is one of the most valuable methodologies for research subfields in data mining [23]. Text mining of patient records can provide valuable information and assist decision making [24]. Yang et al. (2009) utilized text mining to analyze clinical data and predict disease condition from clinical discharge summaries [25]. They explained the possibility of text mining for more precise prediction of disease status. Kocbek et al. (2016) documented a text mining for detecting marks as positive for some diseases [26]. Using text mining in the medical field can help advance medical technology by making it easier to acquire and utilize new knowledge from medical literature [23].

2.5 Hospital literatures related to balanced scorecard technique

Hospital studies related to the balanced scorecard technique are as follows. Stewart and Bestor (2000) designed an integrated performance measurement using balanced scorecard [27]. They developed a balanced scorecard of hospitals consisting of one composite financial performance proxy and 12 nonfinancial performance proxies. These performance proxies are used in terms of actual performance as an expected performance percentage and an overall performance score is calculated through a subjective weighting scheme. Pink et al. (2001) reported on hospitals in Canada using the balanced scorecard framework [28]. Indicators of balanced scorecard performance were developed in four areas—patient satisfaction, system integration and change, clinical utilization and outcomes, and financial performance and condition. Chen et al. (2006) explained how a balanced scorecard was effective for identifying chances for improvement and underlining existing problems [29]. A balanced scorecard reveals the contribution of hospitals to improve performance in the health system. Zhijun et al. (2014) investigated the current application of the balanced scorecard and the effect upon hospital performance in China [30]. They reported that some of Chinese hospitals had used a balanced scorecard system in hospital administration. The findings showed that balanced scorecard application contributed to improved organizational performance. In addition, they found that the balanced scorecard system is affected by technological quality, operational scope, and comprehensiveness of medical resources. Walker and Dunn (2006) insisted that a balanced scorecard system can measure productivity of hospitals and improve hospital management at reduced cost without loss of quality [31].

2.6 Hypothesis development

Hospital homepage content contains valuable and meaningful information. In particular, introduction sections on homepages offer a summary of important information about the hospital. This content usually mentions vision and performance outcomes [9]. Hence, we expect that some of the words on hospital homepages might have significant association with financial performance.

In this study, to examine the association between introduction words of hospital homepages and financial information, we use the text mining method and balanced scorecard frame. Text mining plays an important role in obtaining useful implication and meaning from hospital information systems [26]. A balanced scorecard can be key in achieving strategic goals and operating hospitals effectively [32]. Further, performance measurement using the balanced scorecard system provides an effective way for hospitals to achieve strategic goals [27].

We extract keywords from introduction content on hospital homepages and classify meaningful words into groups using a balanced scorecard. Then, we conduct a t-test to analyze the relationship between introduction words and financial information. We ex-

pect that if we find significant results, introduction text on hospital homages can be indicators of hospital financial information. Below, <Figure 1> explains the concept and methodology of this research. Therefore, we develop a hypothesis:
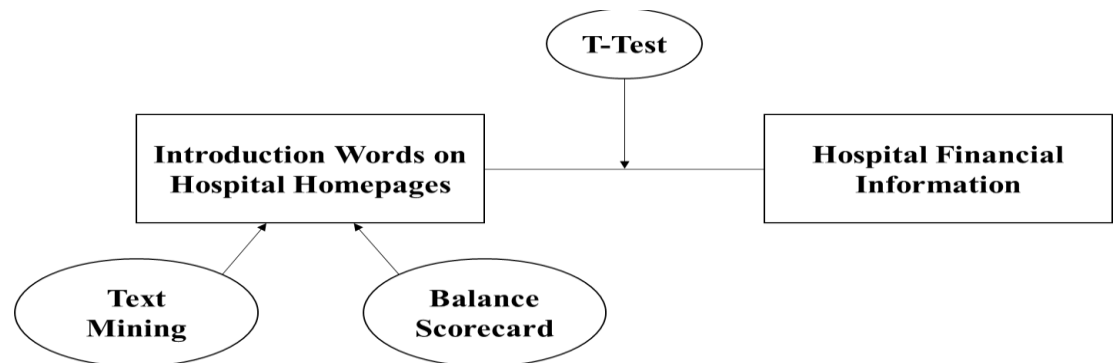

Figure 1. The concept and methodology of this research

H: There might be some significant relationships between CEO's message of hospital homepage and financial information.

3. Material and Methods

This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation, as well as the experimental conclusions that can be drawn.

3.1. Sample selection

3.1.1. Text mining methodology

This study obtains homepage text data and financial statements of 65 public general hospitals that have more than one hundred beds. We collect introduction text from 'About Us' or 'CEO's Greeting' pages on their websites. To quantify the text, we utilize the text mining method. Using Java-contained functions, the text is automatically collected. Morpheme division work is done with the codes contained in Java using Porter's (1980) algorithm. To remove disused words, we conduct POS tagging, a process that classifies words as nouns, verbs, adjectives, adverbs, and so on. Next, we remove verbs, adjectives, adverbs, propositions, and special marks. The symbols that have no lexical meaning, such as sentence marks, are removed. Basic HTML tags are also removed during the text extracting process using a Java-contained function.

Next, we calculate each word's TF-IDF value used in About Us or CEO's Greeting content. The progress is based on term frequency information of the Bag of Words model. TF-IDF is a weighted value that allows us to understand how important a given word is in various documents [21]. The reason we do not simply calculate frequency is that frequency itself cannot determine importance in the content. Meaningless words can appear in high frequency and may not play a proper role in distinguishing the difference by their attentive repeat. Therefore, in this paper, we use the values of TF (Term Frequency) and IDF (Inverse Document Frequency).

Words with more than 1% frequency are selected. Including all words such as subjects of analysis will make the calculation burdensome, so the words with 1% or more frequency are left in each file. The frequency calculation is conducted by file; therefore, a word left in a specific file might not remain in other files. Our purpose is to find words representing the file's traits, so high-frequency words should be analyzed. Hence, we align the words used in each file and record their frequency. As the words appearing in the files are different, we align to avoid overlap after they are collected in one place and record how many times the words are used in each file. Since the TF-IDF can be different in each file, we record the TF-IDF collectively.

3.1.2 Balanced scorecard frame

Balanced scorecard as a management accounting methodology has developed and improved for a general management practice [33]. The balanced scorecard is a comprehensive management framework that constantly interacts with learning and growth, customers, internal processes, and financial perspectives. The aim of the four perspectives is linked together by cause-and-effect relations [34]. That is, the BSC shows how the interaction is taking place not only from a financial perspective but also from a non-financial perspective, such as customer, process, and learning growth, by extracting key success factors based on the four perspectives and developing them in the direction of setting the firm's strategy and goals [35].

We categorize words extracted by text mining, applying the balanced scorecard technique. Many prior studies adopted this method to handle narrative reports [3, 36, 37, 38, 39, 40]. Balanced scorecard consists of four elements: customer, financial, learning and growth perspective, and internal process. However, hospitals can only be established as nonprofit organizations. This means hospitals should not aim for financial profit. Therefore, this study changes 'financial perspective' of the common balanced scorecard to the 'public perspective' in accordance with the purpose of the establishment of hospitals.

In summary, we classify the keywords extracted by text mining into four perspectives based on a balanced scorecard frame modified in accordance with the purpose of nonprofits: public (PB), patient (PT), internal process (IP), and learning and growth (LG). Yet, some words are hard to classify exactly with one of these four perspectives because they have general or vague notion. Thus, they are classified as neutral (NT), and we remove those words from the final balanced scorecard perspective in our analysis

3.2 Model specification

Our research framework is as follows. First, we calculate the mean of TF-IDF included words in each of the four balanced scorecard perspectives (TI_PB, TI_PT, TI_IP and TI_LG). Each mean indicates to what degree a manager focuses on the content of that particular perspective in their homepage text compared with other hospitals. For example, if TI_PB is high, the manager focuses on their public vision in the introduction content. Second, we divide sample to quintile by the mean of TF-IDF in balanced scorecard perspective, defining that first and second quintile samples are 'high balanced scorecard' and fourth and fifth quintile samples are 'low balanced scorecard'. Finally, we compare the mean of financial ratio in the high balanced scorecard group with the low balanced scorecard group by using the t-test method.

Our study aims to examine the relation between non-financial data and a hospital's financial information. We utilize financial proxies to assess hospital financial performance such as profitability, activity, leverage, and liquidity ratio. Furthermore, Korean hospitals can retain their income as EBF within a certain limit ranging from 50% to 100%, depending on their establishment form. This money should only be used for medical purposes such as buying medical equipment or building a new ward, and it classifies as long-term liability in a financial statement. Considering this unique characteristic of Korean hospitals, we add several ratios related to the essential business fund to our analysis.

Commonly used financial information are ratios of profitability, growth, leverage, liquidity, and activity. However, we cannot calculate a growth ratio (the change between this year's revenue and previous year's revenue to the previous year's revenue), as we only have one year of data from financial statements. Therefore, in this study, we compare each high balanced scorecard hospital's financial information with low balanced scorecard hospitals' information, which is profitability, activity, leverage, and retaining an essential business fund.

The proxies in our research are size (SIZE), return on asset (ROA), fixed-asset turnover (FTOV), long-term debt ratio (LTLEV), current ratio (CUR), and transferring earning into essential business fund ratio (TRANS). Each financial ratio is calculated as follows:

Table 1. Variable definition

Financial Information			Variables
Hospital size	:	SIZE	(= natural logarithm of total asset)
Profitability	:	ROA	(= net income divided by total asset)
Activity	:	FTOV	(= medical revenue divided by fixed asset)
Leverage	:	LTLEV	(= long-term debt scaled by total asset)
Liquidity	:	CUR	(= current asset divided by current liability)
Transfer into EBF	:	TRANS	(=Translation to EBF scaled by pre-tax income)

4. Results

4.1 Descriptive statistics

<Table 2> shows the descriptive statistics used in this research. Final hospital samples are 65; according to balanced scorecard frame, there are four classified variables, and there are six financial ratios used for the t-test. TI_PB, TI_PT, TI_IP, and TI_LG are the means of TF-IDF of included words in each of the four balanced scorecard perspectives. The means of TI_PB, TI_PT, TI_IP, and TI_LG in <Table 2> are the means of TF-IDF of text on hospital homepages. The TI_PB, TI_PT, TI_IP, and TI_LG means are 0.044, 0.039, 0.042, and 0.044. The standard deviations of TI_PB, TI_PT, TI_IP, and TI_LG are 0.023, 0.021, 0.023, and 0.024.

The mean of ROA is 0.000, standard deviation is 0.103, and median is -0.003. These numbers indicate that most hospitals do not report profit. The mean of TRANS is 0.535 and the standard deviation is 0.924, which implies that a considerable portion of income is retained as essential business funds.

Table 2. Descriptive statistics

Variable	N	Mean	Std	Min	Q1	Median	Q3	Max
TI_PB	65	0.044	0.023	0.000	0.028	0.043	0.061	0.090
TI_PT	65	0.039	0.021	0.000	0.024	0.036	0.053	0.109
TI_IP	65	0.042	0.023	0.000	0.021	0.043	0.059	0.093
TI_LG	65	0.044	0.024	0.000	0.028	0.044	0.060	0.101
SIZE	65	25.208	1.746	21.416	23.594	25.757	26.357	29.195
ROA	65	0.000	0.130	-0.391	-0.037	-0.003	0.026	0.573
FTOV	65	5.377	8.077	0.000	1.547	2.395	4.614	36.906
LTLEV	65	0.456	0.576	0.000	0.107	0.304	0.589	3.480
CUR	64	1.407	1.182	0.086	0.788	1.049	1.698	7.379
TRANS	66	0.535	0.924	-2.525	0.000	0.000	1.034	4.306

4.2 Research results

<Table 3> shows t-test results examining the mean differences of financial ratio. First, in terms of profitability, ROA of a high-PB hospital is higher than that of a low-PB hospital, but not significant (t-value=0.75, p=0.458). The mean of activity and leverage ratios, which are represented by FTOV and LTLEV, are higher in low-PB hospitals, but not statistically significant (t-value=-0.21, p=0.834; t-value=-1.18, p=0.243). In contrast, the mean difference of CUR is statistically significant at 5%, which indicates that high-PB hospitals have more liquidity than low-PB hospitals. On the other hand, high-PB hospitals' TRANS ratio is lower, but not statistically significant, similar to the activity and stability ratios. This result implies that hospitals using many words related to their public vision have higher financial liquidity than other hospitals. However, other financial information is almost identical.

Table 3. The mean difference of TF/IDF on PB between high (1,2) and low (4,5) quintile

Variable	High			Low			Difference	t-value	Prob-t	Significance
	N	Mean	Std	N	Mean	Std				
SIZE	26	25.186	1.689	26	25.602	1.474	-0.416	-0.95	0.349	
ROA	26	0.017	0.116	26	-0.004	0.077	0.020	0.75	0.458	
FTOV	26	5.250	7.197	26	5.765	10.155	-0.515	-0.21	0.834	
LTLEV	26	0.315	0.337	26	0.439	0.419	-0.125	-1.18	0.243	
CUR	25	1.666	1.226	26	1.084	0.466	0.582	2.22	0.034	**
TRANS	27	0.479	0.950	26	0.540	0.594	-0.061	-0.28	0.780	.

<Table 4> presents the t-test result examining the mean differences of financial ratio between high-PT hospital and low-PT hospitals. First, profitability (ROA) of high-PT hospitals is higher than that of low-PT hospitals, but not significant, although there is a relatively high t-value (t-value=1.60, p=0.115). Similar to <Table 3>, the mean of activity and leverage ratios, which are represented by FTOV and LTLEV, are higher in low-PT hospitals, but not statistically significant (t-value=-1.13, p=0.267; t-value=-0.21, p=0.838). Unlike <Table 3>, the liquidity ratio of high-PT hospitals is lower than that of low-PT hospitals, but the mean is not significantly different. However, the gap in the TRANS ratio between the two is statistically significant at 10% (t-value=-1.82, p=0.08). This result indicates that hospitals using many words related to the patient perspective transfer earnings into essential business funds for future medical investment more than other hospitals.

Table 4. The mean difference of TF/IDF of Patient perspective between high (1,2) and low (4,5) quintile

Variable	High			Low			Difference	t-value	Prob-t	Significance
	N	Mean	Std	N	Mean	Std				
SIZE	27	25.387	1.881	26	25.119	1.682	0.268	0.55	0.587	.
ROA	27	0.476	0.668	26	0.352	0.436	0.124	0.80	0.427	.
FTOV	27	0.044	0.140	26	-0.013	0.120	0.057	1.60	0.115	.
LTLEV	27	3.949	4.546	26	6.526	10.747	-2.577	-1.13	0.267	.
CUR	26	1.260	1.022	26	1.314	0.868	-0.054	-0.21	0.838	.
TRANS	27	0.738	0.977	26	0.286	0.822	0.452	1.82	0.075	*

<Table 5> shows the t-test result examining the mean differences of financial ratio between high-IP hospitals and low-IP hospitals. As can be seen in <Table 4>, no statistical significance of mean difference in any financial ratio exist except for SIZE. This result can be interpreted to mean that hospital managers do not allude to financial information using words related to an internal process perspective in homepage content. On the contrary, the mean of a high-IP hospital's size is bigger than that of a low-IP hospital's size, and it is statistically significant at 1% level (t-value=-3.06, p=0.004). This means that small hospitals' managers try to overcome the relative weakness derived from their scale by emphasizing the effectiveness of their internal process rather than communicating other factors on their homepages.

Table 5. The mean difference of TF/IDF of Patient perspective between high (1,2) and low (4,5) quintile

Variable	High			Low			Difference	t-value	Prob-t	Significance
	N	Mean	Std	N	Mean	Std				
SIZE	26	24.354	1.788	26	25.687	1.322	-1.332	-3.06	0.004	***

ROA	26	0.542	0.786	26	0.396	0.430	0.146	0.83	0.411	.
FTOV	26	0.039	0.170	26	-0.014	0.072	0.052	1.45	0.156	.
LTLEV	26	7.134	9.282	26	4.801	8.277	2.333	0.96	0.343	.
CUR	25	1.567	1.471	26	1.298	1.053	0.269	0.75	0.455	.
TRANS	27	0.405	0.711	26	0.600	0.751	-0.195	-0.97	0.336	.

<Table 6> shows the t-test result examining the mean differences of financial ratio between high-LG hospitals and low-LG hospitals. Unlike IC, there is no significance in the mean difference of hospital size between high- and low-LG hospitals. The means of LTLEV, FTOV, and TRANS ratios are significantly different between high- and low-LG hospitals at 10%, 5%, and 10%, respectively. Further, although not statistically significant, the mean difference of the ROA ratio is almost five percentage points between high- and low-LG hospital. This empirical result suggests that the high-LG hospitals have relatively insufficient debt financing capability, inefficient usage of fixed assets, and lack of capacity for future medical investment. Financial information of low-LG hospitals is generally more favorable than that of high-LG hospitals. Hence, managers of high-LG hospitals are likely to promote their future growth possibilities by writing homepage text related to the learning and growth perspective rather than to existing financial performance.

Table 6. The mean difference of TF/IDF of LG between high (1,2) and low (4,5) quintile

| Variable | High | | | Low | | | Difference | t-value | Prob-t | Significance |
	N	Mean	Std	N	Mean	Std				
SIZE	26	25.569	1.725	26	24.945	1.879	0.623	1.25	0.219	.
ROA	26	0.341	0.329	26	0.642	0.807	-0.301	-1.76	0.087	*
FTOV	26	-0.019	0.061	26	0.028	0.173	-0.047	-1.31	0.200	.
LTLEV	26	3.179	3.180	26	8.847	11.537	-5.668	-2.42	0.022	**
CUR	26	1.560	1.405	26	1.365	1.159	0.195	0.55	0.588	.
TRANS	27	0.385	0.598	26	0.814	0.982	-0.430	-1.92	0.063	*

5. Discussion

The empirical results and implications of this paper are as follows. First, low-PB hospitals have low financial liquidity. This means managers of hospitals in financial distress do not focus on public value on their homepages. Second, high-PT hospitals tend to transfer their earnings to essential business funds. Hospitals focused on patient satisfaction are likely to pursue future medical investment over present performance. Third, there is no significant difference in financial ratio between high-IP and low-IP hospitals, but the size of high-IP hospitals is smaller than that of low-IP hospitals. This result implies that managers of small hospitals account for relative weakness due to size by emphasizing their internal operating efficiency. Finally, high-LG hospitals have relatively low long-term leverage, activity of fixed assets, and transferring earnings into EBF.

This study has the following limitations: The results of the empirical analysis of this study are difficult to generalize due to the small number of samples. Also, since we analyzed only general hospitals that met certain criteria, the results could be different when analyzing smaller hospitals.

Despite this limitation, this paper has the following contributions. First, as far as we know, this paper is the first study to investigate the relation between hospital financial information and hospital CEO's message using text mining technique. In particular, we expect this study to provide a broad understanding of the text mining methodology in the hospital research area. Second, this study did not rely on subjective viewpoint, but instead

used a balanced scorecard frame to classify CEO's messages on the hospital's homepage. Based on this, we find a significant relationship with the hospital's financial ratio. Namely, this study is expected to contribute to further research development by presenting a new methodology for hospital management research. Also, it is expected to remind hospital managers of the importance of qualitative data disclosure online in hospital management.

6. Conclusions

We examine the association between CEO message of hospital homepage and hospitals' financial status. The hospital's CEO communicates variety of information with customer, such as the achievement of the organization's goals, current status of the organization, and organizational policy. The introduction words of a homepage serve as a summary of important information about the organization [41]. Particularly, the introduction section of a homepage is most accessible to customers, containing the aims and ideals of hospitals and reflecting their values and visions [2]. In addition, in view of financial status, they can either emphasize financial strength or focus on other areas to mask weakness of financial information [42].

This study analyzed CEO's messages on the hospital's website and investigated their relationship with hospital reload status. To examine this relationship, we used text mining method and balanced scorecard frame. We calculated TF-IDF of each word extracted by the text mining method, which suggests the importance of words used in the homepage text written by every hospital's manager. To categorize extracted words into several groups, we adopted a balanced scorecard perspective. Since all hospitals in Korea are non-profit organizations, we adjusted the financial performance perspective of the balanced scorecard to a public perspective.

The findings of this paper are expected to provide useful information for hospital managers in developing policies. It reminds us of the importance of hospital website's disclosure, and it can be inferred from the financial status of the hospital. It also highlights the need for harmonization between quantitative data, financial statements, and qualitative data, CEO's messages.

Appendix A

Table 7. The words categorized to balance scorecard

Category name	n of words	Details
Public (PB)	16	Domestic, nation, contribution, Republic of Korea, resident, welfare, service, society, life, initiative, citizen, action, inhabitant, the first, vulnerability, happiness
Patient (PT)	15	home, family, appreciation, client, interest, origin, mind, visit, love, trust, appointment, use, information, support, sincerity
Internal Process (IP)	18	open, base, faculty, establishment, international, introduction, business, foundation, world, facilities, system, safety, high quality, role, management, equipment, cutting edge, securement

Learning and Growth (LG)	18	care, education, change, ward, sickbed, sanitation, study, medical personnel, medicine, specialty, spirit, enhancement, continuation, disease, illness, high-tech, treatment, environment
Neutral (NEU)	13	class, national, university hospital, basis, level, performance, beginning, history, we, best, the best, the latest, the present

References

1. Lau, K. N.; Lee, K. H.; Ho, Y.; Lam, P. Y. Mining the web for business intelligence: Homepage analysis in the internet era. *Journal of Database Marketing & Customer Strategy Management* **2004**, 12, 32-54.
2. Lee, S. M.; Kwon Lee, C. E-company CEO websites: contents and information value. Management Decision **2002**, 40, 158-167.
3. Kapian, R. S.; Norton, D. P. The Balanced Scorecard: measures that drive performance. *Harvard Business Review* **1992**, 1.
4. Quigley, J. V. Vision: How leaders develop it, share it, and sustain it. *Business Horizons* **1994**, 37, 37-42.
5. Lucas, J. R. Anatomy of a vision statement. *Management Review* **1998**, 87, 22-27.
6. Raynor, M. E. That vision thing: Do we need it?. *Long range planning* **1998**, 31, 368-376.
7. Berson, Y.; Shamir, B.; Avolio, B. J.; Popper, M. The relationship between vision strength, leadership style, and context, *The Leadership Quarterly* **2001**, 12, 53-73.
8. Gulati, R.; Mikhail, O.; Morgan, R. O.; Sittig, D. F. Vision statement quality and organizational performance in US Hospitals, *Journal of Healthcare Management* **2016**, 61 335-350.
9. Kantabutra, S.; Avery, G. C. The power of vision: statements that resonate. *Journal of Business Strategy* **2010**, 31, 37-45.
10. Kim, T. S.; Lee, H. J. Contents and operational situation of internet homepage in general hospital. *Korea Journal of Hospital Management* **1999**, 4, 192-218.
11. Lee, J. W.; Ahn, S. Y. A study on public relation activation plan through hospital homepage contents analysis. *The Korean Journal of Health Service Management* **2012**, 6, 13-27.
12. Jin, L. X.; Ibrahim, A. M.; Newman, N. A.; Makarov, D. V.; Pronovost, P. J.; Makary, M. A. Robotic surgery claims on United States hospital websites, *Journal for Healthcare Quality* **2011**, 33(6), 48-52.
13. Lee, H. S. The impact of hospital selection factors between internal and external customers on hospital image. *Health Policy and Management* **2013**, 23, 281-288.
14. Nelson, E. C.; Rust, R. T.; Zahorik, A.; Rose, R. L.; Batalden, P.; Siemanski, B. A. Do patient perceptions of quality relate to hospital financial performance?. *Journal of Health Care Marketing* **1992**, 12, 6-13.
15. Zeller, T. L.; Stanko, B. B.; Cleverley, W. O. A new perspective on hospital financial ratio analysis. *Healthcare Financial Management* **1997**, 51, 62-67.
16. Watkins, A. L. Hospital financial ratio classification patterns revisited: Upon considering nonfinancial information. *Journal of Accounting and Public Policy* **2000**, 19, 73-95.
17. Alexander, J. A.; Weiner, B. J.; Griffith, J. Quality improvement and hospital financial performance. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior* **2006**, 27, 1003-1029.
18. Kaissi, A. A.; Begun, J. W. Strategic planning processes and hospital financial performance. *Journal of Healthcare Management* **2008**, 53, 197-209.
19. Upadhyay, S.; Sen, B.; Smith, D. The cash conversion cycle and profitability: A study of hospitals in the state of Washington. *Journal of Health Care Finance* **2015**, 41, 1-9.
20. Dobrzykowski, D. D.; McFadden, K. L.; Vonderembse, M. A. Examining pathways to safety and financial performance in hospitals: A study of lean in professional service operations. *Journal of Operations Management* **2016**, 42, 39-51.
21. Wang, T.; Wang, Y.; McLeod, A. Do health information technology investments impact hospital financial performance and productivity?. *International Journal of Accounting Information Systems* **2018**, 28, 1-13.
22. Hahn, U.; Romacker, M.; Schulz, S. Creating knowledge repositories from biomedical reports: the MEDSYNDIKATE text mining system, *In Biocomputing 2002*, **2001**, 338-349.
23. Zhou, X.; Peng, Y.; Liu, B. Text mining for traditional Chinese medical knowledge discovery: a survey. *Journal of biomedical informatics* **2010**, 43, 650-660.
24. Lucini, F. R.; Fogliatto, F. S.; da Silveira, G. J.; Neyeloff, J. L.; Anzanello, M. J.; Kuchenbecker, R. D. S.; Schaan, B. D. Text mining approach to predict hospital admissions using early medical records from the emergency department. *International journal of medical informatics* **2017**, 100, 1-8.
25. Yang, H.; Spasic, I.; Keane, J. A.; Nenadic, G. A text mining approach to the prediction of disease status from clinical discharge summaries. *Journal of the American Medical Informatics Association* **2009**, 16, 596-600.
26. Kocbek, S.; Cavedon, L.; Martinez, D.; Bain, C.; Mac Manus, C.; Haffari, G.; Zukerman, I.; and Verspoor, K. Text mining electronic hospital re cords to automatically classify admissions against disease: Measuring the impact of linking data sources. *Journal of Biomedical Informatics* **2016**, 64, 158-167.
27. Stewart, L. J.; Bestor, W. E. Applying a balanced scorecard to health care organizations. *Journal of Corporate Accounting & Finance* **2000**, 11, 75-82.
28. Pink, G. H.; McKillop, I.; Schraa, E. G.; Preyra, C.; Montgomery, C.; Baker, G. R. Creating a Balanced Scorecard for a hospital

system. *Journal of Health Care Finance* **2001**, 27, 1-20.

29.   Chen, X. Y.; Yamauchi, K.; Kato, K.; Nishimura, A.; Ito, K. Using the balanced scorecard to measure Chinese and Japanese hospital performance. *International Journal of Health Care Quality Assurance* **2006**, 19, 339-350.

30.   Zhijun, L. I. N.; Zengbiao, Y. U.; Zhang, L. Performance outcomes of balanced scorecard application in hospital administration in China. *China Economic Review* **2014**, 30, 1-15.

31.   Walker, K. B.; Dunn, L. M. Improving Hospital Performance and Productivity with the Balanced Scorecard. *Academy of Health Care Management Journal* **2006**, 2, 85-110.

32.   Souissi, M.; Ito, K. Implementing and Cascading the Balanced Scorecard: The Case of a Danish Hospital. *Journal of Accounting, Ethics and Public Policy* **2018**, 19, 203-214.

33.   Cooper, D. J.; Ezzamel, M.; Qu, S. Q. Popularizing a management accounting idea: The case of the balanced scorecard. *Contemporary Accounting Research* **2017**, 34, 991-1025.

34.   Kaplan, R. S.; Norton, D. P. Measuring the strategic readiness of intangible assets. *Harvard business review* **2004**, 82, 52-63.

35.   Kaplan, R. S.; Norton, D. P. Transforming the balanced scorecard from performance measurement to strategic management: Part I. *Accounting horizons* **2001**, 15, 87-104.

36.   Hoque, Z.; James, W. Linking balanced scorecard measures to size and market factors: impact on organizational performance. *Journal of management accounting research* **2000**, 12, 1-17.

37.   Lipe, M. G.; Salterio, S. E. The balanced scorecard: Judgmental effects of common and unique performance measures. *The Accounting Review* **2000**, 75, 283-298.

38.   Figge, F.; Hahn, T.; Schaltegger, S.; Wagner, M. The sustainability Balanced Scorecard–theory and application of a tool for value-based sustainability management. *In Greening of Industry Network Conference, Gothenburg,* **2002**, 2.

39.   Gumbus, A.; Lussier, R. N. Entrepreneurs use a balanced scorecard to translate strategy into performance measures. *Journal of Small Business Management* **2006**, 44, 407-425.

40.   Hansen, E. G.; Schaltegger, S. The sustainability balanced scorecard: A systematic review of architectures. *Journal of Business Ethics* **2016**, 133, 193-221.

41.   Na, H. J.; Choi, S.; Kwon, O. The Association of Institutional Information on Websites with Present and Future Financial Performance. *Journal of Society for E-business Studies* **2018**, 23, 63-85.

42.   Barkemeyer, R.; Comyns, B.; Figge, F.; Napolitano, G. CEO statements in sustainability reports: Substantive information or background noise?. In *Accounting Forum* **2014**, 38, 241-257.