KSCE Journal of Civil Engineering (2018) 22(5):1535-1543

DOI 10.1007/s12205-017-0767-7

Construction Management

pISSN 1226-7988, eISSN 1976-3808
www.springer.com/12205

Analysis of Fire-Accident Factors Using Big-Data Analysis Method for Construction Areas

Joon-Soo Kim* and Byung-Soo Kim**

Received March 25, 2016/Revised 1st: October 28, 2016, 2nd: May 19, 2017/Accepted June 19, 2017/Published Online August 15, 2017

Abstract

The Ministry of Employment and Labor releases its annual report on the present conditions of industrial disasters by aggregating and summarizing negligent accidents that occur at construction sites. Industry-specific accident and fatality rates, and disaster classification and statistics are aggregated in this report, but its effectiveness is low. This is due to the fact that it does not sufficiently present the direct causes of accidents or related information on their causal relation. However, this study utilizes a big-data method that has recently gained significant attention throughout all industrial and academic areas to collect Internet articles on fire-accidents that have occurred at construction sites over the last decade. In addition, principal component analysis was conducted to deduce season-specific factors according to time, location, inducer, and accident pattern. Based on this analysis, as for common factors, direct spark and oil mist were deduced. As work-related factors, negligent supervision and violations of the safety regulations were shown to cause fire-accidents, illustrating the man-made nature of such accidents. It was also found that secondary accidents such as collapses, burials, explosions, and suffocation have occurred when fires have broken out. The big-data analysis method utilized in this study is considered to be very effective and can be successfully utilized in the future for deducing high volumes of text data.

Keywords: *big-data, negligent accident, web crawling, text mining, data mining, principal component analysis*

1. Introduction

Along with the advancements in information and communication technology (ICT), big data are receiving significant attention. The processing and analytical capacity of big data have indicated their future competitiveness, and the 2012 World Economic Forum pointed at big data as the most notable technological advancement. In accordance with this trend, high volumes of information are needed and are being generated in current construction sites, which are becoming larger in scale, more complex, and highly specialized, and the development of an according system for the management and processing of such sites is garnering significant attention.

Korea Occupational Safety & Health Agency (KOSHA) establishes the industrial disaster cause investigation report ever year. This report analyzes the investigation result to the cause of work accidents after survey and the process of accidents for works. But, the statistics data only provide the stats for industrial, scale, occurrence type and matter of cause etc. of accidents, but fail to provide the detail cause of each accident type. For example, the data doesn't provide the detail cause of time, season, location and work of fire accident. Therefor it is significantly appropriate to derive a detail cause of fire accident non-supplied in the statistics using big data analysis technique.

Using a big-data analysis method, which is receiving significant attention in all areas of industry and academics, this study collected news information on fire-accidents that occurred during the last decade according to season, and a PCA was applied to deduce the seasonally specific factors of a fire-accident.

A construction site safety management monitoring system was recently developed based on a Ubiquitous Sensor Network (USN), which can be applied to a construction site through the implementation of Information Technology (IT) (Kim. 2009). In addition, learner-centered tools have been developed to provide construction safety education at universities. However, there have not been many studies utilizing large volumes of information generated at construction sites. Accordingly, the big-data method utilized as the core element of this study is a method for discovering new facts, rules, or predictable patterns by collecting large volumes of information and using existing statistical analyses.

Big-data methods can be classified into social big data for analyzing patterns and trends when collecting data from a Social Network Service (SNS), spatial big data for analyzing regional and space-specific data, and the Internet of Things for analyzing sensor data among things. Regarding studies conducted on social big data used in the construction industry, Choi *et al.* (2015) developed a Social Big Board for collecting social media data

*Member, Ph.D. Candidate, Dept. of Civil Engineering, Kyungpook National University, Daegu 41566, Korea (E-mail: kimjoonsoo@knu.ac.kr)
**Member, Professor, Ph.D., Dept. of Civil Engineering, Kyungpook National University, Daegu 41566, Korea (Corresponding Author, E-mail: bskim65@knu.ac.kr)

and analyzing region-specific location disclosure information according to the disaster occurrences through which a reliable disaster response system can be established. Bae *et al*. (2014), also conducted experiments on detecting explosions by analyzing the syntactic patterns in social big data to detect man-made disasters including explosions.

For researches on the Internet of Things pertaining to big data, a big-data system has been developed and implemented by an Australian construction company to prevent negligent accidents in mines or at construction sites.

Through a real-time analysis of the data generated at construction sites, this system provides an early warning of risk to workers. The data this system analyzes include existing accident records and daily task reports, Closed Circuit Television (CCTV) images, and location data on various risk factors. In addition, various forecasting models based on text and video analytics are used.

Shin *et al.* (2014) selected the priority of fire and explosion danger of chemical material, and presented guidance, supervision and safety education method for disaster prevention of middle and small scale site after analyze the prevention data of the fire and explosion disaster. Shin (2015) presented that cause of welding, heater and electric power as current situation each principle factor of fire explosion is 78%, and suggested transmission and distribution, machine facility and chemical products as cause material. Oh (2015) has studied fire disaster factors of 7 major industries and to establish a preventive measure with special reference to apartment building project. For this purpose, countermeasure was planned by analyzing disaster cases and preventive measure was suggested by figuring out progression of work, risk factors of fire by industrial type and inflammable and combustible materials. As mentioned above, research about fire accident cause analysis have studies that have only analyzed fire cause material and each work process. However, there has not been a study that analyzed a variety of fire cause of fire accident (as time, season and official management etc.). Therefore, this study aims to analyze the fire cause of fire accident based on the collected data of news accrued in construction site using big data analysis method.

This study, consisting of information collection and text and data mining, was conducted using R, a statistical analysis program. For information collection, an R library was used to collect SBS Internet news articles on negligent accidents that have occurred at construction sites during the last decade, and to conduct Web crawling modeling as a way to collect the desired information located on the Internet. In addition, Internet news articles released during the period of Jan. 1, 2000 through Dec. 31, 2014 were collected using the keywords “construction site negligent accident.”

For text mining, a morphological analysis was conducted before measuring the frequency to extract Internet news articles consisting of sentences into a vocabulary format by excluding the post position and suffix for processing them into a format useful for analysis prior to the data mining. The frequently measured words include those indicating the types of accidents that occur at construction sites and words indicating the cause of such accidents.

During the text mining, the frequency of words indicating various accident types was used to deduce six types of negligent accidents that have frequently occurred at construction sites during the last decade (fires, explosions, collapses, heavy equipment problems, falls, and suffocation). A total of 170 factors considered to have affected such accidents frequently appeared. The aggregate was then dubbed the Factor Dictionary, which was used to effectively extract words related to negligent accidents when conducting text mining during the collection of reports on accident-specific negligence. Data mining was then conducted on frequently occurring fire-accidents at construction sites identified through text mining. The frequently appearing words consist of factors considered to have affected such fire-accidents and those factors hindering an analysis. After removing factor words inappropriate for this study using the Factor Dictionary, Principle Component Analysis (PCA) was performed on the filtered results to deduce the factors that were considered to have affected fire-accidents.

2. Theoretical Consideration

2.1 Big-data System

Considering the characteristics of a big-data analysis, it is necessary to establish a system for processing and analyzing large volumes of information within a short period of time and in a real-time manner. Fig. 1, shows a big-data system based on Hadoop that has most frequently been established by government agencies and corporations. Hadoop is an open-source framework that supports distributed application programs that operate in a massive computer cluster for conducting high-volume processing (Lee, 2012).

Among the functions supported by Hadoop, MapReduce is a function for collecting high-volume data in the form of parallel processing using multiple node PC units. Because Hadoop is a

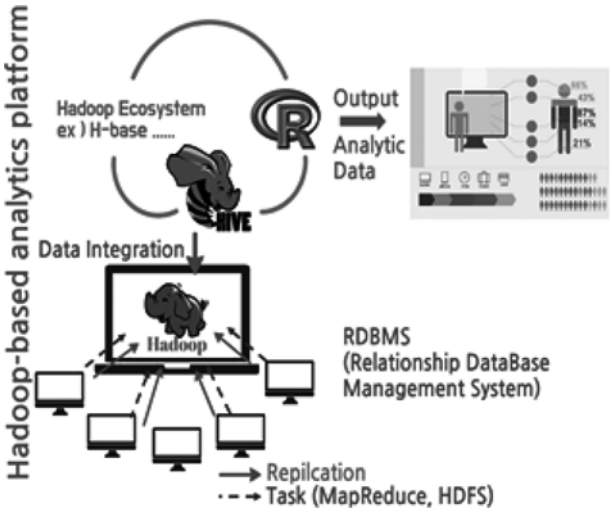

Fig. 1. Big-Data System

specialized program for information collection, it requires a separate analytical program to analyze information. The most widely used programs are Hadoop HIVE and R, an open-source statistical program, and the programs allow users unfamiliar with statistics to easily obtain the results by providing suitable algorithms for cluster, pattern, trend, and forecasting analyses.

2.2 Big-Data Analysis Program, R

R (Version 3.2.3) is a programming language developed based on S Language by Bell Labs, a private R&D institute in the USA. It was developed as a part of free software, i.e., a mass collaboration project (GNU Project), as a programming language for applying statistical computations and graphics. Although R was not widely used in the past, recently it is being frequently used as a big-data analytical tool in various areas as a package of a Comprehensive R Archive Network (CRAN), along with the recent trend in big data. SNS and cloud data integration through an R package, and visualization methods through Wordcloud, are ideal tools for analyzing and expressing big data.

2.3 Web Crawling

As a search engine foundation, web crawler is a technology for tracking web documents on the Internet to collect necessary information (Seo, 2013), and is a core technology used by the majority of the Internet industry, including Internet search systems like Yahoo! and e-commerce product search systems (Kwon, 2003). Web crawler is a program for collecting numerous types of information posted on various websites by circulating Web servers. Instead of having to repetitively follow each website link to obtain information, the program automatically analyzes the contents of the Web pages connected to each URL.

To use web crawling, it is necessary to first select a website containing the desired information. Because the basic concept of web crawling is searching selected websites, it needs to be composed of easily accessible security for its implementation. In addition, web crawling results with a high collection rate can be obtained within a relatively short period of time when composed of a simple Hypertext Markup Language (HTML) structure. The important aspect of web crawling is determining the scope of a webpage search from which information will be collected, as well as determining the cycle of information collection.

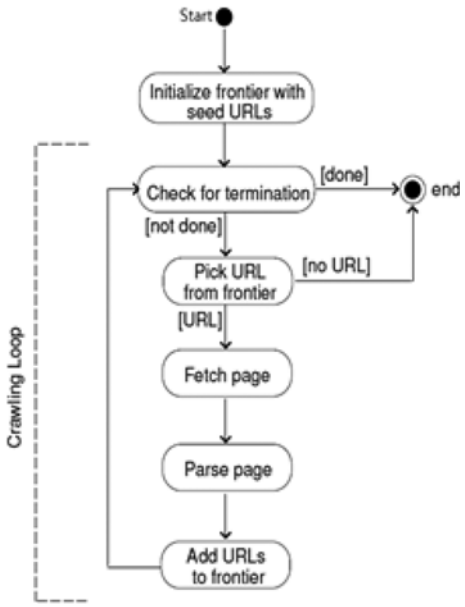

Fig. 2. Web Crawling Model (Cho *et al.*, 2013)

The use of web crawling can be divided into cases in which an entire website is selected as the search scope, and cases in which the search is centered on a particular subject or only particular Uniform Resource Locators (URLs) through the input of URL patterns. To obtain the desired information from collected URLs, a separate parsing is required. Parsing is a process of analyzing an HTML structure to allow effective programming. The final data collected are converted into a text file and saved. To implement web crawling, a core technology of the present study, a library of R, a statistical analysis program, was used.

2.4 Text Mining

Text mining refers to a process of finding useful patterns by applying an algorithm in the area of machine learning and statistics from high volumes of documents. Because data mining method is applied to text, it is also referred to as text mining. What is different from existing data mining techniques, however, is that contrary to programming and control language, non-standardized data such as commonly used human languages are not applied (Oh, 2012).

Research areas related to text mining include information searches, natural language processing, and information extraction. Data mining methods for documents include a classification method for allocating documents based on the given keywords and a clustering method for binding similar documents without advance information. Document classification methods include the use of a decision tree, Bayesian classification, nearest neighbor classifier, and Support Vector Machine (SVM). Finally, clustering methods include hierarchical clustering, K-means, SOM, and EM (Hotho *et al.*, 2005).

2.5 Data Mining

The complexity level of a data mining analysis is high because the amount of data to be processed is massive with a high percentage of a typical data. Based on the existing methods used in data mining, including statistics, the majority of analytical methods apply algorithms used in analytical methods by improving them for processing large volumes of data for a big data processing analysis. Analytical technologies for processing large amounts of data in real-time include text mining for establishing a hypothesis between the extracted information upon extracting the meaning contained in atypical sentences, opinion mining for a reputation index analysis for distinguishing user opinions on particular services and products, social network analytics for examining users centering on word-of-mouth, and a cluster analysis for deducing new user groups by analyzing the differences

between a target group with high similarity and objects belonging to other clusters. As an example of an analysis tool commercialized through packaging, Hadoop, developed by Yahoo!, is a prime example that has the advantage of quickly and reliably allowing an analysis of standardized and non-standardized data to be conducted through the application of distributed cluster technology. In addition, Big Query by Google and Dynamo by Amazon have been developed (Seo, 2015).

2.6 Principal Component Analysis, PCA

A PCA is a method for reducing the dimensionality of a dataset while maintaining variations of related variables when they compose a dataset. For example, it can be arbitrarily defined as a vector consisting of a dimension random variable found in vector, and all data can then be expressed as a matrix. The first objective of a PCA is to find the parameter that maximizes the dispersion of X under normalization, namely, finding that satisfies the following maximization problem.

The result then becomes the first principal component (Principal Component 1). Next, find vector that maximizes the dispersion that is unrelated to the first principal component. Subsequently, when it is found as the solution to the following maximization problem, then it becomes the second principal component (Principal Component 2). The principal component for subsequent problems can be found by continuously applying such a method, but it usually stops at finding the principal components in the number of variables because the objective of a PCA is to reduce the dimensionality of the data. In a PCA, it is presumed that each principal component is a type of underlying variable that cannot be observed. However, it is presumed that such principal components can be expressed through an observed linear combination of variable. In addition, principal components is very useful in that it can be considered for displaying respectively different statistical dimensions because by definition there is no correlation between the principal components (Byeon, 2010).

3. Negligent Accident Factor Analysis using Big-Data Analysis

3.1 Development of Big-Data System

Consequently, it becomes necessary to develop a system to effectively deduce factors of negligent accidents using text and data mining upon the collection of Internet articles on negligent accidents that have occurred at construction sites during the last decade. Accordingly, the open-source statistical analysis program, R, was used to develop this system, as shown in Fig. 3.

R has the advantage of allowing developers to easily obtain a new analysis method and technology based on communities with many users and various libraries supported based on their open-source characteristics.

3.2 Collection of Negligent Accident Information

To obtain desired information on the Web, a library of R was used to conduct a web crawling modeling, as shown in Fig. 4. To ensure effective information collection, the SBS NEWS website (www.news.sbs.co.kr), which has an HTML structure that can be relatively easily accessed, was chosen. To collect news information on negligent accidents that have occurred at construction sites during the last decade from the SBS NEWS website, the keywords "construction site" was used, and a total of 28,263 Internet news articles were found for the period of Jan. 1, 2000 through Dec. 31, 2014. For effective web crawling, it is also necessary to understand the URL and HTML structures.

The URL structure of SBS NEWS consists of the SBS domain, page number, and search date condition and search word code. Notwithstanding, it is possible to collect search results for Internet articles based on revised keywords without having to connect to the website by simply revising the search condition

Fig. 3. Big-Data Analysis System

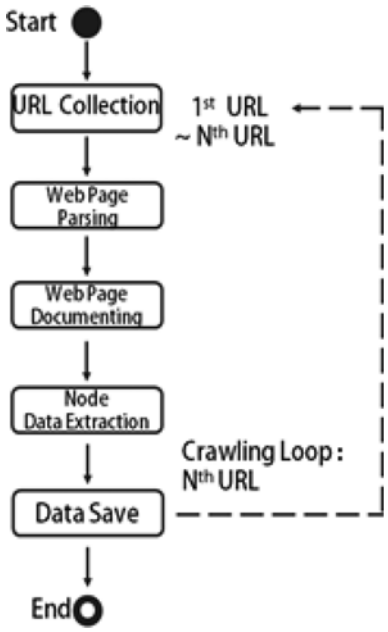

Fig. 4. Web Crawling Process

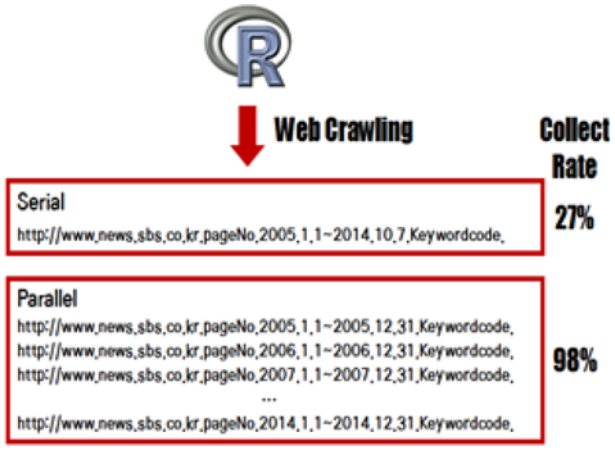

Fig. 5. Comparison of Serial- and Parallel-Structure Web Crawling

and keyword code. To obtain Internet article search results based on the same keywords throughout a period of several years, it is also possible to conduct web crawling based on the year. For collection of the information used in this study, news articles on negligent accidents at construction sites that occurred during the last decade were divided based on the year.

The reason for a high collection rate is an important area in web crawling, and as shown in Fig. 5, the collection rate of serial-structure web crawling without a yearly classification is 27%, whereas that of parallel-structure web crawling with yearly classification is 98%. For this reason, a parallel-structure web crawling model was used in this study, through which 18,490 out of 18,868 Internet news articles were collected, which represents a 98% collection rate. In addition, the articles were saved into text files according to the year of publication.

3.3 Text Mining

3.3.1 Accident Type

News articles collected through web crawling are composed of sentences. Using the R library and count feature, the articles were divided into words, the frequency of which was obtained. By referring to a study by Kim *et al.* (2010), the types of accidents occurring at construction sites were divided into collapses, heavy equipment problems, suffocation, falls, explosions, and fires. The remaining words were considered as factors causing negligent accidents at construction sites. Table 1 shows the frequency of words mentioned, and as the table demonstrates, the frequency of accident types mentioned increases in the period closer to the present date, which is considered to be a result of an increase in the amount of information being recorded through the advancement of media and telecommunication technologies, instead of an increase in the frequency of accident occurrence.

In Table 1, the most frequently occurring negligent accidents at construction sites during the last decade were fire-accidents, which were mentioned 1,077 times, representing 33.11% of the total number. However, falls made up 72% of the total number in the Industrial Disaster Status Report by the Ministry of Employment and Labor, whereas fire-accidents made up only 1%. The reason for such contrasting results may be that falls, which frequently occur, do not receive as much media attention as fires, which are usually at a larger scale and incur heavier damage. Hence, it was determined that fires would be a suitable subject for effectively collecting information and analyzing the various factors in the present study.

Table 1. Types of Accidents for 2005-2014

<SBS NEWS>

	Suffocation	Heavy Equipment	Fall	Explosion	Collapse	Fire
2005	3	28	7	22	23	21
2006	4	14	16	9	28	45
2007	5	76	49	14	77	40
2008	5	33	17	67	34	194
2009	5	30	24	21	112	65
2010	0	56	26	74	54	123
2011	9	42	46	48	116	54
2012	10	26	81	72	60	133
2013	31	31	102	38	142	168
2014	15	18	41	73	241	234
	87	354	409	438	887	1077
%	2.67	10.88	12.57	13.46	27.27	33.11

<Industrial Disaster Status Report, Ministry of Employment and Labor>

	Suffocation	Heavy Equipment	Fall	Explosion	Collapse	Fire
2005	44	1,737	5,260	204	320	61
2006	29	2,274	5,873	240	357	79
2007	45	2,744	5,950	279	319	127
2008	47	1,585	6,976	144	497	78
2009	52	1,943	6,742	104	570	81
2010	74	2,150	7,322	92	516	80
2011	48	1,917	7,489	64	452	72
2012	71	1,820	7,734	70	411	118
2013	67	1,763	7,682	77	468	62
2014	61	2,045	7,908	71	308	91
	538	19,978	68,936	1,345	4,218	849
%	0.6	20.8	71.9	1.4	4.4	0.9

3.3.2 Factor Dictionary

In terms of the type of information that can be obtained through text mining, along with the types of negligent accidents occurring at construction sites, the factors causing such accidents were obtained. A news article can be characterized in that it factually records the exact account of an accident. News articles became an important database in the present study on developing the Factor Dictionary because they include the type and frequency through which the factors causing an accident can be examined. The Factor Dictionary is an aggregate of factors with high frequency that have direct/indirect effects on accidents occurring at construction sites.

Table 2. Factor Classification of Factor Dictionary

Classification1	Classification1	Factor
Natural	Season	Spring, Summer, Fall, Winter
	Weather	Heat Wave, Cold, etc.
Time	Day of the Week	Mon, Tue, Wed, etc.
	Time of Occurrence	Dawn, Morning, Afternoon, etc.
Spatial	Location of Occurrence	Road, Apartment, etc.
	Place of Occurrence	Slant, Slope, etc.
Worker Behavior	Work	Depositing, Construction, etc.
	Task	Enforcement, Design Change, etc.
	Physical Change	Anxiety, Plea, etc.
Site Condition	Gas	Gas, Carbon Dioxide, etc.
	Type	Sheeting, Concrete, etc.
	Dynamics	Torsion, Concentration, etc.
	Occurrence	Spark, Short Circuit, etc.
Heavy Equipment	Heavy Equipment	Truck, Crane, etc.
Other	Other	Instant, High, etc.
No. of Factors: 170		

It consequently helped us prepare the standard and scope of such factors when analyzing negligent accidents at construction sites. Table 2 shows the classification of factors in the Factor Dictionary. The *Natural* classification was divided into the season and weather at the site of interest. The *Time* classification was divided into the day of the week and time of occurrence. The *Spatial* classification was divided into place of occurrence and location. The *Worker Behavior* classification was divided into work, task, and physical change. The *Site Condition* classification was divided into gas, type, dynamics, and occurrence. The *Heavy Equipment* classification was divided into the types of heavy equipment. The remaining were classified as *Other* factors.

3.4 Collection of Fire-Accident Information

Based on the text mining conducted earlier, it was found that fire-accidents were the most frequently occurring negligent accidents at construction sites.

Accordingly, using web crawling for Internet news articles, this study aims to deduce through a PCA the various factors of fire-accidents occurring at construction sites during the last decade. Considering that accident-causing factors vary depending on the season, even for the same type of accident, Internet news articles on fire-accidents were collected based on the season and year. The reason for classifying Winter 1 independently from Winter 2 as part of the search conditions is based on the idea that the accident-causing factors will differ because Winter 1 was a thawing season and Winter 2 was a freezing season.

In addition, the URL structure shown in Table 3 was entered and the yearly season-specific fire-accident keywords were used to search 12,019 Internet articles, among which 10,978 were collected, which represents 91.3% with the parallel-structure model, as shown in Fig. 6. The reason for this rate is that the web crawling model used in the present study inherently has a 2% collection omission rate, thereby causing a cumulative omission rate.

Table 3. Accident Type-Specific Search Method According to Season

Fire-Accident	Search Condition		
URL Structure	Year	Season	
http://www.news.sbs.co.kr.pageNo. search condition: construction site fire-accident code	05	Winter 1	Jan-Feb
	06		
	07	Spring	Mar-May
	08		
	09	Summer	Jun-Aug
	10		
	11	Fall	Sep-Oct
	12		
	13	Winter 2	Nov-Dec
	14		

Fig. 6. Filtering Diagram

3.5 Fire-Accident Text Mining

The 10,798 Internet news articles on construction sites obtained through the web crawling were again summarized according to the season and processed into words suitable for data mining.

Using the R library and its features, Internet news articles composed of sentences were classified into words through a morphological analysis and counted to measure the frequency of the fire-accident factors. Because fire-accident factors (classified into words) include those directly and/or indirectly involved in causing a fire-accident and those hindering an analysis, the Factor Dictionary was used for filtering.

Accordingly, the fire-accident factors were saved into a Comma Separated Value (CSV) file suitable for programming, and the MS Excel built-in function was used to deduce the words considered to represent the direct and/or indirect factors of fire-accidents according their frequency. As for the filtering principle, corresponding words between the words included in Factory Diction, which is an aggregate of negligent accident causing factors at construction site, and fire-accident factor words are considered to be fire-accident causing factors. In addition, unaccounted words and words with low cumulative frequency were removed to obtain favorable values in the Kaiser-Meyer-Olkin measure of sampling adequacy test, which showed a significant correlation with the collected data. The elimination criterion is that the word with the cumulative frequency of less than 90% is removed after the words with high frequency were listed in order. The factors deduced through text mining are summarized in Tables 4 through 8.

Table 4. Principal Components of Construction Site Fire during Jan-Feb

Division	Factor	Principal component		
		1	2	3
Time	Dawn	0.78		
	Morning	0.75		
	Afternoon	0.75		
Gas	Gas	0.81		
	Toxic gas	0.94		
Occurrence	Spark	0.95		
	Explosion	0.88		
	Short Circuit	0.96		
Task	Illegality	0.72		
	Construction	0.95		
	Work	0.83		
Heavy equipment	Equipment	0.75		
	Machine	0.89		
Type	Facilities	0.96		
Location	Subway	0.73		
Time	Morning		0.83	
Weather	Weather		0.79	
	Below zero		0.82	
	Dryness		0.89	
Location	Apartment		0.87	
Type	Fire protection system			0.95
Location	Tower			0.77
	Site			0.91
Eigenvalue		2.60	2.49	1.80
Cumulative dispersion ratio (91%)		76%	10%	5%
KMO		0.64		

Table 5. Principal Components of Construction Site Fire during Mar-May

Division	Factor	Principal Component		
		1	2	3
Time	Morning	0.96		
	Afternoon	0.89		
Gas	Toxic gas	0.97		
	Gas	0.95		
	Smoke	0.92		
Occurrence	Spark	0.92		
Task	Work	0.85		
	Control	0.94		
	Man-made	0.98		
Location	Site	0.70		
	Subway	0.88		
Task	Renovation		0.97	
	Illegality		0.91	
Weather	Wind			0.77
Task	Reconstruction			0.79
Type	Facilities			0.81
Eigenvalue		4.59	2.54	1.36
Cumulative dispersion ratio (81%)		46%	23%	12%
KMO		0.61		

3.6 Fire-Accident Data Mining

In the data obtained through fire-accident text mining, the season-specific fire-accident causing factors and frequency are included. Because high-frequency factors were not considered sufficiently convincing to be regarded as the causes of a fire-accident, a PCA was used to deduce the principal components of a season-specific fire-accident. Therefore, the validity of the data

Table 6. Principal Components of Construction Site Fire during Jun-Aug

Division	Factor	Principal Component		
		1	2	3
Time	Morning	0.78		
Gas	Gas	0.93		
	Smoke	0.85		
Occurrence	Spark	0.82		
Task	Work	0.91		
	Construction	0.90		
	Complete	0.74		
Type	Machine	0.84		
Task	Illegality		0.55	
Location	Subway		0.53	
Time	Afternoon			0.76
Location	Apartment			0.56
Eigenvalue		2.37	1.95	1.53
Cumulative Dispersion Ratio(85%)		52%	21%	12%
KMO		0.60		

Table 7. Principal Components of Construction Site Fire during Sep-Oct

Division	Factor	Principal Component		
		1	2	3
Gas	Gas	0.84		
Occurrence	Collapse	0.95		
	Damage	0.98		
Task	Completion	0.98		
	Illegality	0.72		
Location	Roof	0.93		
	High Building	0.98		
Occurrence	Crack		0.98	
	Hazard		0.98	
Task	Construction		0.98	
Location	Tower		0.93	
Time	Morning			0.84
Eigenvalue		6.86	5.14	3.61
Cumulative Dispersion Ratio(81%)		48%	24%	9%
KMO		0.60		

Table 8. Principal Components of Construction Site Fire during Nov-Dec

Division	Factor	Principal Component		
		1	2	3
Time	Morning	0.86		
Gas	Gas	0.76		
	Toxic Gas	0.86		
Weather	Dry	0.78		
Occurrence	Collision	0.92		
	Cigarette fire	0.95		
Task	Construction	0.78		
Location	Sandwich panel	0.96		
	Tower	0.83		
Task	Control		0.93	
	Reconstruction		0.74	
	Man-made		0.72	
Location	Structure		0.72	
	Temporary building		0.76	
	Subway		0.82	
Time	Afternoon			0.76
Weather	Wind			0.81
Task	Illegality			0.74
Eigenvalue		4.38	2.09	1.20
Cumulative dispersion ratio (83%)		37%	30%	16%
KMO		0.65		

should be verified before the principal component analysis. In this study, we verified the usefulness of the collected data by using the sample fitness measure with the minimum standard of 0.6. Then rotated the Euclidean coordinate axes to see clearly the relationship between the data and the extracted principal components. Varimax Rotation, which is an orthogonal rotation that can be easily analyzed and avoids multi-collinearity problem, is applied. There are several methods for selecting principal components, but the kaiser's high value 1 or more and the cumulative dispersion ratio are used as a criterion. In this study, principal component results for each season are shown in Table 4 ~ 8, showing only significant factors for principal components respectively. In addition, the cumulative dispersion ratio is more than 80%, indicating that seasonal fire accidents are more explanatory.

The first components of fire accidents during the January to February construction period were dawn, morning, and afternoon, and gas, toxic gas, and flames were also major factors in the fire. Also, it was found that the fire occurred on construction sites, consumed heavy equipment, facilities and machines. The temporal factors of the second main component were morning, dry and subzero weather, and it was found that it occurred mainly in the apartment. The third factor is the construction site or tower facilities with insufficient firefighting facilities.

The time factors of the first main component of the fire safety accident during the construction period from March to May are morning and afternoon, and fire are caused by gas or flame when the control is not smooth and when working on construction site and subway. The second main ingredient is illegal and unreasonable modification of the construction site. The third main component appeared to be the aging of facilities at the redevelopment site.

The temporal factor of the first main component of the fire safety accident in the construction site during the period from June to August was morning and it was found that the fire occurred due to the gas, flame and machine at the work. The second main ingredient was illegal at the subway construction site, and the third main ingredient appeared at the apartment site in the afternoon.

The first main ingredient in September and October was the illegal work of the roof and high-rise buildings, damage or collapse due to unreasonable completion, and gas. The second main component appeared as cracks and defects at the construction site of the tower, and the time factor of the third main component appeared in the morning.

The first major component of the fire safety accident at the construction site in November and December was the tower construction site, sandwich panel, dry weather, cigarette, gas, and impact. The second main component is the unstable control of the tall buildings and subways. The third major component appeared to be illegal acts and winds in the field.

The main features of each season are summarized as follows: In all seasons, it is found that a fire occurs due to the contact between gas and flame. In the case of spring, summer, and autumn, where the construction is actively underway, it was found that a fire occurred because of insufficient construction or on-site control.

In addition, it was found that the fire was generated because the management of firefighting and facilities at the redevelopment site was insufficient. As a result, it was found that fire accidents occurred in the construction site due to the lack of supervision by the supervisor and violation of safety rules of the workers.

4. Conclusions

Using a big-data analysis method, which is receiving significant attention in all areas of industry and academics, this study collected news information on fire-accidents that occurred during the last decade according to season, and a PCA was applied to deduce the seasonally specific factors of a fire-accident. Common factors were direct gas and flame in the fire, and seasonal factors were subzero temperature and dry weather in winter. And it was found that the work factors were caused by the supervisor's negligence and violation of safety regulations. In addition, it was found that the fire and the management of the facilities were insufficient and the fire occurred. It is thought that effectively utilizing a big-data analysis method based on this study will allow negligent accident factors to be analyzed from various angles. Just, this research has the limitation that can't know intensive information of fire accident and can't verify the effectiveness of web-crawling by compare to accident statistics.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. NRF-2014R1A2A1A11051065).

References

Bae, B. G., Bae, J. H., and Choi, S. H. (2014). "A study of analysis based on social big data for disaster prediction." *Proc., of the Korean Society for Geospatial Information System Conference*, Vol. 2014, pp. 273-274, DOI: 100002372.

Byun, Y. G. (2010). "Measuring labor market flexibility and security based on principal components." *Journal of Korea Labor Institute of Korea*, Vol. 10, No. 4, pp. 1-37, DOI: 82525464.

Cho, W. S., Lee, J. E., and Choi, C. H. (2013). "Refresh cycle optimization for web crawlers." *Journal of Korea Contents Association*, Vol. 13, No. 6, pp. 30-39, DOI: 99761699.

Choi, S. H., Bae, B. G., and Lee, B. R. (2015), "The sensing model of disaster issues based on relevance to disaster from social big data." *Proceedings of Korea Institute of Information Scientists and Engineers, Korea*, Vol. 2015, No. 6, pp. 4-6, DOI: 100640256.

Hotho, A., Nurnberger, A., and Paab, G. (2005). "A brief survey of text mining." *Journal for Computational Linguistics and Language Technology*, Vol. 20, pp. 19-62.

Kang, S. H. (2015). "Big data in construction industry." *Journal of Korean Society Civil Engineers,* Vol. 63, No. 7, pp. 14-20, DOI: 100561706.

Kim, J. W., Kim, Y. H., Kim, J. H., and Kim, J. J. (2010). "A study on the methods of reducing the safety accident through pattern analysis of construction disasters." *Proceedings of Korea Institute of Building Construction, Korea*, Vol. 10, No. 2, pp. 137-140, DOI: 82428238.

Kim, K. T. (2009). "A study on the Implementation of USN technologies for safety management monitoring of architectural construction sites." *Journal of Korea Institute of Building Construction of Korea*, Vol. 9, No. 4, pp. 103-109, DOI: 76531452.

Korea Occupational Safety & Health Agency (2015). *2014 Industrial Disaster Cause Investigation Report (Accident for Works).*

Kwon, S, H., Lee, Y. T., Kim, Y. J., and Lee, Y. J. (2003). "Design and implementation of a high performance web crawler." *Journal of Korea Society of Industrial Information Systems,* Vol. 8, No. 4, pp. 64-72, DOI: 60301415.

Lee, H. J. (2012). "Take advantage of Big Data Hadoop platform." *Journal of Korean Institute of Communications and Information Sciences of Korea*, Vol. 29, No. 11, pp. 43-47, DOI: 100391834.

Nam, K. S. (2015). "Highway Traffic openness and take advantage of Big Data." *Journal of Korea Research Institute of Human Settlement of Korea*, Vol. 7, No. 405, pp. 25-30.

OH, S. W. (2012). *Text Mining Study on the practical utilization*, Master Thesis, Korea University, Korea.

Oh, Y. T. (2015). *Problem Analysis and Response of Securing Fire Protection of Apartment Building Projects,* Master Thesis, Kyung-Hee University (in Korean).

Park, H. J., Choi, B. G., Na, Y. W., and Roh, S. L. (2014). "Method of environmental complaints analysis using spatial big data mining in Incheon." *Proceedings of Korean Society Civil Engineers,* Vol. 2014, No. 10, pp. 619-620.

Seo, D. M. and Jung, H. M. (2013). "Intelligent web crawler for supporting big data analysis services." *Journal of Korea Contents*, Vol. 13, No. 12, pp. 575-584, DOI: 99862503.

Seo, J. H. (2015). *Design of Opinion Sensitivity Dictionary Model for Big Data Management*, PhD Dissertation, Incheon University, Korea.

Shin, W. C. (2015). "The characteristics on construction site fire that caused fatalities and Injures." *Proceedings of Korean Institute of Fire Investigation,* p. 249.

Shin, W. C., Kwon, J. H., and Lee, Y. S. (2014). "Case study on prevention of Fire/Explosion accidents caused by chemical substances in Small/Medium sized construction sites." *Journal of Korea Safety Management Science*, Vol. 16, No. 3, pp. 63-70, DOI: 10.12812/ksms.2014.16.3.63 (in Korean).

Yong, Y. H. (2012). "Era of big data." *Journal of Advanced Information Technology and Convergence*, Vol.10, No. 3, pp. 39-43.