Measure for the Improvement of Constructio Work Accident Informaiton Service Contents in CPMS

Focused on Analysis of Construction Work Accidents Big Data

Sung-Hoon Yang, Jin-Uk Kim, Young-Jin Kim, Hyun Ok
ICT Convergence and Integration Research
Korea Institute of Civil Engineering and Building Technology, KICT
Goyang-si, Gyeonggi-do, Korea
{ispecter, jukim, yjkim73, okhyun} @kict.re.kr

***Abstract*—This paper analyzes 14,578 cases of construction industry accidents based on big data for the improvement of the construction work accident information service contents of the Construction Project Management System ("CPMS"). The data were gathered from South Korea's major portal site that provides the accident causes, workers' behavior, injury areas, and injury factors. The data were surveyed in the following order. Factors (words) were extracted and classified into various categories, factors with predicates and adverbs were re-extracted, the target data were verified, factors were standardized and classified, and the rate and correlational rate by factor were analyzed. The analysis revealed that crash accidents involving high-place work accounted for 47% of all the accidents; arm and leg injuries, 29% and 26%, respectively; and bone fractures, 47%. For the injury causes by injury factor, crashes from ladders and roofs occurred most frequently; and for the injury areas by injury factor, leg injuries associated with materials, reinforcing bars, and excavators occurred most frequently.**

Keywords-CPMS; Big Data, Construction Industry Accidents

I. INTRODUCTION

A. Research Necessity and Objectives

With the spread of Internet use since 1990, ICT and ICT device technologies have rapidly developed, and the digital information production and acquisition methods are increasingly shifting from spatial environments to time environments. Notably, mobile ICT devices, social networks, and wireless Internet technologies are optimizing the use of any form of data anytime and anywhere, and the quantity of the data being processed is also rapidly surging, so by 2020, the quantity of worldwide digital data is estimated to reach approximately 40 zeta bytes [1]. Such surging data have brought about a shift to the existing data processing technology, creating the term big data. Big data, based on the data size, speed, diversity, accuracy, and values, are emerging as an attention-arresting issue not only for the ICT industry but for all other industries in Korea and abroad.

US federal agencies have invested USD200 million in formulating big-data response plans, and have opened up the public data of the federal government to the general public to help create business opportunities. The U.K. has discussed measures for public-data convergence and use, and has pursued the "Show us a better way" project under which it will open medical, education, tax, employment, and weather data, among other data [2].

South Korea’s Ministry of Land Infrastructure and Transport to be able to provide spatial big-data-based smart administration and tailored people services, has established the Spatial Big Data Utilization Support Center, through which the ministry will prepare a system designed to combine, jointly use, and provide public administration information, real-time situation information, humanities and social information, and private-sector information [3]. With the government announcing policies for the use of big data, diverse sectors are researching on big data. The construction sector is also pushing ahead with big data development plans to analyze and use the construction project life-cycle information data in the construction project information system, which consists of five unit systems (the construction project information portal, construction project management, construction approval, facility maintenance, and land compensation). Of these systems, the Construction Project Management System (CPMS) is the main system for linking information between construction sites and construction project orderers, and handles all information, from construction project order information to project completion information. However, due to the poor information input environments of construction sites and field managers' low information evaluation skills, some information data are not inputted or skipped, which is lowering the quality and use of the system-providing content. Notably, construction-site accident information are typical examples of poor information, to the extent that of 545 construction work accidents from 2000 to 2015, only 40 cases have been updated as to their accident circumstances and follow-up information. As such, the information quality needs to be enhanced. Against such backdrop, this study seeks to gather inquiries on construction site accidents registered in the country's large search portal sites, to analyze the accident causes, types, and injury areas based on big data, to promote understanding and use of social data in the construction sector, and to improve the construction work accident information service contents of the CPMS, which manages domestic construction work life cycle information.

978-1-4673-9795-7/15 $31.00 © 2015 IEEE
DOI 10.1109/CSCI.2015.51

B. Research Outline and Methodology

This study gathered inquiries on construction work accidents, registered on a domestic large portal site, for 10 years from 2005 to 2014, and analyzed the accident causes, behavior, types, and injury areas, considering that the Internet search portal, which targets Internet users, focuses more on accident occurrence rates by type rather than on statistics on the number of accidents. This portal site was selected because such data could be gathered as they are, without needing to be processed and edited, and because the rate of duplication of accidents and incidents is lower than in other news media, thereby enhancing the accuracy of the analyzed data. Moreover, the HTML structure of the website pages is simpler, which made it easy to access and gather the target data.

The research methodology consisted of four stages. First, the target sources and keywords were selected to gather the target data. Second, the duplication of the gathered webpage data and effective data were sorted out. Third, through text-mining, effective texts were extracted and a vocabulary dictionary was created. Fourth, based on the vocabulary dictionary, effective data were verified and content services were provided.

II. Preliminaries

A. Construction Site Safety Accidents

Construction businesses and technologies have advanced alongside the development of information technology, and construction sizes are also increasingly becoming massive, from individuals' houses to large public and private structures. Nonetheless, construction site accidents happen year after year and corresponding economic losses have reached KRW4.8 trillion (25% of the total industry loss of KRW19 trillion). These problems are occurring not only in South Korea but also elsewhere around the world. Diverse studies are being conducted along with the proposal of solutions to prevent construction site accidents and incidents [4].

Hong Kong's Labor Department are analyzing, arranging, and distributing cases of construction site accidents to sensitize construction workers and managers on the need for construction site safety. The data classify the causes of construction accidents into falls from high places, falls of objects, neglect of materials, narrow spaces, and electrical management failures, and present illustrations of accident types and prevention methods according to detailed works, thereby promoting awareness of safety accidents [5].

G. Karunasena has indicated that Sri Lanka has a high rate of construction industry accidents because time and cost are emphasized more than safety management. He has proposed a framework for estimating and evaluating safety management. Towards that end, he surveyed six construction specialist groups, derived safety management factors, and analyzed inter-factor correlations using the AHP technique. He further proposed a new safety management framework designed to monitor risks of construction site accidents [6].

J. Dumrak examined 24,764 cases of construction work accidents in southern Australia from 2002 to 2011 to analyze the seriousness of the accidents by worker, time, work type, equipment, and injury area. The he derived the construction accident factors and proposed preventive measures [7].

H. Lubega examined construction projects in five regions in Uganda to analyze the causes of accidents, and proposed a safety program to prevent accidents. He indicated eight major accident causes as detailed factors, including supervisors' inappropriate management techniques, unqualified engineers, inappropriate construction methods, and lack of awareness of safety rules. He thus proposed nine improvement methods, which include the amendment of existing safety regulations, to minimize construction work accidents and prevent their recurrence [8].

To analyze the major causes and factors of fall accidents in construction sites in the U.S., Sohaib Siddiqui used OSHA's IMIS to gather and analyze 9,141 cases of fall accidents in construction sites over 20 years. Among the causes he found were low-budget projects, frequent fall accidents in residential housing and commercial building construction sites, and inappropriate fall prevention facilities in under 30-feet-deep construction works. Also, the fall accidents that happened during roof, ladder, and scaffolding works accounted for 55% of all the construction industry accidents [9].

Domestically, various studies on construction industry accidents have been conducted. In such studies, accidents were described according to the manager-worker relationship status and patterns, construction site safety education problems, measures for formulating and implementing plans for stacking of materials and safety activities, and types of construction accidents. However, they analyzed mainly construction site accident types, etc.; and unlike overseas studies, which proposed accident recurrence prevention measures, they only mentioned cases of deaths using the industrial accident statistics provided by the Korea Occupational Safety & Health Agency, but did not properly deal with details on accident causes and types, injury areas, etc. Thus, this study gathered and analyzed the information on the construction industry accidents based on big data, and provided rates for the causes, types, and injury areas, as well as the inter-factor relations, as CPMS construction work accident information contents.

B. CPMS Construction Accidents Information Services

CPMS is designed to manage all life-cycle information on domestic construction works, and to link the information to clients and construction sites. The total quantity of the data registered with the system reaches a huge 50 GB per month, but the quantity of the accident information data stands at a mere 100 KB, and the relevant services are nearly unused. And CPMS construction work accident information service, which shows the names of the construction projects with accidents by construction site, the accident occurrence date, the number of accident cases, the accident causes, and the action taken, but does not provide diverse information, including on the type, area, and seriousness of the injury. Due to the failure of field staff' to input such information and the lack of such staff, even basic information is hardly provided. This situation needs to be rectified [10].

III. ANALYSIS OF CONSTRUCTION INDUSTRY ACCIDENTS

A. Gathering of Construction Industry Accident Data

To accurately analyze construction work accident data, sources of accident-time situations, types, etc. are needed. Relevant information is partially provided by various Internet news and newspapers, individual blogs, SNSs, construction-related associations, and Statistics Korea. However, Internet news and newspapers present only industrial accidents in large construction projects or medium- and large-scale industrial accidents, and limit their analysis of data to a certain category. Also, statistics provided by various relevant associations, Statistics Korea, etc. focus on deaths, seriousness of injuries, work types, etc., and it is difficult to acquire detailed information on the project sizes, periods, numbers of casualties, accident causes and types, injury areas, and other details from portal sites, depending on the site's characteristics. As an information data source, this study selected South Korea's large search portal site, Knowledge IN, which mentions the construction time accident situations, causes, types, seriousness of injuries, areas of injuries, etc. Knowledge IN, which is available to all Internet users, provides a Q&A on diverse issues, including social, political, and economic problems, tastes, religions, and foods. As such, the site provides a diverse and huge amount of information.

To gather a wide range of data, this study selected the “*site*&*construction*&accident” and searched for diverse information by including searches for words encompassing the symbol ‘*’ before and after the keywords accumulated for 10 years from 2005 to 2014. As a result, data from 14,578 cases, consisting of Webpage 1 (the subject, writing date, summary, and category) and Webpage 2 (the subject, writing date, details, experts' replies, etc.), were searched for; and based on these data, crawling work was performed. Crawling is a technique designed to access Internet web pages to automatically gather necessary information. In this study, the script tool “Tool bar” was used to gather data from 14,286 cases; rearrange the data according to their writing date, subject, and details; and save them in text-format "csv" files.

B. Construction Industry Accident Data-mining

The data gathered through crawling, including on traffic accidents and construction fraud, which are irrelevant to construction industry accidents needed to be removed through data-mining. First, to sort out data irrelevant to the target data, target-related keywords were selected, and the gathered data and keywords were compared to count the number of retrieved keywords. More frequent retrieval of a word meant the word was closer to the target, and contents with a count value of 1 or higher were classified as target data. Next, when data with the same count value were compared as to their subjects and text amounts, if the value was 2, the data was considered 100% duplicated, and if the value was 1, the data was re-verified to delete duplicated data. The target data thus sorted out totaled 6,483 or 45.38% of the total data gathered.

C. Construction Industry Accident Text-mining

Text-mining work involves extracting construction industry accident factors (words) from target data, and creating the vocabulary dictionary needed to re-verify the target data. First, using the R tool, words, from which conjunctives, adverbs, and predicates had been deleted, were extracted, and words equivalent to injury areas, symptoms, injury causes, and injury factors were re-classified. Likewise, the R tool, given its nature, cannot recognize grammatical errors (spacing) or words unlisted in the dictionary (dialect, mistyping, vulgar words, etc.), so re-classified words were compared with the target data gathered through data-mining, so as to add words not searched through R to the dictionary. Alongside this, to distinguish passive from active behavior, predicates and post-position particles were added to the words, and ungrammatical words were transformed into standard words and classified, thereby boosting the accuracy of the vocabulary dictionary Data Analysis.

TABLE I. WORD CONVERSION AND CLASSIFICATION TABLE

Category	Classification & Transformation	Words Subjected to Transformation
Injury area	Head	Forehead, eye, nose, mouth, ear, etc.
	Neck	Neck, cervical vertebra, etc.
	Chest	Ribs, thorax, pneumothorax, etc.
	Back	Spine, thoracic vertebra, etc.
	Arm	Shoulder, hand, elbow, etc.
	Abdomen	Belly, etc.
	Waist	Lumbar vertebra, backbone, etc.
	Hip	Tailbone, pelvis, hip, etc.
	Leg	Thigh, shin, foot, etc.
Symp-tom	Fracture	Broken, crushed, cracked bone, etc.
	Brain disease	Cerebral hemorrhage, cerebral infarction, etc.
	Paralysis	Palsy, unconsciousness, etc.
	Death	Fall death, electric shock death, etc.
	Vision loss	Vision loss
	Lacerated wound	Skin tear and bursting
	Sprain	Extended or torn ligament, muscle, etc.
	Cut	Cut body part
	Bruise	Skin scratch, bruise, or edema
	Burn	Wound from heat
	Cut injury	Cut from sharp materials and tools
Injury cause	Electric shock	Electric shock from high voltage, electricity, etc.
	Fall	Fall of a material, tool, etc.
	Cutting	Cut from a material or temporary facility
	Burial	Being buried in soil, roof, etc.
	Collapse	Collapse of soil, roof, etc.
	Falling over	Falling or tripping over, or crashing against, a target or material
	Overturning	Overturning of a heavy equipment, etc.
	Heavy object	Lifting or moving of a heavy object
	Pricking	Being pricked by a sharp material, etc.
	Fall	Drop to a lower place from a high place
	Collision	Being hit by a material, column, heavy equipment, etc.
	Explosion	Explosion of an explosives, ga, etc.
	Being caught	Being caught between or crushed by materials, heavy equipment, etc.
Injury factor	Poclain	Excavator, excavator track, excavator bucket, etc.
	Sash	Sash, etc.
	Foothold	Scaffold, etc.
	Etc.	Etc
Target	Worker	Worker, father, mother, laborer, etc.

Using the R tool, the factors classified and grouped by item were counted by year, item, and factor, and the total factor appearance frequency by item (upper 10 factors) and the relationships between the factors were analyzed. First, for the appearance frequency, the arms (area of injury) showed the highest frequency (29%), followed by fractures (symptoms) (47%), fall (cause of injury) (47%), and materials (injury factor) (8%).

TABLE II. RATES OF CONSTRUCTION INDUSTRY ACCIDENTS BY ITEM

Injury Area	Head	Neck	Chest	Back	Arm	Abdomen	Waist	Hip	Leg	
	16%	9%	5%	4%	29%	1%	7%	3%	26%	
Symptoms	Fracture	Brain disease	Paralysis	Death	Lacerated wound	Sprain	Cut	Bruise	Burn	Vision loss
	47%	3%	2%	9%	11%	14%	4%	6%	1%	1%
Injury cause	Electric shock	Fall (Object)	Cutting	Collapse	Falling over	Overturning	Pricking	Fall (Human)	Collision	Being caught
	1%	11%	1%	1%	16%	1%	1%	47%	12%	6%
Injury factor	Object	Foothold	Ladder	Material	Roof	Vehicle	Iron	Crane	Pipe	Excavator
	3%	2%	5%	8%	3%	3%	4%	6%	3%	3%

Furthermore, the graph in Fig. 1. shows the relationship between the injury area and the symptom according to the causative behavior of the injury factor and of the relationship. Then the factors with a high appearance frequency (materials, cranes, and ladders) were analyzed.

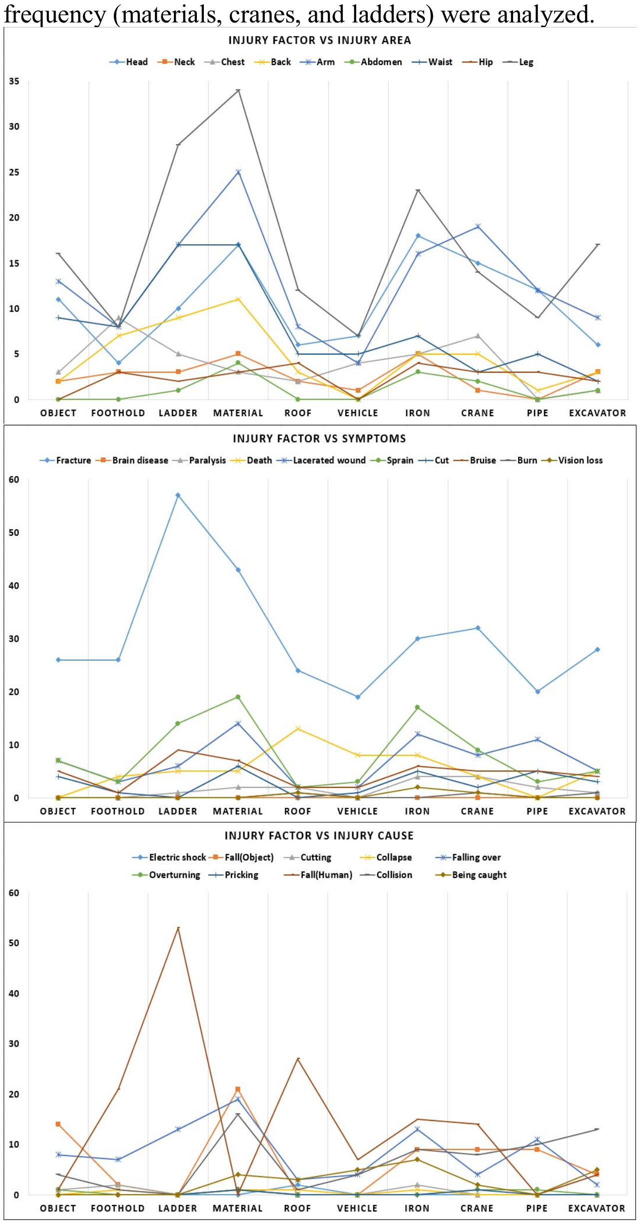

Figure 1. Relationship between Areas, Symptoms, Cause by Factor

First, for the causative factor materials, falls, overturns, and collisions occurred with a high frequency. Due to these, injuries in the leg, arm, head, and waist occurred often, such as fractures, sprains, and lacerated wounds. For the crane factors, crashes, falls, and collisions occurred in that order according to their frequency magnitude, and injuries in the arm, head, and leg, and fractures, sprains and lacerated wounds occurred with a high frequency. Finally, for the ladder factors, the rate of crashes and overturns was high; for the injury areas, the leg, arm, waist, and head sustained injuries most significantly, in that order; and for the types of injuries, fractures, sprains, and bruises occurred most frequently. In conclusion, for construction site accidents, crashes, falls, and overturns dominated the causes of the accidents, and fractures mainly in the leg, arm, head, and waist occurred most significantly. Also, the main factor and cause of death were falls from roofs, which suggests that managers and workers have low awareness of safety management (for works in not very high places) and thus are negligent, and that safety facilities are not properly secured.

IV. CONCLUSION

To boost understanding of the use of social media data in the construction field, and to improve CPMS construction work accident information services, this study proposed measures for improving the contents thereof. Towards this end, 6,483 construction industry accidents from a big domestic portal site were analyzed based on big data. The analysis showed that fractures in workers legs and arms occurred most frequently due to construction site factors involving worker falls and material falls. The death rate was the fourth highest of the total injury factors, and the danger was highest mainly on roofs (in falls), followed by motor vehicles (in collisions and being caught) and reinforcing steel bars (in falls).

REFERENCES

[1] IDC, "The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things", April 2014.*(references)*

[2] Dong-Hee, Shin, Jea-Kil, Lee,"Big Data Trends and Implications", Journal of Korean Society for Internet Information, vol. 14, no. 2, pp. 5-17, Jun 2013.

[3] Min-Soo, Kim, "Korea and Abroad of Space Big Data Policy and Technology Trends", Journal of Korea Research Institute for Human Settlements, vol. 389, pp. 30-39, Mar 2014.

[4] Ha-Joong, Yun, "A Study on Ways to Improve Construction Safety Management for a Safer Society", Brief of Korea Research Institute for Human Settlements, No. 520, Jun 2015.

[5] Occupational Safety and health Council, Hong Kong,"Brief Analysis of Site Accident Cases", Feb 2013.

[6] K. Priyadarshani, G. Karunasena, S. Jayasuriya,"Construction Safety Assessment Framework for Developing Countries: A Case Study of Sri Lanka", Journal of Construction in Developing Countries, vol. 18, no. 10, pp.33-51, 2013.

[7] J. Dumrak, S. Mostafa, I. Kamardeen, R. Rameezdeen,"Factors Associated with the Severity of Construction Accidents: The Case of South Australia", Journal of Construction Economics and Building, Vol. 13, No. 4, 2013.

[8] H. Lubega, B.M. Kigundu and D. Tindiwensi,"An Investigation into the Causes of Accidents in the Construction Industry in Uganda", 2nd International Conference On Construction in Developing Countries, pp.1-12, 2000.

[9] S. Siddiqui, "U.S. Construction Worker Fall Accidents: Their Causes And Influential Factors", FIU Electronic Theses and Dissertations. Paper 1157, Feb 2014.

[10] Construction Project Information System, "*http://www.calspia.go.kr*