Linked Data System for Sharing Construction Safety Information

Akeem Pedro [a], Do Yeop Lee [a], Rahat Hussain [a] and Chan Sik Park [a]

[a] Department of Architectural Engineering, Chung Ang University, Seoul, South Korea
E-mail: lanrepedro3@gmail.com, cpark@cau.ac.kr

Abstract –

Despite the consistent growth of global markets and emergence of cutting-edge technologies, construction jobsites remain among the most dangerous workplaces with extremely high accidents rates. In the safety management process, it is necessary to review and assess a variety of safety documents and data. Numerous public agencies provide open access to accident data, safety checklists, templates and other useful documents for safety management. However, these safety related contents are multifarious, fragmented and often stored in unstructured ways. As a result, the process of finding and utilizing required safety information for specific safety tasks is challenging and inefficient. In order to address this issue, this paper proposes an innovative approach using linked data and semantic web technologies for integrating and sharing construction safety information from diverse sources. A construction safety contents ontology is developed, and construction safety information from accident cases, Job Hazard Analyses (JSA), safety rules and training contents is integrated through RDF formats and SPARQL queries. In order to enable convenient and effective retrieval of information, this study develops a categorization and classification of safety-related contents for safety management processes. The proposed system is expected to improve access and sharing construction safety information, reduce data search time and improve data accuracy. Through this, the approach may enable reductions in safety incidents and overall improvements in safety management and performance.

Keywords –

Building Construction Safety, Linked Data, Safety Information Management;

1 Introduction

Construction remains one of the most accident-prone industries, with recurring injuries and fatalities, even though it is one of the largest contributors to national economies worldwide [1]. Despite strongly-enforced safety policies and continuous efforts from safety professionals, the construction industry reports an incident rate which doubles the industrial average. In addition to injuries and fatalities, safety incidents also result in serious project delays and cost overruns. During the various stages of safety management, it is necessary to access and review a plethora of safety documents and data. The advent of the internet and web-based services allows the convenient online publication and sharing of safety contents such as accident cases, rules and regulations, training documents and checklists. In this regard, various public agencies such as the Occupational Safety and Health Administration (OSHA) [2] and National Institute for Occupational Safety and Health (NIOSH) make an array of useful safety resources available to the public domain. However, even though multifarious safety contents are accessible, finding pertinent safety information for specific safety tasks is difficult. This issue can be attributed to the fact that safety information is unstructured, fragmented and dispersed across various sources on the web.

In order to improve safety management and safety outcomes in construction, numerous research studies have been conducted, focusing on visualization technology for enhanced safety and health training [3]; real-time safety monitoring systems [4] and BIM-integrated, automated safety planning and management systems [5] . However, despite the advancement of information technology in the building and construction industry, a missing link exists between safety management and information models [6]. Very few research efforts have considered improving construction safety information management in order to enable more efficient and systematic retrieval, utilization and sharing of safety contents. Furthermore, traditional database systems and keyword searches are limited in improving information discovery and retrieval [7]. Hence, the development of an accurate and efficient information management approach and system is crucial to enable users to conveniently search and access safety data and information.

This study takes advantage of state-of-the art information technologies such as ontologies, linked data and the semantic web to address the current issues in construction safety information sharing. Ontologies are considered an ideal tool for the explicit formal specification of domain knowledge and modelling of semantic relationships. Moreover, ontologies play a central role in publishing and integrating structured data on the Web through linked data [8]. Through the innovative application of these technologies, this study aims to develop a linked data system for retrieving and sharing construction information from diverse data sources. To this end, the structure and characteristics of construction safety information are first analyzed; safety contents are classified; and then formalized through the development of a construction contents ontology. The paper presents a preliminary study to address how construction safety data can be systemized through the RDF format, and how information can be retrieved efficiently through SPARQL queries. This paper is organized as follows: After the introduction, section 2 highlights the current state and issues regarding construction safety information sharing and management. Next, section 3 reviews prominent information management techniques such as ontology, semantic web and linked data in related research. Section 4 analyzes and classifies construction safety information, and then a linked data system framework for sharing and retrieving construction safety information is presented in section 5. Lastly, the paper is concluded with a discussion of findings and considerations for future works.

2 Current Issues in Sharing Construction Safety Information

During the safety management process, a myriad of safety data, information and documents are generated, referred to and stored in various ways. As a result of the dynamic and complicated nature of construction work, safety management tasks tend to be challenging [6]. Moreover, due to the unstructured nature and diverse formats and characteristics of construction safety data, it is difficult to retrieve information related to specific safety management task. This section discusses the current issues in sharing construction safety information based on literature reviews and usage analyses of safety data.

Firstly, it is necessary to clarify several key terminologies such as construction safety data, safety information and safety contents in order to understand the paper. Data can be defined as statistics or values of quantitative or qualitative variables, belonging to a set of items. As pointed out by Le et al [1], data is the lowest level of abstraction, followed by information, which is defined as a message, or collection of messages in an organized sequence, justifying a change in a construct that represents physical or mental experience or another construct [9]. In the context of this paper, construction safety data includes accident statistics, records and reports etc., while construction safety information is defined as useful descriptions and facts including rules and regulations, safety training materials and so on. Lastly, safety contents refers to a variety of safety related documents such as checklists, news articles, templates, and reports etc.

Construction is an information-driven industry, producing and requiring copious amounts of data throughout project tasks. Some information and data are stored in corporations' databases, while some are dispersed in unstructured formats across various agencies systems, online databases and websites. Safety information and knowledge remains fragmented in safety regulations, accident records, best practices and safety experts experiences [6]. Unstructured safety data are stored with differing formats, hence difficulties are encountered in comprehensively searching, analyzing and accessing desired information. Moreover, without a machine-readable format, it is tedious and difficult to manually generate insights from the vast amounts of existing information.

Within this study, conventionally unstructured construction safety information and data from accident and fatality cases, rules and regulations, safety training contents, news, document templates and statistics are considered. Accident cases are usually analyzed and published as casebooks, or shared online as news articles, or individually as data files for safety reporting procedures. However, these files, articles and casebooks exist in unstructured text-based formats, such as PDFs, Word documents, PowerPoint presentations, spreadsheets, HTML pages and so on. Similarly, safety training materials such as posters, lecture presentations, videos and interactive games are unstructured and scattered over private domain databases and publicly on the web. This issue is further exacerbated by language barrier problems faced by migrant workers [10], and the need for educational resources in various languages and mediums. In order to address these issues, it is necessary to understand the broad structure and characteristics of safety data, information and contents.

Information from accident cases, rules and regulations and training contents typically contain construction context information. Context is defined as a situation in which construction entities exist. An entity can be a person involved in a process, an action to complete a task, a resource, or a building component. Regarding construction safety, context information that needs to be stored includes work trade or type, work phase related work space, work elements, and hazard/accident type. In addition, detailed safety information pertaining to the severity and impact of safety incidents can enable more specific analysis of cases, articles and so on. Similarly, in the context of educational contents, document format, medium and language information can enable more accurate content retrieval. Through context information and additional detailed information, project personnel can acquire valuable resources, knowledge not only for retrieving data at the specific task level, but also for interpreting safety data and generating insights at the corporation level. In order to enable users to systematically retrieve safety information, it is necessary to not only store safety data in a machine readable format, but also to describe safety domain information with a standardized structure. Through this, safety information from accident cases, rules and regulations, training contents, news, document templates and statistics can be integrated in a structured manner.

In addition to the lack of structure, available methods for finding pertinent safety information are also limited. In order to find specific safety information, personnel are required to manually search through casebooks, online databases, rule and regulations and so on. However, because these contents have differing categorizations according to hazard type, work type etc. the information retrieval process tends to be tedious. Safety data stored as HTML or document files can be accessed through keyword searches, based on content titles and descriptions. However, keyword based search is problematic because various forms of safety information are shared subjectively without any coherent structure or standard vocabulary. Hence, the development of improved safety information search mechanisms which take these issues into account is necessary.

3 Information Management Technologies and Approaches in Construction

3.1 Linked Data Approaches

The Semantic Web, which is an extension of the current Web not only allows the upload of data on the web, but also supports making links and connections to enable manual or machine based exploration of web data. In essence, the semantic web of data provides a framework that data sharing and reuse between applications, corporations and community boundaries (W3C). Linked data provides a publishing paradigm whereby the web is used to create typed links between data from different sources [11]. As noted by Berners-Lee [12], linked data is based on the following four basic principles: 1) the use URIs as names for things, 2) the use HTTP URIs so that people can look up those names, 3) When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL) and lastly, 4) include links to other URIs so that they can discover more things. In linked data, all information is encoded with the Resource Description Framework (RDF), enabling anyone to refer to anything. Project participants can look up any URL in an RDF graph over the global web of data, and RDF triples can be used to explore this data space. Furthermore, data publishers may provide RDF dataset dumps for local replication of data, and SPARQL endpoints for querying the data directly [8].

Various domains have applied linked data techniques to describe resources such as media and scientific publications. In the public sector, the Open Government Data (OGD) movement emerged with the goal of meeting the diverse needs of data consumers, and encouraging government transparency by releasing government data on the web [13]. However, the heterogeneous formats in which OGD raw datasets are typically published create interoperability and usability challenges. To address these limitations, Linked Open Government Data (LOGD) has emerged based on linked data technologies [13].

3.2 Ontology and Related Technologies

Ontologies are capable of resolving the aforementioned ambiguity issues in knowledge sharing and reuse through the explicit specification of concepts, attributes and relationships within a domain. The e-COGNOS research project is a prominent example of domain ontology development for the construction industry. It supports consistent knowledge representation and access and use of knowledge at various organizational levels. Some works have also focused on developing specific ontologies at the task level. For instance, Niu and Issa [14] presented a claim ontology to formally share claim knowledge with various project participants and stakeholders within the claim workflow. Le et al [1] proposed integrating semantic wiki, ontologies and social networks to facilitate construction safety knowledge sharing. More recently, Lee et al [8] developed a defect ontology as part of a linked data system for sharing and managing information from construction defect cases. In construction, and other industries alike, access to knowledge-based software for organizational modelling is now easier than ever. For instance, free and open source ontology editors such as Protégé enable users to not only establish and populate hierarchical ontologies, but also to build entirely new ontology classes. In essence, ontologies are powerful technologies which can contribute to the systemization of construction safety information, and through this facilitate enhanced information retrieval, sharing and management. They provide means to semantically represent knowledge with great expressive capacity through defining concepts, relationships, and axioms which facilitate query and reasoning under a variety of conditions. The following section discusses the analysis and categorization of construction safety information, which is an essential process for developing a construction safety contents ontology.

4 Analysis and Categorization of Construction Safety Information

In order to limit the scope of this initial study, safety contents solely from OSHA were considered, since OSHA is recognized as a reputable authority on occupational safety and health, and recognized as a reliable provider of widely used safety data, information and contents. The OSHA website was analyzed, and various safety contents were examined. These include, but are not limited to accident cases, fatality reports, safety rules and regulations, safety training requirements, training materials, training requirement guidelines, safety management templates, safety statistics and so on. The main goal of the analysis was to identify the various classification categories the safety data on the web could be grouped into. Some categories pertained to background information, which described specific dates, times and locations. This type of data is pertinent to specific incidents and statistics data, hence it is not broadly applicable to most safety contents. Next, context information was defined to include work type, work trade, elements, material, space and work phase. Context information focuses on the circumstances around a safety matter. Currently, there are overlaps between categories such as work type and work trade. However, for the purpose of comprehensiveness, both categories are included in this interim study. Lastly, detailed information comprises of accident and hazard types, severity and impact, cause, prevention measure, related personnel/responsible personnel, medium and language categories.

Furthermore, in order to improve interoperability, several defined categories can be based on the Omniclass strategy for classifying the built environment. The Omniclass construction classification system incorporates extant systems such as Masterformat and Uniformat, and provides a classification structure useful for organizing information for electronic databases (Omni site). In the context of this study, Omniclass table 21 (elements) can serve as a basis for the elements category, table 22 (work results) can cover work type and work trade categories, while table 13 (spaces by function) pertains to the related spaces category. The aforementioned categories form a foundation for developing a safety contents ontology to facilitate the sharing of safety information through linked data.

5 Framework for Sharing Safety Information through Linked Data

In order to link the fragmented information flows of construction safety information, an initial framework is proposed by adopting ontologies and linked data technologies. Since this framework is based on ongoing research, the details of some processes such as information sharing between project participants are excluded from this paper. The proposed framework for sharing construction safety information is illustrated in figure 1. Firstly, the development of a standardized safety data structure is necessary to serve as a basis for the systems desired information retrieval and sharing features. As discussed in the previous section unstructured web based safety data are analyzed, classified and categorized. In order to structure construction safety data, a preliminary safety contents ontology has been proposed with the Protégé ontology authoring tool. This process is described as safety contents ontology in figure 1. Lastly, data search functions are proposed through query functionality in the ontology editing tool. This process is shown as query in figure 1. The linked data system underpinnings and features and are described in further detail as follows.

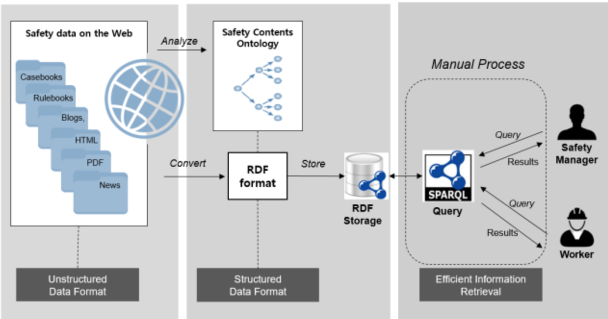

Figure 1: Framework for Sharing Safety Information through Linked Data

5.1 Development of Safety Contents Ontology

Establishing an agreement to use the same terms coherently for a domain of interest is one of the core ontology principles. In the specification and conceptualization stages of ontology development, concepts are organized into a superclass–subclass hierarchical structure, otherwise known as a tree structure. For this study a common vocabulary and an ontological structure are proposed to explain interrelations between safety contents. These are established based on content analyses, comprehensive literature reviews and expert interviews.

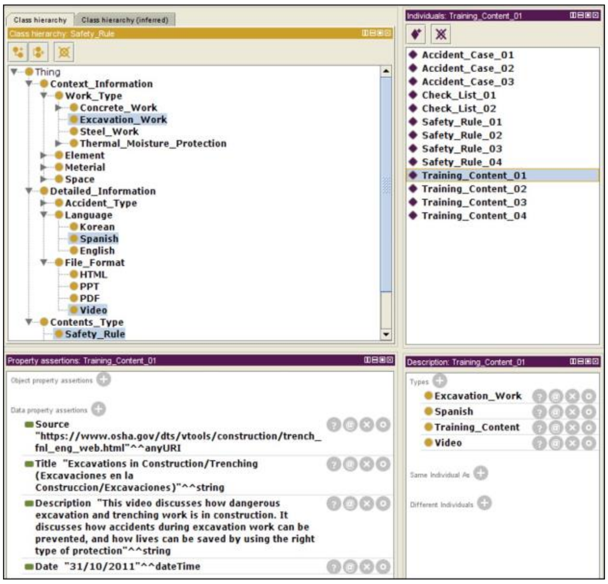

Figure 2. Development of Safety Contents Ontology

Thus far, the safety contents ontology comprises of ten classes and seven data properties. Each of the classes have several subclasses, which also have further subclasses. For instance, *Fabricated_steel_column* and *Structural_steel_for_buildings* would fall under the *Steelwork* subclass in the *Work_type* class. The terms used in each subclass are based on literature reviews and Omniclass terminology and vocabulary. The data properties include a title, description, date, specification and source. Data properties have different *DataPropertyValues*, depending on the data type, such as text data, numerical data and URI data. For example, the Title and Description data properties refer to a *String*, *Impact_cost and Time* refer to an integer, Date refers to a *dateTime* value, and Specification and Source refer to an *anyURI* value for linking actual safety data, documents or web pages. In the context of an OSHA accident case, text data would include the title of an accident case, e.g. "Employee is Struck By Crane Boom and Killed"; with the following description data "an employee was removing connecting pins from a crane boom, which fell and struck him, causing crushing injuries and multiple traumas which lead to his death". *Impact cost* and *Time* would provide numerical values for the financial cost or temporal impact in the case of a lost time incident. The *dateTime* value would be the date the accident occurred and specification and source would be the *anyURI* link which would connect to the actual data on the OSHA website.

5.2 Accessing Safety Information through SPARQL Queries

In order to maximize data sharing, efficient data publishing and search techniques are also necessary. To implement this, safety data should be converted to the RDF format, and published in an RDF store, and then contents would be accessible through the use of SPARQL endpoints. All information is defined using the classes in the safety contents ontology illustrated in figure 2. As the ontology schema is published online, their relationship can be traced whenever SPARQL queries are made on the web. Furthermore, aside from type information, data properties can be expressed explicitly in the RDF file. Data properties are formed as triples in the file and are easy to retrieve in response to various data queries and further analysis.

In order to assess the possibility of retrieving safety information through the proposed system, a variety of safety contents were analyzed and categorized based on the classes hierarchy presented in figure 2. Two example query scenarios for specific safety related purposes are tested in the Protégé platform. The preliminary results show the potential of the approach to retrieve beneficial information from organized RDF data. The first scenario considers a case whereby a safety manager is looking for safety rules, regulations and requirements related to falls during steelwork for the construction of a new building. In this query, a variable representing the cases of interest is used and listed as *?Case*. SELECT is a SPARQL reserved word for listing the query results of the variable. All the conditions are specified in the WHERE section. It is stated that data of interest should all belong to the classes, *dm:steelwork* subclass of *Work_type*, and *dm:falls* under the *occupational accident* subclass of *accident_type, and* the *dm:Safety_rule* under *Contents_type*. As shown in figure 3, the prefix dm: represents the URI resource of the defect ontology. The prefix is similar to that of other RDF standard ontologies like rdf:, rdfs:, and so on. The second query example refers to a scenario whereby a safety manager is looking for educational video materials to help a Spanish worker understand the safety issues related to excavations (figure 3). It is stated that data of interest should all belong to the classes, *dm:excavation* under the subclass of *Work_type*, the *dm:spanish* subclass of *Language, the dm:video* under the *File_format* subclass and lastly the *dm:Training_content* under *Contents_type*.

SPARQL Query

```
SELECT    ?Case    ?anyURL
   WHERE {
      ?Case    rdf:type     dm:excavation
      ?Case    rdf:type     dm:spanish
      ?Case    rdf:type     dm:video
      ?Case    rdf:type     dm:file_format
      ?Case    dm:source    ?anyURL
   }
```

Query Result:

Case	anyURL
Training_content_01	*https://www.osha.gov/dts/vtools/construction/trench_fnl_eng_web.html*

Figure 3: SPARQL Query and Result

In both query scenarios, specific relevant information relating to excavation work and falls during steelwork was retrieved, in the *anyURL* format, through the *?anyURL* query. For the training contents query, a link to an educational excavations video from OSHA is retrieved. Through this approach, safety managers and construction personnel can directly access specific safety data, information and contents conveniently and accurately. This paper only presents a preliminary trial of the proposed linked data system, hence more thorough and in depth evaluation of the approach will be necessary.

6 Conclusion

This study presents a novel framework forming part of an ongoing research for sharing construction safety information through linked data. The paper reviews the characteristics of construction safety information and discusses the current issues encountered with regards to integration, storage, retrieval and management of safety data. Construction accident cases, rules and regulations, training documents and statistics are analyzed and an interim categorization for safety information is presented. A construction safety ontology is proposed to formalize construction safety knowledge, and a linked data based system framework is presented to enable the convenient and efficient retrieval of contents for specific safety tasks. The proposed system is expected to not only improve access to construction safety information, but also to reduce data search time and improve data accuracy. Through this, the approach may enable reductions in safety incidents and overall improvements in safety management and performance. Future research will focus on the full scale system development, with a comprehensive safety contents ontology. In addition, it will be worthwhile to consider integrating the proposed system with BIM environments.

Acknowledgement

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIP) (No. NRF-2016R1A2B3016694).

References

1. Le, Quang Tuan, Do Yeop Lee, and Chan Sik Park. "A social network system for sharing construction safety and health knowledge." *Automation in construction* 46, 30-37 2014.
2. Occupational Safety and Health Administration. Online: https://www.osha.gov/, Accessed: 20/03/2017
3. Sacks, Rafael, Amotz Perlman, and Ronen Barak. "Construction safety training using immersive virtual reality." *Construction Management and Economics* 31.9, 1005-1017,2013.
4. Teizer, Jochen, et al. "Autonomous pro-active real-time construction worker and equipment operator proximity safety alert system." *Automation in Construction* 19.5, 630-640 2010.
5. Zhang, Sijie, et al. "Building information modeling (BIM) and safety: Automatic safety checking of construction models and schedules." *Automation in Construction* 29, 183-195, 2013.
6. Zhang, Sijie, Frank Boukamp, and Jochen Teizer. "Ontology-based semantic modeling of construction safety knowledge: Towards automated safety planning for job hazard analysis (JHA)." *Automation in Construction* 52, 29-41 2015
7. Wang, Han-Hsiang, Frank Boukamp, and Tarek Elghamrawy. "Ontology-based approach to context representation and reasoning for managing context-sensitive construction information." *Journal of Computing in Civil Engineering* 25.5, 331-346, 2010.
8. Lee, Do-Yeop, et al. "A linked data system framework for sharing construction defect information using ontologies and BIM environments." *Automation in Construction* 68, 102-113. 2016
9. Merriam-Webster Dictionary. Online , https://www.merriam-webster.com/ Accessed 21/03/2017
10. Bust, Philip D., Alistair GF Gibb, and Sarah Pink. "Managing construction health and safety: Migrant workers and communicating safety messages." *Safety Science* 46.4 585-602, 2008.
11. Kauppinen, Tomi, and Giovana Mira de Espindola. "Linked open science-communicating, sharing and evaluating data, methods and results for executable papers." *Procedia Computer Science* 4 726-731, 2011.
12. Berners-Lee, Tim. "Linked data-design issues (2006)."
13. Ding, Li, et al. "TWC LOGD: A portal for linked open government data ecosystems." *Web Semantics: Science, Services and Agents on the World Wide Web* 9.3. 325-333 2011.
14. Niu, Jia, and Raja RA Issa. "Framework for production of ontology-based construction claim documents." *Computing in Civil Engineering*. 9-16, 2012.